

Asymptotic theory of outlier detection algorithms for linear time series regression models

Rejoinder

SØREN JOHANSEN

Department of Economics, University of Copenhagen and CREATES, University of Aarhus

BENT NIELSEN

Nuffield College and Department of Economics, University of Oxford

Running headline : Outlier detection algorithms: Rejoinder

6 December 2015

We would like to thank the discussants for their stimulating observations and questions. Most of the discussants point out that we need to analyze the situation when there are outliers, and we of course wholeheartedly agree with that. We need to formulate relevant cases of outliers and discuss the efficiency with which the various methods can identify the outliers. We need to discuss the methods in the presence of outliers, using influence functions and breakdown points and we need to go beyond the concept of the gauge. We therefore define the detection probability for declaring observation y_i an outlier by

$$\delta_{i,n} = \mathbf{P}(y_i \text{ declared outlier}). \quad (\text{R.1})$$

Considering a situation with n variables of which the subset \mathcal{O} are outliers in the population, we define the population gauge of an outlier detection algorithm as

$$\gamma_n = \frac{1}{n - \#\mathcal{O}} \sum_{i \notin \mathcal{O}} \mathbf{P}(y_i \text{ declared outlier}), \quad (\text{R.2})$$

see equations (45), (51), and Doornik & Hendry (2014, p. 122). If there is only one outlier, the detection probability for that observation is the same as the potency concept, see Doornik & Hendry (2014, p. 122), and the average power studied by Riani, Cerioli, Atkinson & Perotta (2014). Ideally, we would like the detection probability $\delta_{i,n}$ to be close to the gauge for non-outliers, $i \notin \mathcal{O}$ and close to one for outliers, $i \in \mathcal{O}$.

Outliers can be defined using two strategies: by generating innovations ε_i and then add a constant d to one or more of these or by replacing ε_i by a constant d . Both methods appear in the discussion, so we summarize these here for convenience,

$$\text{‘stochastic outlier’}: \varepsilon_i \leftarrow \varepsilon_i + d \quad \text{or} \quad \text{‘deterministic outlier’} \varepsilon_i \leftarrow d. \quad (\text{R.3})$$

Atkinson, Cerioli and Riani refer to a ‘complicated empirical rule’ for stopping the forward search, see Riani, Atkinson & Cerioli (2009). The idea is to use different pointwise confidence bands at different points in the Forward Search. The analysis in our Theorem 9 can be adopted to this situation by varying the factor q with m . One approach would be to choose q_m exactly as specified in their paper and then compute the implied gauge. The other approach would be to choose q_m up to a scaling constant and then derive that scaling constant from the desired gauge. The ‘complicated empirical rule’ involves F quantiles. This may be useful in a range of finite samples.

We agree that finite sample corrections are needed, and that an F modification is a good starting point. But it appears that the heuristic argument leading to the ‘complicated empirical rule’ is too simple. In the forward search, the order statistic \hat{z}_ψ estimates σc_ψ , where c_ψ is the population quantile and σ the unknown scale. Thus, we compare the scaled order statistic $\hat{z}_\psi / \hat{\sigma}_{\psi, \text{corr}}$ with c_ψ , which has expansion

$$\frac{\hat{z}_\psi}{\hat{\sigma}_{\psi, \text{corr}}} - c_\psi = \frac{\hat{z}_\psi - \hat{\sigma}_{\psi, \text{corr}} c_\psi}{\hat{\sigma}_{\psi, \text{corr}}} = \frac{\hat{z}_\psi - \hat{\sigma}_{\psi, \text{corr}} c_\psi}{\sigma + o_P(1)},$$

since $\hat{\sigma}_{\psi, \text{corr}}$ is consistent for σ . Corollary 3 shows that the asymptotic variance has two components, of which the first is a Bernoulli component $\psi(1 - \psi)$ arising from \hat{z}_ψ / σ . The ‘complicated empirical rule’ does not appear to take the variance of $\hat{\sigma}_{\psi, \text{corr}}$ into account. It may therefore not perform so well in large samples, see the discussion of Theorem 3.6 in Johansen & Nielsen (2015a).

ACR reanalyze the Fulton fish data using a Forward Search with a 10% initial sample. The gauge is not reported. The asymptotic theory in Theorem 9 and Table 4 shows that the cut-off values should be chosen larger, when the initial sample is smaller. This may be part of the reason that the ACR results appear less unstable. Indeed, if we re-run the analysis in Figure 4 with a 50% initial sample then the later part of the Forward Search path is unchanged, but the cut-off values are moved further out. In any case the holiday dummies appear to remain

candidates for outliers. Even if they are not particular large outliers, they have an interesting institutional interpretation.

ACR recommend that cut-off values should be chosen large, so that the gauge is small to avoid too many false signals. This reasoning motivated us in developing the Poisson limit results for gauges.

Croux and Wilms are interested in the situation with a non-symmetric reference distribution. The empirical process theory underlying this paper allows for non-symmetry. So there is potential for extending the results. However, in the first instance we would consider non-symmetry along with the situation with outliers.

CW also ask for analysis of cut-offs that are depending on the sample size in other ways than the Poisson exceedance theory. A relevant theory of empirical processes is work in progress.

CW provide a simulation study of gauges and power for a range of different methods and $n = 100$. The main points are that the sample gauges are relatively stable across a range of different m -step methods, but leverage points are only detected well if the initial estimator is robust. We would like to supplement this study.

First, we consider the situation without outliers to show that the asymptotic results are attained for large n . To simplify the discussion we choose the data generating process to be a first order autoregression

$$y_i = \alpha + \phi y_{i-1} + \varepsilon_i, \quad y_0 = \alpha/(1 - \phi), \quad (\text{R.4})$$

with $\phi = 0.4$ and $\alpha = 0.5$ and ε_i independent $\text{N}(0, 0.1^2)$. Table R.1 reports the gauge when there are no outliers. The initial empirical gauge and 1-step empirical gauge are denoted $\hat{\gamma}^{(0)}$, $\hat{\gamma}^{(1)}$, see (45). In particular, $\hat{\gamma}^{(0)}$ is a Monte Carlo average of $n^{-1} \sum_{i=1}^n 1_{(|y_i - x'_i \hat{\beta}_{LS}| > \hat{\sigma}_{LSc})}$. The consistency factor ς is needed when considering iterations, see (15). Ignoring this gives the 1-step gauge $\hat{\gamma}_{bias}^{(1)}$. The corresponding sample detection frequencies for declaring observation 25 an outlier are denoted $\hat{\delta}_{25}^{(0)}$, $\hat{\delta}_{25}^{(1)}$, $\hat{\delta}_{25,bias}^{(1)}$. For instance, $\hat{\delta}_{25}^{(0)}$ is a Monte Carlo average of $1_{(|y_{25} - x'_{25} \hat{\beta}_{LS}| > \hat{\sigma}_{LSc})}$. The simulations use 10^4 repetitions and a target gauge of $\gamma = 0.01$. The results support the asymptotic theory and in particular the role of the consistency corrections.

[Table R.1 about here]

Secondly, we consider 2 types of a single ‘deterministic outlier’ where we replace random variables with constants along the line of (R.3).

Innovative outlier: generate $\varepsilon_1, \dots, \varepsilon_{100}$, replace ε_{25} by d , generate y_i from equation (R.4).

Leverage point: generate $\varepsilon_1, \dots, \varepsilon_{100}$, generate y_i from equation (R.4). Write

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_{100} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & y_0 \\ \vdots & \vdots \\ 1 & y_{99} \end{pmatrix},$$

so that the usual least squares estimator is $\hat{\beta} = (X'X)^{-1}X'Y$. Now, replace $Y_{25} = y_{25}$ and $X_{25,2} = y_{24}$ by d . In a time series analysis that uses lagged dependent variables, such leverage points do not seem common.

[Table R.2 about here]

Table R.2 reports gauge and detection probabilities when there is a single outlier at observation 25. Innovative outliers are considered in the left panel with nearly identical, non-reported figures for additive outliers. Leverage points are reported in the right panel. The gauges are now computed from the remaining $n - 1$ observations, so that for instance $\hat{\gamma}^{(0)}$ is a Monte Carlo average of $(n - 1)^{-1} \sum_{i \neq 25} 1_{(|y_i - x_i' \hat{\beta}_{LS}| > \hat{\sigma}_{LS})}$, while the sample detection probability $\hat{\delta}_{25}^{(0)}$ is defined as before.

For innovative outliers we see that robustified least squares works as intended. The initial gauge $\hat{\gamma}^{(0)}$ decreases to zero with increasing outlier size d . This is presumably because a large outlier distorts the initial variance estimator. However, the outlier is identified with frequency $\hat{\delta}_{25}^{(0)}$ that increases with increasing outlier size d . As a consequence, the gauge $\hat{\gamma}^{(1)}$ recovers to more stable values in the next step. Continued iteration improves the gauge $\hat{\gamma}^{(2)}$ further, although it keeps below the target of $\gamma = 1\%$. This could be a finite sample issue. A simulation for $n = 1000$, but still with one outlier, gives gauges $\hat{\gamma}^{(4)}$ in the interval 0.98% to 1.00% for $0 \leq d \leq 100$.

For leverage points the situation is different. This type of outlier is less of a concern in time series analysis when lagged dependent variables are included in the model. The gauge is now quite stable in the outlier size even in the initial round. The detection probability increases until $d = 1$ and then it drops to zero, which is the leverage effect.

Doornik and Hendry apply the impulse indicator saturation to model the savings rate in the US. They remark that the consistency factors are modest when the gauge is small. While this is correct, let us just point out that there are two factors in play. Consider a normal reference distribution and gauge $\gamma = 1\%$. Then the innovation standard deviation estimator $\hat{\sigma}$ includes the factor $1/\varsigma = 1.039$, see (9, 15). The standard error for the 1-step robustified least squares estimator for β includes a further consistency factor of $\{\eta_{\beta}^{(1)}\}^{1/2} = 1.040$, see (37). Thus, when using standard regression software, the 1-step estimators for the residual variance and for the standard error on β have to be multiplied by 1.039 and 1.080, respectively.

DH also apply the forward search and find more outliers than expected. Could it be a finite sample issue with the asymptotic theory?

Finally DH remark that outliers can be used to test for super exogeneity. Recently, Bazinas & Nielsen (2015) have shown that the super exogeneity concept can be relaxed somewhat and used for analyzing causal transmission channels.

Doornik analyzes an interesting simulation where the data contains two clusters. He shows that by changing the data slightly, the robust estimator switches from describing one cluster to describing the other. The original use of the robust estimators assumed that there was a good model for "the bulk of the data" and the robust estimator should rely on these, so as to avoid contamination. The example of D shows nicely that clusters should be discussed differently, and that the forward search could be useful here. He also demonstrates, that in this example a test for linearity will show that it is not a useful hypothesis.

Oja gives a interesting suggestion for classifying outliers, by allowing either the regressor or the error term or both to be contaminated. He also points out that further results are needed for multivariate linear regression. Such results would be useful for the Taskinen & Oja (2015) model and for the forward search, see Riani, Atkinson & Cerioli (2009).

Further, he comments on the link between outlier detection and the multiple testing problem with a hypothesis of the form $H_0 = \cap_{t=1}^n H_{0t}$. Normally the hypotheses H_{0t} would be different aspects of a substantial theory, whereas in this context H_{0t} is the hypothesis that observation t is not an outlier. Statistical tests trade off errors of the first and the second type. That trade-off gets more extreme in the case of multiple testing, so it may be more useful to control gauge rather than size at conventional levels. We would like to illustrate this issue.

Consider a sample of n standard normal observations y_i , where we assume known location and scale. Let $\xi_i = |y_i|$ have order statistics $\xi_{(1)} \leq \dots \leq \xi_{(n)}$. We will consider the test with critical region $\{\xi_{(n)} > c\}$ instead of the more complicated, but very elegant Simes (1986) test mentioned by Oja.

At first, suppose we only have a few observations of which one happens to be $d = 3$. Compared to standard normal distribution this observation appears to be somewhat large. It is so large that we may wonder if the sample is not normal. The question is how we should react to such an observation in a larger sample with $n = 100$.

[Figure R.1 about here]

We compute the size of the test with critical region $\{\xi_{(n)} > c\}$. Assuming that the random variables y_i are independent standard normal the size is

$$\alpha_n = \mathbf{P}\{\xi_{(n)} > c\} = 1 - \mathbf{P}\bigcap_{i=1}^n (\xi_i \leq c) = 1 - \mathbf{P}(\xi_1 \leq c)^n. \quad (\text{R.5})$$

Solving for c for $n = 100$ and exploiting normality yields $c_\alpha = \Phi^{-1}[\{1 + (1 - \alpha_n)^{1/n}\}/2]$ so that $c_{\alpha=50\%} = 2.70$, $c_{\alpha=10\%} = 3.28$, $c_{\alpha=5\%} = 3.47$ see also Figure R.1. We see that the size is very sensitive to the choice of the critical value c just around the value $d = 3$, where we may be worried about outliers. This is because of the exponential dependence on n in (R.5). Thus, if we want the ability to detect small outliers of magnitude $d = 3$, we need to choose a rather large size.

Now, we turn to the gauge of the test $\{\xi_{(n)} \geq c\}$. This test rejects if any of ξ_1, \dots, ξ_n exceeds c . We now choose the critical value c to control the population gauge

$$\gamma_n = \mathbf{E} \frac{1}{n} \sum_{i=1}^n 1_{(\xi_i > c)} = \mathbf{P}(\xi_1 > c). \quad (\text{R.6})$$

Solving for c and exploiting normality we get $c_\gamma = \Phi^{-1}(1 - \gamma_n/2)$ so that $c_{\gamma=5\%} = 1.96$, $c_{\gamma=1\%} = 2.58$, $c_{\gamma=0.5\%} = 2.81$, see also Figure R.1. We see that the gauge is much less sensitive to the choice of the critical value around the value $d = 3$, and unless the gauge is set very low, we will detect the observation with value $d = 3$ as an outlier. The price we pay is that we typically detect a few ‘good’ observations falsely.

We could interpret the choice between controlling size and gauge in multiple decision problems as a choice between controlling the frequency of rejection in ‘repeated samples’ and in ‘extended observations’. To be specific, when controlling size we control the frequency of rejecting the hypothesis of no outliers in repeated samples. The formula (R.5) with its exponential dependence on n then leads to large critical values for small choices of the size and moderate to large sample lengths. Instead, when controlling gauge we control the frequency of falsely detecting outliers, when we take more observations. We then replace the exponential dependence on n with the formula (R.6) that is independent on n .

Ronchetti asks for an analysis of outlier detection algorithms in terms of the false discovery rate used in multiple testing problems. These were developed by Benjamini & Hochberg (1995) in the tradition of Simes (1986). On p. 290 they discuss the trade-off between the two types of errors in conventional multiple testing and note: ‘This is probably why a per comparison error rate (PCER) approach, which amounts to ignoring the multiplicity problem altogether, is still recommended by some (e.g. Saville (1990))’. When, in the above notation, H_0 is satisfied, Benjamini & Hochberg define the per comparison error rate in the same way as the population gauge. Our analysis indicates that the recommendation of Saville and the later approaches of Hoover & Perez (1999) and Hendry & Santos (2010) can be analyzed formally. In our view the gauge concept adds a quite interesting dimension to the usual conundrum of choosing the level of a test.

Benjamini & Hochberg recognize that controlling size, which they call familywise error rate, may be too stringent in multiple testing problems. This is one of their motivations for the false discovery rate. It is defined as the expectation of $Q = V/R$, where V is the number of falsely rejected hypothesis H_{0t} and R is the total number of rejected hypothesis, see also our Table 1. In the situation where H_0 is satisfied then $R = V$ so that $Q = 1$ when $R > 0$ while Q is not defined when $R = 0$. For that situation Benjamini & Hochberg recommend using $P(R > 0)E(Q|R > 0)$. This, of course reduces to $P(R > 0)$ when H_0 is satisfied so that we are back to controlling size as discussed above.

R also emphasizes that soft rejection methods are often useful and gives an empirical example. He also mentions that the empirical process techniques may be useful in the analysis of quantile regression. Indeed, this is what we are finding in ongoing work.

Finally, R asks if impulse indicator saturation is more robust than the robustified least

squares. Table R.2 indicates that robustified least squares is not robust to leverage, although we note that such leverage points are not so relevant in time series. We now investigate the impulse indicator saturation using the same simulation design. Table R.3 reports the gauge, so that $\hat{\gamma}^{(split)}$ is the gauge based on the two split half least squares estimators. Eliminating the outliers and re-running least squares gives an updated gauge of first $\hat{\gamma}^{(0)}$ and then $\hat{\gamma}^{(1)}$. The overall performance is similar to that of robustified least squares. Table R.4 reports gauges and detection probabilities in the presence of an outlier at $i = 25$. Impulse indicator saturation now deals with a single leverage point nearly as well as with an innovative outlier.

Curiously, the gauge control is now poor for very large innovative outliers. The issue seems to be that the next few observations after the outlier also get picked up as outliers, and the algorithm only sheds these in part. For instance, the detection probability for the ‘good’ innovation $i = 26$ are $\hat{\delta}_{26}^{(split)} = \hat{\delta}_{26}^{(0)} = 0.91$, $\hat{\delta}_{26}^{(1)} = 0.77$, $\hat{\delta}_{26}^{(2)} = 0.66$, $\hat{\delta}_{26}^{(3)} = 0.61$. The reason for falsely detecting innovation $i = 26$ could be as follows. Initially, we estimate α, ϕ based on the second sample half and use that to find outliers in the first sample half. The estimator, $\hat{\phi}$ say, is consistent so that $\hat{\phi} - \phi = O_P(n^{-1/2})$. The residual $\hat{\varepsilon}_{26} = y_{26} - \hat{\alpha} - \hat{\phi}y_{25}$ can be rewritten as $\hat{\varepsilon}_{26} = \varepsilon_{26} - (\hat{\alpha} - \alpha) - (\hat{\phi} - \phi)y_{25}$. Now, if y_{25} is diverging at a rate faster than $n^{1/2}$ then $(\hat{\phi} - \phi)y_{25}$ diverges. The residual is then large and observation $i = 26$ is declared an outlier. Simulations for additive outliers, which are not reported, indicate a somewhat better gauge control.

[Table R.3 about here]

[Table R.4 about here]

Zwanzig asks for a more direct comparison of m -step Huber skip algorithms and the forward search. At present we do not have any preference. We would analyze the question by choosing a common gauge and compare their detection probabilities for different types of outliers. We suspect that in many situations they will perform similarly, but in other situations none of the methods will dominate uniformly.

Z points out that outliers can be of a different nature in time series and in cross section. To that end she discusses the case of innovative outliers. To follow up on that point we would

interpret an innovative outlier as an outlier in the conditional distribution of y_t given the past, not in the marginal distribution of y_t . To be specific, suppose a time series y_t , for instance for the oil price, typically follows a first order autoregression $y_t = \alpha y_{t-1} + \varepsilon_t$, but at time t_0 it is given by $y_{t_0} = \alpha y_{t_0-1} + d + \varepsilon_{t_0}$, which is a ‘stochastic outlier’ in the sense of (R.3). The oil prices $y_{t_0-1}, y_{t_0}, y_{t_0+1}$ are the actual oil prices and should not be dummied out. What is an outlier is the shock at time t_0 , which we remove by a dummy variable. In the Gaussian likelihood we have the terms

$$(y_{t_0-1} - \alpha y_{t_0-2})^2 + (y_{t_0} - \alpha y_{t_0-1} - d)^2 + (y_{t_0+1} - \alpha y_{t_0})^2 \quad (\text{R.7})$$

Whatever the estimate, $\hat{\alpha}$, of α , the estimate of d is $\hat{d} = y_{t_0} - \hat{\alpha} y_{t_0-1}$ giving residual $\hat{\varepsilon}_{t_0} = 0$. Note that the outlying observation y_{t_0} appears in the next term of the likelihood function explaining the mean of the next observation. The various algorithms seem quite reliable in finding such innovative outliers. Indeed, Tables R.2, R.4 indicate that robustified least squares and impulse indicator saturation will find the innovative outlier quite reliably while controlling the gauge.

Z also makes the point that the type II error of not detecting an outlier may be more important than the type I error of falsely classifying a good observation as an outlier. For this reason we may not worry about the issue discussed in the reply to Ronchetti, where we found that the gauge of the impulse indicator saturation is not controlled under extreme outliers. We also note that type II errors are a problem, when forecasting in the presence of an undetected level shift, see Hendry & Nielsen (2007, §21) for an overview.

Finally, Z points out that outlier detection could be done by a LASSO algorithm. More generally, this would allow a simultaneous search over variables and outliers. The Autometrics algorithm by Doornik (2009) does the same. Hendry & Doornik (2014, p. 215f) compare the algorithms in terms of their ability to search over variables. One of their findings is that LASSO does not control gauge, but then it was not designed to do so. This brings our discussion to a full circle in that the Autometrics algorithm is what got us interested in outlier detection algorithms in the first place.

References¹

¹Those marked * are used in main paper. Those marked † are used by discussants.

- Bazinas, V. & Nielsen, B. (2015). Causal transmission in reduced-form models. Nuffield College Discussion Paper.
- * Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57**, 289–300.
- * Doornik, J.A. (2009). Autometrics. In *The methodology and practice of econometrics: a festschrift in honour of David F. Hendry* (eds Castle, J.L. & Shephard, N.), 88–121. Oxford University Press, Oxford.
- * Hendry, D.F. & Nielsen, B. (2007). *Econometric modelling*, Princeton University Press, Princeton NJ.
- * Hendry, D.F. & Santos, C. (2010). An automatic test of super exogeneity. In *Volatility and time series econometrics: essays in honor of Robert F. Engle* (eds Bollerslev, T., Russell, J.R. & Watson, M.W.), 164–193. Oxford University Press, Oxford.
- * Hoover, K.D. & Perez, S.J. (1999). Data mining reconsidered: encompassing and the general-to-specific approach to specification search (with discussion). *Econom. J.* **2**, 167–191.
- * Johansen, S. & Nielsen, B. (2015a). Analysis of the forward search using some new results for martingales and empirical processes. To appear in *Bernoulli J.* DOI: 10.3150/14-BEJ689.
- † Riani, M., Atkinson, A.C. & Cerioli, A. (2009). Finding an unknown number of multivariate outliers. *J. R. Statist. Soc. B* **71**, 447–466.
- † Riani, M., Cerioli, A., Atkinson, A.C. & Perrotta, D. (2014). Monitoring robust regression. *Electronic J. Statist.* **8**, 642–673.
- Saville, D.J. (1990). Multiple comparison procedures: The practical solution. *Amer. Statistician* **44**, 174–180.
- † Simes, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**, 751–754.

† Taskinen, S. & Oja, H. (2015). Influence functions and efficiencies of k-step Hettmansperger-Randles estimators. To appear.

Table R.1: Gauges and detection probabilities for robustified least squares when there are no outliers and $\gamma = 0.001$, $MCse = 0.001$.

n	$\hat{\gamma}^{(0)}$	$\hat{\gamma}^{(1)}$	$\hat{\gamma}_{bias}^{(1)}$	$\hat{\delta}_{25}^{(0)}$	$\hat{\delta}_{25}^{(1)}$	$\hat{\delta}_{25,bias}^{(1)}$
100	0.0085	0.0092	0.0108	0.0100	0.0110	0.0126
1000	0.0099	0.0100	0.0130	0.0100	0.0099	0.0128

Table R.2: Gauges and detection probabilities when there is one outlier at $i = 25$ and $n = 100$, $\gamma = 0.01$, $MCse = 0.001 - 0.005$.

d	Innovative Outlier					Leverage point				
	$\hat{\gamma}^{(0)}$	$\hat{\gamma}^{(1)}$	$\hat{\gamma}^{(2)}$	$\hat{\delta}_{25}^{(0)}$	$\hat{\delta}_{25}^{(1)}$	$\hat{\gamma}^{(0)}$	$\hat{\gamma}^{(1)}$	$\hat{\gamma}^{(2)}$	$\hat{\delta}_{25}^{(0)}$	$\hat{\delta}_{25}^{(1)}$
0.1	0.0082	0.0089	0.0090	0.052	0.054	0.0084	0.0090	0.0092	0.023	0.024
0.2	0.0073	0.0081	0.0083	0.248	0.255	0.0081	0.0088	0.0088	0.071	0.074
0.3	0.0061	0.0076	0.0079	0.610	0.618	0.0076	0.0084	0.0086	0.161	0.167
0.4	0.0047	0.0072	0.0077	0.895	0.899	0.0071	0.0081	0.0083	0.289	0.298
0.5	0.0034	0.0069	0.0076	0.986	0.987	0.0066	0.0079	0.0082	0.426	0.436
1	0.0002	0.0062	0.0075	1.000	1.000	0.0058	0.0078	0.0083	0.755	0.761
2	0.0000	0.0062	0.0075	1.000	1.000	0.0068	0.0076	0.0078	0.211	0.212
3	0.0000	0.0062	0.0075	1.000	1.000	0.0077	0.0082	0.0083	0.000	0.000

Table R.3: Gauges for impulse indicator saturation when there are outliers and $\gamma = 0.01$, $MCse = 0.001$.

n	$\hat{\gamma}^{(split)}$	$\hat{\gamma}^{(0)}$	$\hat{\gamma}^{(1)}$	$\hat{\gamma}_{bias}^{(0)}$	$\hat{\gamma}_{bias}^{(1)}$
100	0.0153	0.0115	0.0110	0.0144	0.0126
1000	0.0104	0.0101	0.0101	0.0132	0.0131

Table R.4: Gauges and detection probabilities for impulse indicator saturation when there is one outlier at $i = 25$ and $n = 100$, $\gamma = 0.01$, $MC_{se} = 0.001 - 0.005$.

ϵ	Innovative Outlier					Leverage point				
	$\hat{\gamma}^{(split)}$	$\hat{\gamma}^{(0)}$	$\hat{\gamma}^{(1)}$	$\hat{\delta}_{25}^{(split)}$	$\hat{\delta}_{25}^{(0)}$	$\hat{\gamma}^{(split)}$	$\hat{\gamma}^{(0)}$	$\hat{\gamma}^{(1)}$	$\hat{\delta}_{25}^{(split)}$	$\hat{\delta}_{25}^{(0)}$
0.1	0.0149	0.0112	0.0107	0.073	0.063	0.0150	0.0114	0.0109	0.042	0.032
0.2	0.0138	0.0105	0.0101	0.298	0.279	0.0144	0.0110	0.0106	0.120	0.096
0.3	0.0122	0.0100	0.0097	0.652	0.641	0.0137	0.0106	0.0102	0.284	0.232
0.4	0.0107	0.0096	0.0096	0.913	0.910	0.0131	0.0104	0.0101	0.498	0.429
0.5	0.0097	0.0094	0.0095	0.989	0.989	0.0128	0.0103	0.0100	0.698	0.635
1	0.0091	0.0093	0.0096	1.000	1.000	0.0127	0.0102	0.0101	0.993	0.991
2	0.0122	0.0104	0.0105	1.000	1.000	0.0135	0.0102	0.0101	1.000	1.000
100	0.0496	0.0387	0.0324	1.000	1.000	0.0142	0.0103	0.0101	1.000	1.000

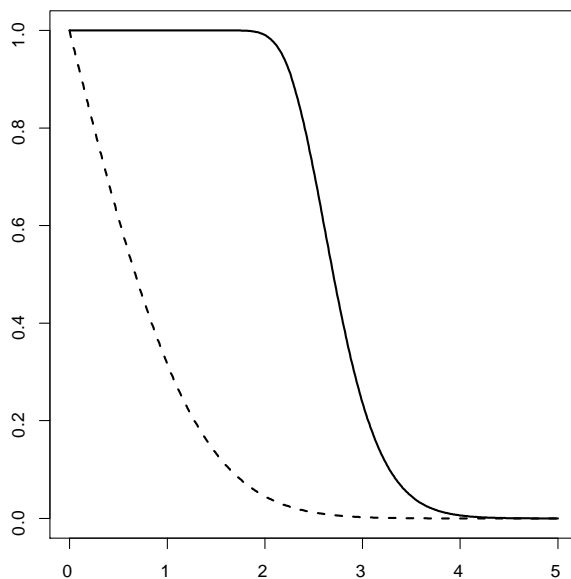


Figure R.1: Size (solid line) and gauge (dashed line) as a function of cut-off value for the test with critical region $\{\xi_{(n)} > c\}$.