HIGHLY ROBUST ESTIMATION OF THE AUTOCOVARIANCE FUNCTION

By Yanyuan Ma and Marc G. Genton

Massachusetts Institute of Technology

First version received August 1998

Abstract. In this paper, the problem of the robustness of the sample autocovariance function is addressed. We propose a new autocovariance estimator, based on a highly robust estimator of scale. Its robustness properties are studied by means of the influence function, and a new concept of temporal breakdown point. As the theoretical variance of the estimator does not have a closed form, we perform a simulation study. Situations with various size of outliers are tested. They confirm the robustness properties of the new estimator. An S-Plus function for the highly robust autocovariance estimator is made available on the Web at http://www-math.mit.edu/~yanyuan/Genton/Time/time.html. At the end, we analyze a time series of monthly interest rates of an Austrian bank.

Keywords. autocovariance, breakdown point, influence function, robustness, scale estimation.

1. INTRODUCTION

The autocovariance function plays an important role in time series analysis. For example, it is often used to study the underlying dependence structure of the process (Box and Jenkins, 1976; Brockwell and Davis, 1991). This is an important step towards constructing an appropriate mathematical model for the data. Therefore, it is important to have a sample autocovariance function which remains close to the true underlying autocovariance function, even when outliers, i.e. faulty observations, are present in the data. Otherwise, important goals of the time series analysis such as inference or forecasting can be non-informative. In fact, experience from a broad spectrum of applied sciences shows that measured data may contain 10–15% of outlying values (Hampel, 1973) due to gross errors, round-off errors, measurement mistakes, faulty recording, etc., and this proportion can even reach 30% (Huber, 1977). Unfortunately, the widely used sample autocovariance function based on the methods of moments is not robust against outlying values in the data. In this paper, we propose and study a new estimator for the autocovariance function, based on a highly robust estimator of scale.

Consider a time series $\{X_t: t \in \mathbb{Z}\}$ and assume that it satisfies the hypothesis of second-order stationarity:

0143-9782/00/06 663–684 JOURNAL OF TIME SERIES ANALYSIS Vol. 21, No. 6

(i) $\mathrm{E}(X_t^2) < \infty \quad \forall t \in \mathbb{Z}$
(ii) $\mathrm{E}(X_t) = \mu = \text{constant} \quad \forall t \in \mathbb{Z}$
(iii) $\mathrm{cov}(X_{t+h}, X_t) = \gamma(h) \quad \forall t, h \in \mathbb{Z}$

where $\gamma(h)$ is the autocovariance function of X_t at lag h. The classical estimator for the autocovariance function, based on the method of moments, on a sample $\mathbf{x} = (X_1, \ldots, X_n)^{\mathrm{T}}$, is

$$\hat{\gamma}_{\mathrm{M}}(h, \mathbf{x}) = \frac{1}{n-h} \sum_{i=1}^{n-h} (X_{i+h} - \overline{X})(X_i - \overline{X}) \qquad 0 \leqslant h \leqslant n-1 \tag{1}$$

where

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Note that, to ensure positive definiteness of the sampling autocovariance matrix, each entry should be estimated by

$$\frac{n-h}{n} \hat{\gamma}_{\mathrm{M}}(h, \mathbf{x})$$

In the next section, we propose a new sample autocovariance estimator based on a highly robust scale estimator. Section 3 introduces a concept of temporal breakdown point of an autocovariance estimator and discusses its link with the classical breakdown point. The influence function for autocovariance estimators is computed in Section 4, as well as the formula for their asymptotic variance. These results are completed with a simulation study in Section 5, on AR(1) and MA(1) models. The behaviour of the classical and highly robust autocovariance estimator in the presence of outliers is also studied. In Section 6, a time series of monthly interest rates of an Austrian bank is analysed.

2. THE ROBUST ESTIMATOR

The autocovariance function describes the covariance between observations at different time lag distances h. Traditionally, covariance estimation between two random variables X and Y is based on a location approach, because

$$\mathrm{cov}(X, Y) = \mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))]$$

yielding, for example, the sample autocovariance function (1). However, covariance estimation can also be based on a scale approach, by means of the following identity (Huber, 1981; Gnanadesikan, 1997):

$$\mathrm{cov}(X, Y) = \frac{1}{4\alpha\beta} [\mathrm{var}(\alpha X + \beta Y) - \mathrm{var}(\alpha X - \beta Y)] \qquad \forall \alpha, \beta \in \mathbb{R} \tag{2}$$

In general, X and Y may be measured in different units, and the choice

$$\alpha = \frac{1}{\sqrt{\text{var}(X)}} \qquad \text{and} \qquad \beta = \frac{1}{\sqrt{\text{var}(Y)}}$$

is recommended (Gnanadesikan and Kettenring, 1972). However, in the time series setting, X and Y represent the same variable and we set $\alpha = \beta = 1$.

In the context of scale estimation, Rousseeuw and Croux (1992, 1993) proposed a simple, explicit and highly robust estimator Q_n:

$$Q_n(\mathbf{z}) = c\{|Z_i - Z_j|;\ i < j\}_{(k)} \tag{3}$$

where $\mathbf{z} = (Z_1, \ldots, Z_n)^{\mathrm{T}}$ is the sample

$$k = \text{int}\left[\frac{\binom{n}{2} + 2}{4}\right] + 1$$

and int(·) denotes the integer part. The factor c is for consistency: at the Gaussian distribution $c = 2.2191$. This means that we sort the set of all absolute differences $|Z_i - Z_j|$ in increasing order for $i < j$ and then compute its kth order statistic (approximately the $\frac{1}{4}$ quantile for large n). This value is multiplied by c, thus yielding Q_n. Note that this estimator computes the kth order statistic of the $\binom{n}{2}$ interpoint distances. It is of interest to remark that Q_n does not rely on any location knowledge and is therefore said to be location-free. This is in contrast to the classical sample autocovariance function. At first sight, the estimator Q_n appears to need $O(n^2)$ computation time, which would be a disadvantage. However, it can be computed using no more than $O(n \log n)$ time and $O(n)$ storage, by means of the fast algorithm described in Croux and Rousseeuw (1992).

Using the identity (2) and the definition (3) of the scale estimator Q_n, we define the highly robust autocovariance function estimator as follows. Extract the first $n - h$ observations of $\mathbf{x} = (X_1, \ldots, X_n)^{\mathrm{T}}$ to produce a vector $\mathbf{u}$ with length $n - h$ and the last $n - h$ observation of $\mathbf{x}$ to produce a vector $\mathbf{v}$ of length $n - h$, as shown in Figure 1. Then:

$$\hat{\gamma}_{\mathrm{Q}}(h, \mathbf{x}) = \tfrac{1}{4}[Q^2_{n-h}(\mathbf{u} + \mathbf{v}) - Q^2_{n-h}(\mathbf{u} - \mathbf{v})] \tag{4}$$

This turns out to be a highly robust estimator of autocovariance. As shown at the end of Section 3, it has a temporal breakdown point of 25%, which is the highest possible value in the autocovariance case. Note that the highly robust autocovariance estimator $\hat{\gamma}_{\mathrm{Q}}(h, \mathbf{x})$ can also be carried out with $O(n \log n)$ time and $O(n)$ storage.

Another approach to obtain a robust estimator for the autocovariance is by truncating large terms in the sum of equation (1). However, we prefer the scale approach suggested by (2), because it allows the use of the highly robust estimator of scale Q_n, which has a remarkably high asymptotic Gaussian efficiency of 82.27%. For instance, Q_n has already been successfully used in the context of regression (Hössjer *et al.*, 1994; Croux *et al.*, 1994), as well as for variogram estimation (Genton, 1998a) in spatial statistics. Note that rank-

based methods to estimate autocovariances have been proposed by Hallin and Puri (1994) and may have some robustness properties.

Finally, to obtain a highly robust estimator of the autocorrelation function, one could divide the estimator $\hat{\gamma}_{\mathrm{Q}}(h, \mathbf{x})$ in (4) by $Q_n(\mathbf{u})$ and $Q_n(\mathbf{v})$. However, this would not be a natural autocorrelation estimator because it would not be bounded between -1 and 1. Therefore, a highly robust autocorrelation estimator $\hat{\rho}_{\mathrm{Q}}$ would be:

$$\hat{\rho}_{\mathrm{Q}}(h, \mathbf{x}) = \frac{Q^2_{n-h}(\mathbf{u}+\mathbf{v}) - Q^2_{n-h}(\mathbf{u}-\mathbf{v})}{Q^2_{n-h}(\mathbf{u}+\mathbf{v}) + Q^2_{n-h}(\mathbf{u}-\mathbf{v})}$$

so as to insure $|\hat{\rho}_{\mathrm{Q}}(h, \mathbf{x})| \leqslant 1$. Note that $\hat{\gamma}_{\mathrm{Q}}$ depends upon the choice of the constant c appearing in (3), whereas $\hat{\rho}_{\mathrm{Q}}$ is independent of the choice of c. Nevertheless, c can be computed for various distributions, although the Gaussian case is usually preferred.

3. TEMPORAL BREAKDOWN POINT

Outliers in time series can seriously affect the estimation and inference of parameters (Martin and Yohai, 1985; Bustos and Yohai, 1986). The main problem is that estimators which take account of the time series structure are not invariant under permutation of the data, as in the case of estimators for i.i.d. observations. Consequently, distinction between outliers occurring in isolation, in patches, or periodically, becomes important. Three types of outliers are generally considered (Denby and Martin, 1979): innovation outliers (IO), which affect all subsequent observations, and additive outliers (AO) or replacement outliers (RO), which have no effect on subsequent observations. Consider a second-order stationary ARMA(p, q) process $(X_t\colon t \in \mathbb{Z})$ such that for every t:

$$X_t - \rho_1 X_{t-1} - \cdots - \rho_{\ p} X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \qquad (5)$$

where $\rho_1, \ldots, \rho_{\mathrm{p}}$ and $\theta_1, \ldots, \theta_{\mathrm{q}}$ are real parameters, and the innovations are white noise $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$. Subsequently, we assume that the parameters of the ARMA process are defined such that the process is causal and invertible. More details on these notions, as well as necessary and sufficient conditions for causality and invertibility, can be found in Brockwell and Davis (1991).

The ARMA(p, q) process $\{X_t\colon t \in \mathbb{Z}\}$ is said to have innovation outliers (IO) if it satisfies (5), but the innovations $\{Z_t\}$ have a heavy-tailed distribution, for instance

$$F_\varepsilon = (1-\varepsilon)F + \varepsilon H$$

where ε is small and H is an arbitrary distribution with greater dispersion than F. The important characteristic of this kind of outliers is that even when the Z_t have outliers, (5) is satisfied and therefore $\{X_t\colon t \in \mathbb{Z}\}$ is a perfectly observed

ARMA(p, q) process. Robust estimators, like M-estimators, can typically cope with IO (Bustos and Yohai, 1986).

The process $\{X_t: t \in \mathbb{Z}\}$ is said to have additive outliers (AO) if it is not itself an ARMA(p, q) process, but rather defined by

$$X_t = V_t + B_t W_t$$

where V_t is an ARMA(p, q) process satisfying (5), B_t is a Bernoulli process with $P(B_t = 1) = \varepsilon$, $P(B_t = 0) = 1 - \varepsilon$, and W_t is an independent sequence of variables, independent of the sequences V_t and B_t. Therefore, the ARMA(p, q) process V_t is observed with probability $1 - \varepsilon$, whereas the ARMA(p, q) process V_t plus an error W_t is observed with probability ε. AO are known to be much more dangerous than IO. Note also that additive outliers have the same effect as replacement outliers (RO), where

$$X_t = (1 - B_t)V_t + B_t W_t$$

This means that the ARMA(p, q) process V_t is observed with probability $1 - \varepsilon$, and replaced by an error W_t with probability ε. In the sequel, we consider RO.

In the context of robust statistics, the breakdown point of an estimator is an important feature of reliability. It indicates how many data points need to be replaced by arbitrary values to destroy the estimator. In time series, autocovariance estimators are based on measurements taken at various time lag distances. Therefore, the classical notion of breakdown point needs to be extended to a temporal one depending on the construction of the most unfavourable configurations of perturbation. The classical notion of breakdown point of a scale estimator is given in Definition 1.

DEFINITION 1. *Let* $\mathbf{z} = (Z_1, \ldots, Z_n)^{\mathrm{T}}$ *be a sample of size* n *and let* $\tilde{\mathbf{z}}$ *be obtained by replacing any* m *observations of* $\mathbf{z}$ *by arbitrary values. The sample breakdown point of a scale estimator* $S_n(\mathbf{z})$ *is:*

$$\varepsilon_n^*(S_n(\mathbf{z})) = \max\left\{\frac{m}{n} : \sup_{\tilde{\mathbf{z}}} S_n(\tilde{\mathbf{z}}) < \infty \text{ and } \inf_{\tilde{\mathbf{z}}} S_n(\tilde{\mathbf{z}}) > 0\right\}$$

Roughly speaking, the classical breakdown point gives the maximum fraction of outliers with which the scale estimator can cope. It indicates how many data points can be replaced by arbitrary values before the scale estimator explores (tends to infinity) or implodes (tends to zero). Further discussions of this concept can be found in Hampel (1971, 1974, 1976), Huber (1981, 1984), Donoho and Huber (1983), Lucas (1997), and Genton (1998c). The sample breakdown point ε_n^* of most scale estimators is known, or can be computed. However, by using a scale estimator to compute the covariance, it is on the level of sums and differences (see (2)) that the estimator is applied. Similarly, we can define the sample breakdown point of a scale-based covariance estimator, using (2).

DEFINITION 2. *Let*

$$\mathbf{x} = (X_1, \ldots, X_n)^{\mathrm{T}} \qquad \text{and} \qquad \mathbf{y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$$

be two samples of size n. Let $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ *and let* $\tilde{\mathbf{z}}$ *be obtained by replacing any m pairs of* $\mathbf{z}$ *by arbitrary values. The sample breakdown point of a covariance estimator*

$$\hat{C}_{S_n}(\mathbf{z}) = \frac{1}{4\alpha\beta}[S_n^2(\alpha\mathbf{x} + \beta\mathbf{y}) - S_n^2(\alpha\mathbf{x} - \beta\mathbf{y})]$$

based on a scale estimator S_n *is:*

$$\varepsilon_n^*(\hat{C}_{S_n}(\mathbf{z})) = \max\left\{\frac{m}{n} : \sup_{\tilde{\mathbf{z}}} \hat{C}_{S_n}(\tilde{\mathbf{z}}) < \infty \text{ and } \inf_{\tilde{\mathbf{z}}} \hat{C}_{S_n}(\tilde{\mathbf{z}}) > -\infty \text{ and } \right.$$

$$\left. \inf_{\tilde{\mathbf{z}}} S_n(\tilde{\mathbf{z}}\cdot(\alpha \quad \beta)^{\mathrm{T}}) > 0 \text{ and } \inf_{\tilde{\mathbf{z}}} S_n(\tilde{\mathbf{z}}\cdot(\alpha \quad -\beta)^{\mathrm{T}}) > 0\right\}$$

In the particular case of time series, we typically have

$$\mathbf{u} = (X_1, \ldots, X_{n-h})^{\mathrm{T}} \qquad \text{and} \qquad \mathbf{v} = (X_{h+1}, \ldots, X_n)^{\mathrm{T}}$$

to form $\mathbf{z} = (\mathbf{u}, \mathbf{v})$. In this case, we denote $\varepsilon_{n-h}^*(\hat{C}_{S_{n-h}}(\mathbf{z}))$ by $\varepsilon_n^*(\hat{\gamma}_{S_n}(h, \mathbf{x}))$, where $\mathbf{x} = (X_1, \ldots, X_n)^{\mathrm{T}}$ (see Proposition 1). However, in time series, one is much more interested in the breakdown point related to the initial data, which are located in time. Therefore, the previous definition loses its meaning because the time location of the outlier becomes important. In fact, the effect of the perturbation of a point located close to the boundary of the time domain can be quite different from one located in the middle of the time domain, and the effect depends notably on the time lag distance h. Therefore, we introduce the following definition of a temporal sample breakdown point of an autocovariance estimator based on (2).

DEFINITION 3. *Let* $\mathbf{x} = (X_1, \ldots, X_n)^{\mathrm{T}}$ *be a sample of size n of a time series and let* $\tilde{\mathbf{x}}$ *be obtained by replacing any m observations of* $\mathbf{x}$ *by arbitrary values. Denote by* I_m *a subset of size m of* $\{1, \ldots, n\}$. *The temporal sample breakdown point of an autocovariance estimator* $\hat{\gamma}(h, \mathbf{x})$ *is:*

$$\varepsilon_n^t(\hat{\gamma}(h, \mathbf{x})) = \max\left\{\frac{m}{n} : \sup_{I_m}\sup_{\tilde{\mathbf{x}}} S_{n-h}(\tilde{\mathbf{u}} + \tilde{\mathbf{v}}) < \infty \text{ and } \inf_{I_m}\inf_{\tilde{\mathbf{x}}} S_{n-h}(\tilde{\mathbf{u}} + \tilde{\mathbf{v}}) > 0 \right.$$

$$\left. \text{and } \sup_{I_m}\sup_{\tilde{\mathbf{x}}} S_{n-h}(\tilde{\mathbf{u}} - \tilde{\mathbf{v}}) < \infty \text{ and } \inf_{I_m}\inf_{\tilde{\mathbf{x}}} S_{n-h}(\tilde{\mathbf{u}} - \tilde{\mathbf{v}}) > 0\right\}$$

where $\tilde{\mathbf{u}}$ *and* $\tilde{\mathbf{v}}$ *are derived from* $\tilde{\mathbf{x}}$ *as in (4) (t is used to emphasize temporal).*

Note that, in opposition to Definition 2, the configuration (i.e. the temporal location) of the perturbation is now taken into account, by adding the supremum

and infimum on I_m. This definition is justified by the fact that an autocovariance estimator can be destroyed by a single configuration of perturbation, indexed in I_m. Therefore, it is quite possible to find other configurations, with more than $\varepsilon_n^t(\hat{\gamma}(h, \mathbf{x}))$ of perturbations, which do not demolish the estimator. Notice, furthermore, that this definition is local, in the sense that it is valid for a fixed h.

Consider a fixed temporal lag distance $h \in \mathbb{R}$. For $m = 1$ perturbed data point, it follows that, if $h < n/2$, one perturbation at time i, with $h < i \leqslant n - h$, generates the perturbation of two sums $\mathbf{u} + \mathbf{v}$ and two differences $\mathbf{u} - \mathbf{v}$, whereas for $0 < i \leqslant h$ or $n - h < i \leqslant n$, a single sum (resp. difference) is perturbed. Finally, if $h \geqslant n/2$, one perturbation at time i, with $0 < i \leqslant n - h$ or $h < i \leqslant n$, affects one sum (resp. difference), and none in the other cases. Therefore, to one perturbed observation corresponds at most two perturbed sums (resp. differences). For general $m \geqslant 1$, we are interested in finding the most unfavourable configuration of perturbed data for a fixed h. Such a configuration is shown in Figure 1 for the case $h = 3$, $m = 7$ and $n = 21$. White points represent unperturbed observations, whereas black points represent perturbed observations. There are m black points. Construction of this configuration consists in placing h unperturbed observations, followed by h perturbed observations, followed by h unperturbed observations, and so on until exhaustion of the m black points. This configuration ensures that the most possible sums (resp. differences) are perturbed (i.e. each black point perturbs 2 sums (resp. differences)). Moreover, perturbations do not overlap for a given lag distance h, which means that no sum (resp. difference) between two perturbed observations is ever taken. Let $v_{\max}(h, m, n)$ be the maximal number of perturbed sums (resp. differences) for given h, m and n. This function depends on the relation between m and h. Let p and q be the two non-negative integers such that $m = ph + q$ and $q < h$. By disjunction of cases, it is then possible to compute the function $v_{\max}(h, m, n)$ explicitly:

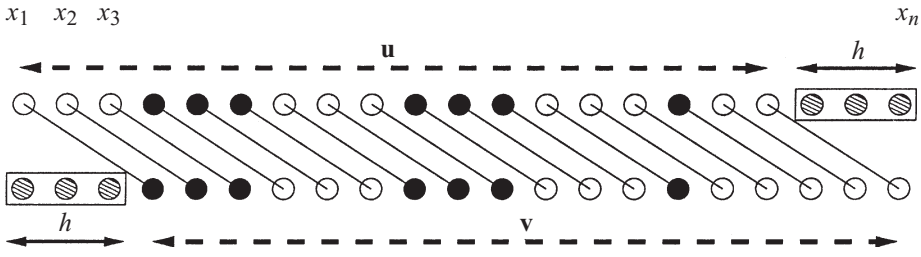

FIGURE 1. The most unfavourable configuration of perturbation for the case $h = 3$, $m = 7$ and $n = 21$; white points represent unperturbed observations, whereas black points represent perturbed observations

$$v_{\max}(h, m, n) =$$

$$\begin{cases} n-h & \text{if } m = \dfrac{n}{2} \\ & \quad \text{or } \dfrac{n}{2} > m \geqslant h,\ q = 0,\ n - 2m < h \\ & \quad \text{or } \dfrac{n}{2} > m \geqslant h,\ q \geqslant 1,\ h + q > n - 2ph \geqslant 0 \\ & \quad \text{or } m < h,\ m + 2h > n,\ m \geqslant n - h \\ 2m & \text{if } \dfrac{n}{2} > m \geqslant h,\ q = 0,\ n - 2m \geqslant h \\ & \quad \text{or } \dfrac{n}{2} > m \geqslant h,\ q \geqslant 1,\ n - 2ph \geqslant 2h + q \\ & \quad \text{or } m < h,\ m + 2h \leqslant n,\ n - 2m < h \\ n - 2h + q & \text{if } \dfrac{n}{2} > m \geqslant h,\ q \geqslant 1,\ 2h + q > n - 2ph \geqslant 2h \\ 2ph + q & \text{if } \dfrac{n}{2} > m \geqslant h,\ q \geqslant 1,\ 2h > n - 2ph \geqslant h + q \\ m + n - 2h & \text{if } m < h,\ m + 2h > n,\ m < n - h,\ h < \dfrac{n}{2} \\ m & \text{if } m < h,\ m + 2h > n,\ m < n - h,\ h \geqslant \dfrac{n}{2} \end{cases}$$

Notice that the case $m > n/2$ makes no sense because it implies that more than half of the differences are perturbed. No equivariant scale estimator can be that resistant (Huber, 1981). The following proposition examines the relation between the classical sample breakdown point (usually known) and the temporal one.

PROPOSITION 1. *For each* $h \in \{0, \ldots, n-1\}$ *and for each integer* $M = n\varepsilon_n^*(\hat{\gamma}(h, \mathbf{x})) \leqslant n/2$, *the sample breakdown point and the temporal sample breakdown point of an autocovariance estimator* $\hat{\gamma}(h, \mathbf{x})$ *satisfy the double inequality*

$$2\varepsilon_n^t(\hat{\gamma}(h, \mathbf{x})) \leqslant \varepsilon_n^*(\hat{\gamma}(h, \mathbf{x})) \leqslant \frac{2n}{n-h}\varepsilon_n^t(\hat{\gamma}(h, \mathbf{x}))$$

The first equality holds if and only if $h = n/2$ *or* $M = n/2$, *and the second equality holds if and only if* $v_{\max}(h, M, n) = 2M$.

PROOF. To prove the first inequality, consider the function

$$\delta(h, m, n) = \frac{v_{\max}(h, m, n)}{n-h} - 2\frac{m}{n}$$

We have to show that the function δ is non negative for all possible integers m. If $v_{\max}(h, m, n) = n - h$, then

$$\delta(h,\, m,\, n) = 1 - \frac{2m}{n} \geqslant 0$$

because $n/2 > m$. If $v_{\max}(h,\, m,\, n) = 2m$, then

$$\delta(h,\, m,\, n) = \frac{2m}{n-h} - \frac{2m}{n} \geqslant 0$$

because $n - h < n$. If $v_{\max}(h,\, m,\, n) = n - 2h + q$, then

$$\begin{aligned}\delta(h,\, m,\, n) &= \frac{n-2h+q}{n-h} - \frac{2m}{n} \\ &= \frac{n^2 - (p+2)hn + m(2h-n)}{n(n-h)} \\ &\geqslant \frac{n^2 - (p+2)hn + h(p+1)(2h-n)}{n(n-h)}\end{aligned}$$

because $m < h(p+1)$ and $2h - n < 0$. So,

$$\delta(h,\, m,\, n) = \frac{n - 2(p+1)h}{n} \geqslant 0$$

because $n - 2ph \geqslant 2h$. If $v_{\max}(h,\, m,\, n) = 2ph + q$, then

$$\begin{aligned}\delta(h,\, m,\, n) &= \frac{2ph+q}{n-h} - \frac{2m}{n} \\ &= \frac{2mh - nq}{n(n-h)} \\ &\geqslant \frac{2mh - 2hq(p+1)}{n(n-h)}\end{aligned}$$

because $n - 2h(p+1) < 0$. So

$$\delta(h,\, m,\, n) = \frac{2hp(h-q)}{n(n-h)} \geqslant 0$$

because $h > q$. If $v_{\max}(h,\, m,\, n) = m + n - 2h$, then

$$\begin{aligned}\delta(h,\, m,\, n) &= \frac{m+n-2h}{n-h} - \frac{2m}{n} \\ &= \frac{m(2h-n) + n(n-2h)}{n(n-h)} \\ &\geqslant \frac{-h(n-2h) + n(n-2h)}{n(n-h)}\end{aligned}$$

because $m < h$. So,

$$\delta(h, m, n) = \frac{n - 2h}{n} \geqslant 0$$

because $h < n/2$. If $\upsilon_{\max}(h, m, n) = m$, then

$$\delta(h, m, n) = \frac{m}{n - h} - \frac{2m}{n}$$

$$= \frac{m(2h - n)}{n(n - h)} \geqslant 0$$

because $h < n/2$. Finally, if $h = n/2$ or $m = n/2$, then $\delta(h, m, n) = 0$ and therefore equality is reached. The second inequality follows from the fact that a perturbation on a single observation generates the perturbation of at most two sums (resp. differences). Thus, the perturbation of m observations generates the perturbation of at most $2m$ sums (resp. differences), i.e. $\upsilon_{\max}(h, m, n) \leqslant 2m$. Consequently, we have the inequality

$$\varepsilon_n^*(\hat{\gamma}(h, \mathbf{x})) = \frac{\upsilon_{\max}(h, M, n)}{n - h} \leqslant \frac{2M}{n - h} = \frac{2n}{n - h} \frac{M}{n} = \frac{2n}{n - h} \varepsilon_n^t(\hat{\gamma}(h, \mathbf{x}))$$

with equality if and only if $\upsilon_{\max}(h, M, n) = 2M$. ■

REMARK 1. *By writing the inequality in Proposition 1 slightly differently, we can bound the temporal sample breakdown point with the classical sample breakdown point:*

$$\frac{n - h}{2n} \varepsilon_n^*(\hat{\gamma}(h, \mathbf{x})) \leqslant \varepsilon_n^t(\hat{\gamma}(h, \mathbf{x})) \leqslant \frac{1}{2} \varepsilon_n^*(\hat{\gamma}(h, \mathbf{x}))$$

The classical sample autocovariance function is based on the classical scale estimator (standard deviation) whose sample breakdown point is zero. Therefore, by Remark 1, the temporal sample breakdown point of this estimator is also zero, for every lag h. This means that a single outlier in the data can destroy it. Figure 2 shows the temporal sample breakdown point $\varepsilon_{100}^t(\hat{\gamma}_Q(h, \mathbf{x}))$ of the highly robust autocovariance estimator, for each lag distance h, represented by the black curve. The upper and lower bounds given in Remark 1 are represented by the light grey curves. As it was stated, the temporal sample breakdown point equals its lower bound as long as $\upsilon_{\max}(h, M, n) = 2M$, and equals its upper bound if $h = n/2$. The interpretation of this figure is as follows. For a fixed h, if the percentage of perturbed observations is below the black curve, the estimator is never destroyed. If the percentage is above the black curve, there exists at least one configuration which destroys the estimator. This implies that the highly robust autocovariance estimator is more resistant at small time lags h or around $h = n/2$, than at large time lags h or before $h = n/2$, according to Figure 2. In practice, however, lags larger than $n/2$ are seldom used. Note that from Remark 1,

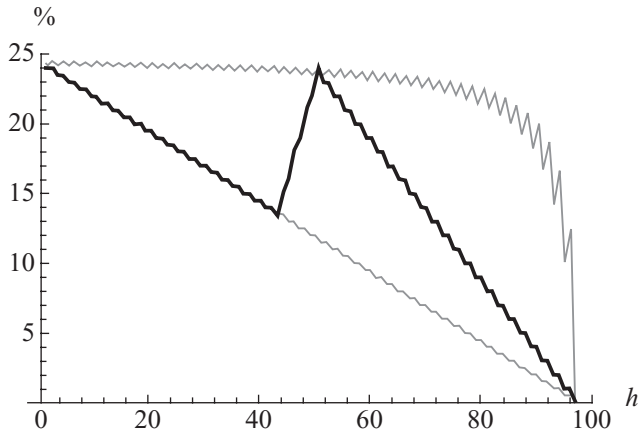

FIGURE 2. The temporal sample breakdown point (in black) as a function of the temporal lag distance h, for the highly robust sample autocovariance estimator $\hat{\gamma}_{\mathrm{Q}}(h, \mathbf{x})$; the upper and lower bounds are drawn in light grey

asymptotically, the temporal breakdown point of the autocovariance estimator is half the classical breakdown point.

Recall that Q_n has classical asymptotic breakdown point 50%, which means we can contaminate half of the observations yet still have a reasonable estimate. Therefore, the highly robust autocovariance estimator has breakdown point 25%. This is because in forming $\mathbf{u}$ and $\mathbf{v}$ from the observation $\mathbf{x}$, most data will appear twice and hence, in the worst case, the number of pairs $(\mathbf{u}_i, \mathbf{v}_i)$ that contain outliers will be twice the number of original outliers in $\mathbf{x}$. Note that this is the highest possible breakdown point for an autocovariance estimator. We can give up such high breakdown point by choosing a different quantile from $\frac{1}{4}$ in (3), with the benefit of higher efficiency (Rousseeuw and Croux, 1992). For example, if we choose the 0.91 quantile, we will reach the highest efficiency ($\approx 99\%$) for Q_n estimator, hence reach the highest efficiency for our estimator too. For the 0.91 quantile, the classical breakdown point is approximately 4.6% for Q_n and therefore the temporal breakdown point of $\hat{\gamma}_{\mathrm{Q}}$ is 2.3%.

4. INFLUENCE FUNCTION AND ASYMPTOTIC VARIANCE

Consider an estimator $T_n(X_1, \ldots, X_n) = T_n(F_n)$, where

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \Delta_{X_i}(x)$$

is the empirical distribution, and let $T(F)$ be the corresponding statistical

functional such that $T(F_n) = T_n(F_n)$. The influence function (Hampel, 1974) is an important tool to describe the robustness properties of an estimator. For a statistical functional T at a distribution F, the influence function IF is defined by:

$$IF(x,\ T,\ F) = \lim_{\varepsilon \to 0^+} \frac{T((1-\varepsilon)F + \varepsilon\Delta_x) - T(F)}{\varepsilon} \tag{6}$$

for x such that this limit exists. The importance of the influence function lies in its heuristic interpretation: it describes the effect of an infinitesimal contamination at the point x on the estimate, standardized by the mass of the contamination, i.e. it measures the asymptotic bias caused by the contamination in the observations. When $T = S$, a statistical functional of scale, there are some useful equalities on the influence function.

PROPOSITION 2. *Suppose* $F = \mathrm{N}(0, \sigma^2)$, $\Phi = \mathrm{N}(0, 1)$. *Let* S *be a statistical functional of scale, and hence* S^2 *be a statistical functional of variance. More generally, consider a functional of* S *denoted by* $h(S)$. *Then*

$$\begin{aligned} IF(x,\ S,\ F) &= \sigma\, IF\left(\frac{x}{\sigma},\ S,\ \Phi\right) \\ IF(x,\ S^2,\ F) &= \sigma^2\, IF\left(\frac{x}{\sigma},\ S^2,\ \Phi\right) \\ IF(x,\ h(S),\ F) &= h'(S(F))IF(x,\ S,\ F) \end{aligned} \tag{7}$$

Moreover, the following equalities on the asymptotic variance hold for independent observations:

$$\begin{aligned} \mathrm{var}(S,\ F) &= \sigma^2\, \mathrm{var}(\mathrm{S},\ \Phi) \\ \mathrm{var}(S^2,\ F) &= \sigma^4\, \mathrm{var}(\mathrm{S}^2,\ \Phi) \end{aligned} \tag{8}$$

PROOF. We prove the first equation in detail and briefly explain the latter ones.

$$\begin{aligned} IF(x,\ S,\ F) &= \frac{\partial}{\partial\varepsilon} S((1-\varepsilon)F(u) + \varepsilon\Delta_x(u))|_{\varepsilon=0} \\ &= \frac{\partial}{\partial\varepsilon} S\left((1-\varepsilon)\Phi\left(\frac{u}{\sigma}\right) + \varepsilon\Delta_{x/\sigma}\left(\frac{u}{\sigma}\right)\right)\Bigg|_{\varepsilon=0} \end{aligned}$$

We use $P(u/\sigma)$ to denote the function

$$(1-\varepsilon)\Phi\left(\frac{u}{\sigma}\right) + \varepsilon\Delta_{x/\sigma}\left(\frac{u}{\sigma}\right)$$

and by the property of equivariance of the scale estimator, we know that

$$S\left(P\left(\frac{u}{\sigma}\right)\right) = \sigma S(P(u))$$

So, we get

$$IF(x, S, F) = \sigma \frac{\partial}{\partial \varepsilon}(S((1-\varepsilon)\Phi(u) + \varepsilon \Delta_{x/\sigma}(u)))|_{\varepsilon=0}$$

$$= \sigma\, IF\left(\frac{x}{\sigma}, S, \Phi\right)$$

The second equality in (7) can be proved similarly as the first one. The third equality is obvious. To prove the equalities in (8), we use the formula

$$\text{var}(S, F) = \int |IF(x, S, F)|^2 \, \mathrm{d}F(x) \quad \text{and} \quad \text{var}(S^2, F) = \int |IF(x, S^2, F)|^2 \, \mathrm{d}F(x)$$

and the equalities in (7). Note that this result is valid only when the estimator is carried on independent observations. ■

For a covariance estimator based on (2), with X and Y replaced by U and V, for given α and β, the influence function is a function from $\mathbb{R}^2$ to $\mathbb{R}$, defined by Genton and Ma (1999):

$$IF((u, v), T, F)$$
$$= \frac{1}{2\alpha\beta}[S(F_1)IF((\alpha u + \beta v), S, F_1) - S(F_2)IF((\alpha u - \beta v), S, F_2)] \quad (9)$$

Here T is estimating γ, S is a scale estimator, with influence function $IF(\cdot, S, F_i)$, $i = 1, 2$, and we assume the following distributions: $\alpha U + \beta V \sim F_1$, $\alpha U - \beta V \sim F_2$, and F is the joint distribution of U and V. In the autocovariance case, we have $\alpha = \beta = 1$. Let γ_Q be the statistical functional corresponding to the highly robust autocovariance estimator $\hat{\gamma}_Q$. Under a bivariate Gaussian distribution F, the influence function of the γ_Q autocovariance estimator is:

$$IF((u, v), \gamma_Q, F) = \frac{1}{2}\left[\sigma^2_{U+V} IF\left(\frac{u+v}{\sigma_{U+V}}, Q, \Phi\right) - \sigma^2_{U-V} IF\left(\frac{u-v}{\sigma_{U-V}}, Q, \Phi\right)\right] \quad (10)$$

where the influence function of Q_n at Φ is (Rousseeuw and Croux, 1993):

$$IF(x, Q, \Phi) = c\frac{\frac{1}{4} - \Phi\left(x + \frac{1}{c}\right) + \Phi\left(x - \frac{1}{c}\right)}{\int \phi\left(y + \frac{1}{c}\right)\phi(y)\, \mathrm{d}y} \quad (11)$$

with $c = 2.2191$ and ϕ is the standard normal density. Figure 3 shows the plot of the influence function of γ_Q when the covariance is zero. The cases of non-zero covariance yield similar graphs. Note that the influence function of γ_Q is bounded between

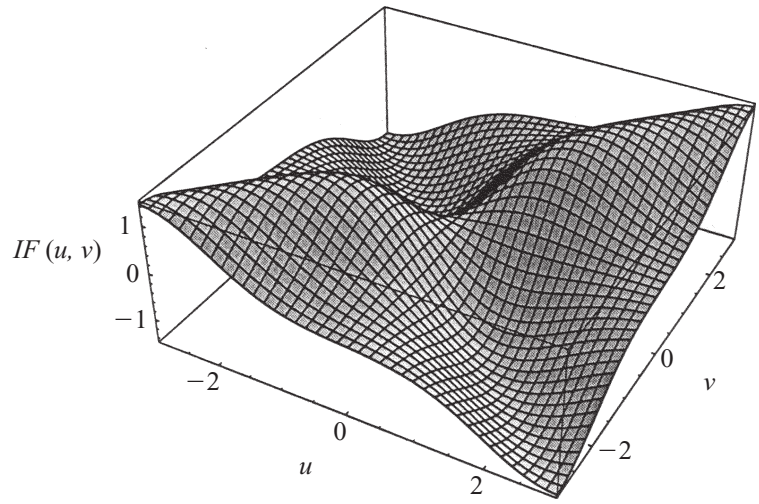

FIGURE 3. The influence function $IF((u, v), \gamma_Q, F)$, where U and V are independent and have identical standard Gaussian distribution

$$\pm\frac{[(a-b)(\sigma_U^2+\sigma_V^2)+2|a+b|\sigma_U\sigma_V]}{2}$$

where

$$a=\max_x IF(x, Q, \Phi) \qquad \text{and} \qquad b=\min_x IF(x, Q, \Phi)$$

The bounds can be computed by writing $\sigma_{U\pm V}^2$ as $\sigma_U^2+\sigma_V^2\pm 2\,\mathrm{cov}(U, V)$ and noticing that $\mathrm{cov}(U, V)$ is bounded between $\pm\sigma_U\sigma_V$. To the contrary, the influence function of γ_M is proportional to uv when the covariance is zero, and therefore unbounded.

Under regularity conditions, $\hat{\gamma}_Q$ is consistent, i.e. $\hat{\gamma}_Q \to \gamma_Q$ in probability as $n\to\infty$, since Q_n is consistent (Rousseeuw and Croux, 1993). Moreover, $\sqrt{n}(\hat{\gamma}_Q-\gamma_Q)$ is asymptotically normal with zero expectation and variance given by (Portnoy, 1977, 1979; Genton, 1998b)

$$\begin{aligned}\mathrm{var}(\gamma_Q, F) &= \iint |IF((u, v), \gamma_Q, F)|^2\,\mathrm{d}F(u, v)\\ &\quad + 2\sum_{k=1}^{\infty}\int\cdots\int IF((u_1, v_1), \gamma_Q, F)\\ &\quad \times IF((u_{1+k}, v_{1+k}), \gamma_Q, F)\,\mathrm{d}F((u_1, v_1), (u_{1+k}, v_{1+k})) \qquad (12)\end{aligned}$$

Regularity conditions for consistency and asymptotic normality are given by Huber (1967) for the independent case and by Portnoy (1977, 1979) and Bustos (1982) for the dependent case. In this latter situation, mixing conditions like α-

mixing or ϕ-mixing are sufficient (Billingsley, 1968; Doukhan, 1994). Equation (12) is valid for any consistent estimator of autocovariance. In particular, for the classical autocovariance estimator $\hat{\gamma}_M$, the equation is equivalent to Bartlett's formula (Priestley, 1981, p. 326). The computation of the asymptotic variance turns out to be tedious for $\hat{\gamma}_Q$ and a closed form is impossible. However, Equation (12) can be used for numerical calculation of the asymptotic variance of $\hat{\gamma}_Q$.

5. SIMULATIONS

In this section, we present some simulations so that we can compare the $\hat{\gamma}_M$ and $\hat{\gamma}_Q$ autocovariance estimator on MA(1) = ARMA(0, 1) and AR(1) = ARMA(1, 0) models, with and without replacement outliers. We start with a brief description of the experiment.

The standard Gaussian MA(1) and AR(1) models $\{V_t\}$ are considered, with or without replacement outliers (RO) defined by

$$X_t = (1 - B_t)V_t + B_t W_t$$

The Bernoulli process satisfies

$$P(B_t = 1) = \varepsilon \qquad \text{and} \qquad P(B_t = 0) = 1 - \varepsilon$$

with $\varepsilon = 0$ and $\varepsilon = 10\%$. The distribution of W_t is chosen to be $N(0, \tau^2)$, where $\tau^2 = k^2 \operatorname{var}(V_t)$ with $k = 3$ and $k = 10$. We generate 1000 samples of sizes 20, 50 and 100 for each model with parameters θ (resp. ρ) equal to 0 and 0.5. The mean of $\hat{\gamma}_M$ and $\hat{\gamma}_Q$ are computed over the 1000 replications, as well as the relative efficiency (REF) of $\hat{\gamma}_Q$ to $\hat{\gamma}_M$, i.e. the ratio of their sample variance. We built an S-Plus function to compute $\hat{\gamma}_Q$, which is available on the Web.

The results are presented in Tables 1, 2, and 3. From the simulation, we can see that when there are no outliers, both estimators yield a mean that is close to the true autocovariance, i.e. are unbiased. The REF is around 80% for large n. This is considered high for a highly robust autocovariance estimator. In the presence of outliers, the classical autocovariance estimator shows a weak resistance in terms of the mean value, and it also has smaller efficiency than the robust estimator. This is particularly clear when the outliers are large ($k = 10$). One can also check that the asymptotic variances of $\hat{\gamma}_M$ given by (12) agree with the ones found in the simulations. Moreover, for the MA(1) model with $\theta = 0$, i.e. the i.i.d. case, the asymptotic variance of $\hat{\gamma}_Q$ can be computed numerically from (12), which yields 2.482 (to be compared with 2 for $\hat{\gamma}_M$). This yields an asymptotic relative efficiency of 80.6%, which is close to the one found by simulation in Table 1.

Note that the classical estimator we took is not modified to ensure positive definiteness of the sample autocovariance matrix. If we use the modified version (divided by n instead of $n - h$), then we should also ensure positive definiteness of the autocovariance matrix obtained with the highly robust

TABLE I

THE MEAN $m(\hat{\gamma}_M)$, $m(\hat{\gamma}_Q)$ AND THE RELATIVE EFFICIENCY REF OF THE AUTOCOVARIANCE ESTIMATORS AT TIME LAG $h = 0$, $h = 1$ AND $h = 2$ ON MA(1) MODEL WITH $\theta = 0$, WITH AND WITHOUT RO OUTLIERS

	$\varepsilon = 0$			$\varepsilon = 10\%,\ k = 3$			$\varepsilon = 10\%,\ k = 10$		
	$m(\hat{\gamma}_M)$	$m(\hat{\gamma}_Q)$	REF	$m(\hat{\gamma}_M)$	$m(\hat{\gamma}_Q)$	REF	$m(\hat{\gamma}_M)$	$m(\hat{\gamma}_Q)$	REF
$\gamma(0) = 1$									
$n = 20$	0.948	1.035	0.566	1.722	1.356	2.913	10.531	1.597	187.1
$n = 50$	0.983	1.012	0.721	1.784	1.290	4.015	10.842	1.514	262.2
$n = 100$	0.983	1.000	0.762	1.766	1.268	4.035	10.612	1.467	277.0
$\gamma(1) = 0$									
$n = 20$	−0.043	−0.056	0.487	−0.098	−0.121	0.743	−0.519	−0.243	9.2
$n = 50$	−0.020	−0.027	0.712	−0.030	−0.026	1.006	−0.231	−0.048	12.9
$n = 100$	−0.005	−0.010	0.728	−0.018	−0.028	1.100	−0.093	−0.044	14.9
$\gamma(2) = 0$									
$n = 20$	−0.054	−0.067	0.497	−0.085	−0.107	0.713	−0.497	−0.237	4.7
$n = 50$	−0.019	−0.019	0.680	−0.049	−0.061	1.039	−0.271	−0.100	14.6
$n = 100$	−0.018	−0.020	0.741	−0.023	−0.021	1.043	−0.142	−0.034	13.3

TABLE II

THE MEAN $m(\hat{\gamma}_M)$, $m(\hat{\gamma}_Q)$ AND THE RELATIVE EFFICIENCY REF OF THE AUTOCOVARIANCE ESTIMATORS AT TIME LAG $h = 0$, $h = 1$ AND $h = 2$ ON MA(1) MODEL WITH $\theta = 0.5$, WITH AND WITHOUT RO OUTLIERS

	$\varepsilon = 0$			$\varepsilon = 10\%,\ k = 3$			$\varepsilon = 10\%,\ k = 10$		
	$m(\hat{\gamma}_M)$	$m(\hat{\gamma}_Q)$	REF	$m(\hat{\gamma}_M)$	$m(\hat{\gamma}_Q)$	REF	$m(\hat{\gamma}_M)$	$m(\hat{\gamma}_Q)$	REF
$\gamma(0) = 1.25$									
$n = 20$	1.134	1.252	0.621	2.120	1.637	2.768	12.700	1.924	167.0
$n = 50$	1.214	1.254	0.754	2.156	1.589	3.135	13.075	1.855	225.0
$n = 100$	1.228	1.244	0.801	2.227	1.575	3.787	13.798	1.838	226.1
$\gamma(1) = 0.5$									
$n = 20$	0.380	0.434	0.509	0.271	0.466	0.795	−0.278	0.780	7.9
$n = 50$	0.464	0.481	0.704	0.338	0.515	1.025	0.106	0.882	10.2
$n = 100$	0.481	0.487	0.800	0.380	0.535	1.061	0.300	0.888	10.8
$\gamma(2) = 0$									
$n = 20$	−0.110	−0.137	0.536	−0.156	−0.196	0.696	−0.839	−0.445	5.0
$n = 50$	−0.046	−0.049	0.711	−0.068	−0.087	0.930	−0.331	−0.115	12.9
$n = 100$	−0.016	−0.018	0.809	−0.027	−0.037	1.174	−0.147	−0.041	19.1

TABLE III

THE MEAN $m(\hat{\gamma}_M)$, $m(\hat{\gamma}_Q)$ AND THE RELATIVE EFFICIENCY REF OF THE AUTOCOVARIANCE ESTIMATORS AT TIME LAG $h = 0$, $h = 1$ AND $h = 2$ ON AR(1) MODEL WITH $\theta = 0.5$, WITH AND WITHOUT RO OUTLIERS

	$\varepsilon = 0$			$\varepsilon = 10\%,\ k = 3$			$\varepsilon = 10\%,\ k = 10$		
	$m(\hat{\gamma}_M)$	$m(\hat{\gamma}_Q)$	REF	$m(\hat{\gamma}_M)$	$m(\hat{\gamma}_Q)$	REF	$m(\hat{\gamma}_M)$	$m(\hat{\gamma}_Q)$	REF
$\gamma(0) = 1.333$									
$n = 20$	1.158	1.279	0.668	2.143	1.664	2.9	14.739	2.007	181.6
$n = 50$	1.252	1.293	0.772	2.292	1.647	3.5	14.433	1.933	199.0
$n = 100$	1.291	1.313	0.803	2.378	1.669	3.3	14.803	1.940	235.1
$\gamma(1) = 0.667$									
$n = 20$	0.486	0.550	0.613	0.328	0.568	0.9	−0.321	1.094	3.9
$n = 50$	0.586	0.609	0.749	0.434	0.668	1.0	0.271	1.126	7.2
$n = 100$	0.628	0.638	0.804	0.501	0.724	1.0	0.404	1.159	7.6
$\gamma(2) = 0.333$									
$n = 20$	0.145	0.172	0.534	0.033	0.121	0.7	−0.499	0.256	5.1
$n = 50$	0.246	0.254	0.763	0.180	0.258	1.0	0.082	0.497	9.6
$n = 100$	0.291	0.296	0.778	0.227	0.334	1.0	0.093	0.543	12.0

autocovariance estimator. This can be done by the shrinking, the eigenvalue or the scaling method (Rousseeuw and Molenberghs, 1993). A typical application is then to use the highly robust autocovariance estimator in the Yule–Walker equations (Brockwell and Davis, 1991) to estimate the parameters of an AR model robustly.

6. EXAMPLE

We compute the classical autocovariance estimator $\hat{\gamma}_M$ and the highly robust autocovariance estimator $\hat{\gamma}_Q$ on 91 monthly interest rates of an Austrian bank; see Figure 4 for the data. This data set has already been analysed by Künsch (1983, 1984). He pointed out the presence of three outliers for the months number 18, 28, 29. In Figure 5, we run $\hat{\gamma}_M$ and $\hat{\gamma}_Q$ on the original data in (a) and (b). Then we replace the three outliers by 9.85 as suggested by Künsch in (c) and (d). Looking at (c) and (d), we can see that the new estimator $\hat{\gamma}_Q$ behaves similarly to $\hat{\gamma}_M$ when no outliers are present. Comparing the difference between (a) and (c) with the difference between (b) and (d), we can see that $\hat{\gamma}_Q$ has better resistance to the outliers than $\hat{\gamma}_M$. This effect is particularly visible for small time lags.

7. CONCLUSION

In this paper, the problem of the robustness of the sample autocovariance function has been addressed, and a new autocovariance estimator, based on a

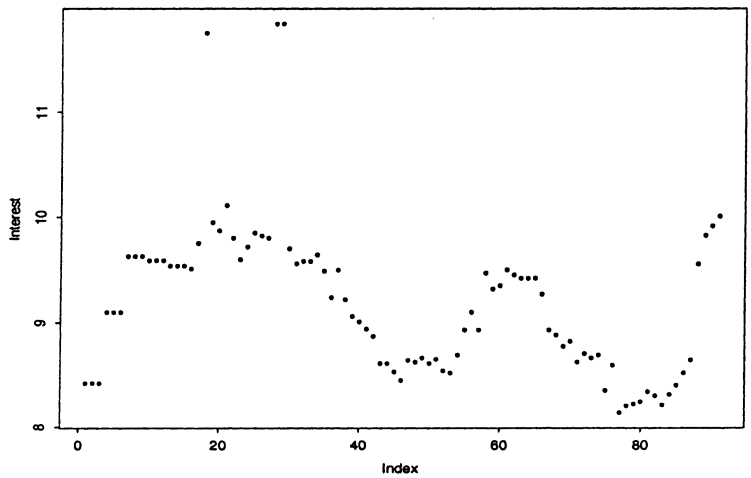

FIGURE 4. Monthly interest rates of an Austrian bank during 91 months

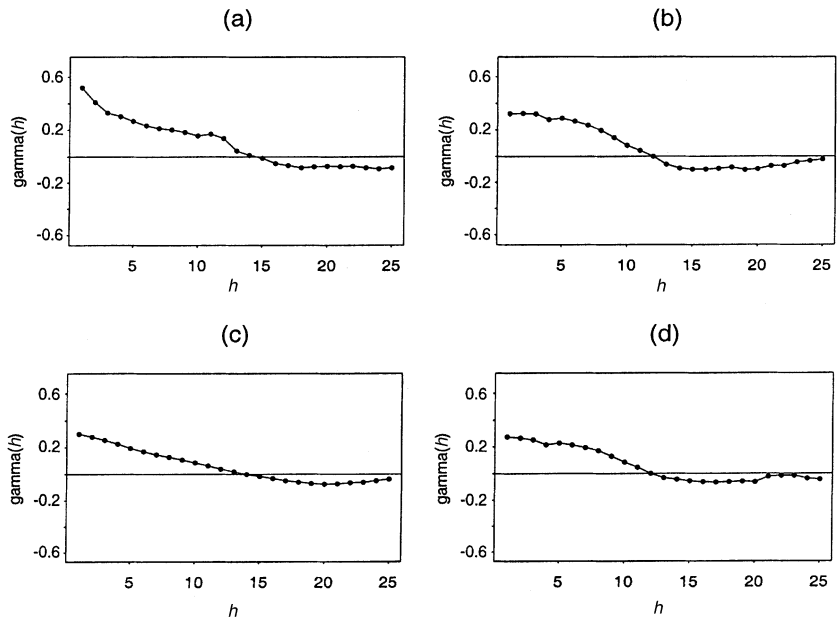

FIGURE 5. Autocovariance estimator on monthly interest rates of an Austrian bank: (a) classical $\hat{\gamma}_M$ on original data; (b) highly robust $\hat{\gamma}_Q$ on original data; (c) classical $\hat{\gamma}_M$ on corrected data; (d) highly robust $\hat{\gamma}_Q$ on corrected data

highly robust estimator of scale Q_n, has been proposed. Its robustness properties were studied by means of the influence function, and a concept of temporal breakdown point. A simulation study has been carried out, showing the behaviour of both the classical and highly robust autocovariance estimator in the presence of outliers. The poor resistance of the classical estimator has been exhibited, in particular with respect to bias and efficiency. Therefore, instead of using only the highly robust autocovariance estimator in practice, we rather suggest computing it along with the classical one. If they are very close to each other, one can assume that outliers had negligible effect. If they are significantly different, one has to think and act with care. For instance, in the time series of monthly interest rates of an Austrian bank, outliers were suspected due to the different results between the two estimators. An S-Plus function for the highly robust autocovariance estimator is available on the Web at http://www-math.mit.edu/~yanyuan/Genton/Time/time.html.

ACKNOWLEDGMENTS

The authors are grateful to Xavier de Luna and two anonymous referees for their comments on an earlier version of the paper. We would like also to thank Cédlic and Chen who were our inspiration for this research.

REFERENCES

BILLINGSLEY, P. (1968) *Convergence of Probability Measures*. New-York: Wiley.

BOX, G. E. P. and JENKINS, G. M. (1976) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden Day.

BROCKWELL, P. J. and DAVIS, R. A. (1991) *Time Series: Theory and Methods*. New York: Springer.

BUSTOS, O. H. (1982) General M-estimates for contaminated p-th order autoregressive processes: Consistency and asymptotic normality. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 59, 491–504.

BUSTOS, O. H. and YOHAI, V. J. (1986) Robust estimates for ARMA models. *J. Am. Stat. Assoc.* 81, 155–68.

CROUX, C. and ROUSSEEUW, P. J. (1992) Time-efficient algorithms for two highly robust estimators of scale. *Computational Statistics* 2, 411–28.

CROUX, C., ROUSSEEUW, P. J. and HÖSSJER, O. (1994) Generalized *S*-estimators. *J. Am. Stat. Assoc.* 89, 1271–81.

DENBY, J. and MARTIN, R. D. (1979) Robust estimation of the first-order autoregressive parameter. *J. Am. Stat. Assoc.* 74, 140–6.

DONOHO, D. L. and HUBER, P. J. (1983) The notion of breakdown point. In *A Festschrift for Erich L. Lehmann*: (eds P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr). Wadsworth, Belmont: Wadsworth 157–84.

DOUKHAN, P. (1994) *Mixing: Properties and Examples*. New York: Springer.

GENTON, M. G. (1998a) Highly robust variogram estimation. *Math. Geology* 30, 213–21.

GENTON, M. G. (1998b) Asymptotic variance of M-estimators for dependent Gaussian random variables. *Statistics and Probability Letters* 38, 255–61.

GENTON, M. G. (1998c) Spatial breakdown point of variogram estimators. *Math. Geology* 30, 853–71.

GENTON, M. G. and MA, Y. (1999) Robustness properties of dispersion estimators. *Statistics and Probability Letters* 44, 343–50.

GNANADESIKAN, R. (1997) *Methods for Statistical Data Analysis of Multivariate Observations*. New York: Wiley.

GNANADESIKAN, R. and KETTENRING, J. R. (1972) Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* 28, 81–124.

HALLIN, M. and PURI, M. L. (1994) Aligned rank tests for linear models with autocorrelated error terms. *Journal of Multivariate Analysis* 50, 175–237.

HAMPEL, F. R. (1971) A general qualitative definition of robustness. *Ann. Math. Stat.* 42 1887–96.

HAMPEL, F. R. (1973) Robust estimation, a condensed partial survey. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 27, 87–104.

HAMPEL, F. R. (1974) The influence curve and its role in robust estimation. *Jour. Am. Stat. Assoc.* 69, 383–93.

HAMPEL, F. R. (1976) *On the Breakdown Points of Some Rejection Rules with Mean*. Research Report 11, Fachgruppe für Statistik, ETHZ.

HÖSSJER, O., CROUX, C. and ROUSSEEUW, P. J. (1994) Asymptotics of generalized S-estimators. *J. Multivariate Analysis* 51, 148–77.

HUBER, P. J. (1967) The behavior of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1, 221–33.

HUBER, P. J. (1977) *Robust Statistical Procedures*. Philadelphia: Society for Industrial and Applied Mathematics.

HUBER, P. J. (1981) *Robust Statistics*. New York: Wiley.

HUBER, P. J. (1984) Finite sample breakdown of M- and P-estimators. *Ann. Math. Stat.* 12, 119–26.

KÜNSCH, H. (1983) Robust estimation for autogressive processes. *Proc. Institute for Statist. Math.* 31, 51–64.

KÜNSCH, H. (1984) Infinitesimal robustness for autoregressive processes. *Ann. Stat.* 12, 843–63.

LUCAS, A. (1997) Asymptotic robustness of least median of squares for autoregressions with additive outliers. *Commun. Statist.-Theory Meth.* 26, 2363–80.

MARTIN, R. D. and YOHAI, V. J. (1985) Robustness in time series and estimating ARMA models. In *Handbook of Statistics* 5 (eds E. J. Hannan, P. R. Krishnaiah and M. M. Rao), 119–55.

PORTNOY, S. (1977) Robust estimation in dependent situations. *Ann. Stat.* 5, 22–43.

PORTNOY, S. (1979) Further remarks on robust estimation in dependent situations. *Ann. Stat.* 7, 224–31.

PRIESTLEY, M. B. (1981) *Spectral Analysis and Time Series. Vol. 1 and 2*. London: Academic Press.

ROUSSEEUW, P. J. and CROUX, C. (1992) Explicit scale estimators with high breakdown point. *L_1 Statistical Analyses and Related Methods*, 77–92.

ROUSSEEUW, P. J. and CROUX, C. (1993) Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* 88, 1273–83.

ROUSSEEUW, P. J. and MOLENBERGHS, G. (1993) Transformation of non positive semidefinite correlation matrices. *Commun. Statist.-Theory Meth.* 22, 965–84.