Aust. N. Z. J. Stat. **43**(4), 2001, 399–419

WEIGHTED WILCOXON ESTIMATES FOR AUTOREGRESSION

JEFFREY T. TERPSTRA[1*], JOSEPH W. MCKEAN[2] AND JOSHUA D. NARANJO[2]

North Dakota State University and Western Michigan University

Summary

This paper explores the class of weighted Wilcoxon (WW) estimates in the context of autoregressive parameter estimation, giving special attention to three sub-classes of so-called WW-estimates. When the weights are constant, the estimate is equivalent to using Jaeckel's estimate with Wilcoxon scores. The paper presents asymptotic linearity properties for the three sub-classes of WW-estimates. These properties imply that the estimates are asymptotically normal at rate $n^{1/2}$. Tests of hypotheses as well as standard errors for confidence interval procedures can be based on such results. Furthermore, the estimates can be computed with an L_1 regression routine once the weights have been calculated. Examples and a Monte Carlo study over innovation and additive outlier models suggest that WW-estimates can be both robust and highly efficient.

Key words: additive outlier model; autoregressive time series; GR-estimates; innovative outlier model; rank-based estimates; robust; weighted Wilcoxon estimates.

1. Introduction

The stationary autoregressive model of order p, denoted here by AR(p), is widely used in time series analysis. The model (with location parameter) is typically written as

$$X_i = \phi_0 + \phi_1 X_{i-1} + \phi_2 X_{i-2} + \cdots + \phi_p X_{i-p} + \varepsilon_i = \phi_0 + \boldsymbol{Y}_{i-1}^{\mathrm{T}} \boldsymbol{\phi} + \varepsilon_i \quad (i = 1, 2, \ldots, n), \tag{1}$$

where $p \geq 1$, $\boldsymbol{Y}_{i-1} = (X_{i-1}, X_{i-2}, \ldots, X_{i-p})$, $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)$ and $\boldsymbol{Y}_0$ is an observable random vector independent of $\{\varepsilon_i\}$. The stationarity assumption requires that the solutions to the following equation,

$$x^p - \phi_1 x^{p-1} - \phi_2 x^{p-2} - \cdots - \phi_p = 0, \tag{2}$$

lie inside the unit circle. Furthermore, the ε_i are typically assumed independent and identically distributed (iid) according to a continuous distribution function F that satisfies

$$\mathrm{E}(\varepsilon_1) = 0 \qquad \text{and} \qquad \mathrm{E}(\varepsilon_1^2) = \sigma^2 < \infty. \tag{3}$$

Recall that (1)–(3) guarantee that the process $\{X_i\}$ is both causal and invertible (see e.g. Brockwell & Davis, 1991 Section 3.1). This, along with the continuity of F, implies that

Received April 2000; revised October 2000; accepted October 2000.
* Author to whom correspondence should be addressed.
[1] Dept of Statistics, North Dakota State University, PO Box 5575 Waldron 201H, Fargo, ND 58105, USA. e-mail: Jeff.Terpstra@ndsu.nodak.edu
[2] Dept of Statistics, Western Michigan University, 3304 Everett Tower, Kalamazoo, MI 49008, USA.
Acknowledgments. The authors thank the referee for providing helpful comments on the initial version of this paper. We also thank Dr Bernard Garel and Dr Marc Hallin for sharing the US Exports dataset with us.

the various inverses appearing in the sequel exist with probability one. The above assumptions also imply that the process is both ergodic and geometrically absolutely regular (see e.g. Hannan, 1970 Theorem 3 p. 204 and Doukhan, 1994 Section 2.4.1, respectively). All these properties play a critical role in the development of Section 4.

Over the years considerable progress has been made regarding robust estimation of the AR parameters in (1). For accounts of M-estimates and GM-estimates see Denby & Martin (1979), Bustos (1982), Kunsch (1984a) and Martin & Yohai (1991). R-estimates and a generalization of R-estimates are discussed by Koul & Saleh (1993) and Koul & Ossiander (1994), respectively. Furthermore, Ferretti, Kelmansky & Yohai (1991) introduce RAR-estimates, which are also rank-based in nature. Lastly, Rousseeuw & Leroy (1987 Section 7.2) discuss an example where re-weighted least squares and least median squares are used to obtain estimates of $\boldsymbol{\phi}$.

In this paper the proposed estimate of $\boldsymbol{\phi}$, say $\hat{\boldsymbol{\phi}}_n$, is a value of $\boldsymbol{\phi}$ that minimizes the dispersion function

$$D(\boldsymbol{\phi}) = \sum_{1 \le i < j \le n} b_{ij} |\varepsilon_j(\boldsymbol{\phi}) - \varepsilon_i(\boldsymbol{\phi})|, \tag{4}$$

where b_{ij} denotes a weight to be used in the (i, j)th comparison and $\varepsilon_i(\boldsymbol{\phi}) = X_i - \boldsymbol{Y}_{i-1}^{\mathsf{T}}\boldsymbol{\phi}$. If the weights are assumed non-negative it can be shown that $D(\boldsymbol{\phi})$ is non-negative, piecewise linear and convex. Hence, a minimum of $D(\boldsymbol{\phi})$ is guaranteed. Although this minimum is not necessarily unique, it turns out that the diameter of the set of solutions is $o_p(n^{-1/2})$. Alternatively, the estimate of $\boldsymbol{\phi}$ can be viewed as an approximate solution of the equation $\boldsymbol{S}(\boldsymbol{\phi}) = -\nabla D(\boldsymbol{\phi}) = \boldsymbol{0}$, where

$$\boldsymbol{S}(\boldsymbol{\phi}) = \sum_{1 \le i < j \le n} b_{ij} (\boldsymbol{Y}_{j-1} - \boldsymbol{Y}_{i-1}) \operatorname{sgn}\big(\varepsilon_j(\boldsymbol{\phi}) - \varepsilon_i(\boldsymbol{\phi})\big)$$

except at those points where the gradient does not exist. The solution is approximate because $\boldsymbol{S}(\boldsymbol{\phi})$ is a simple function that changes values whenever $\boldsymbol{\phi}$ crosses one of the $n(n-1)/2$ hyper-planes,

$$H_{ij} = \{\boldsymbol{\theta} \in \mathbb{R}^p \colon X_j - X_i = (\boldsymbol{Y}_{j-1} - \boldsymbol{Y}_{i-1})^{\mathsf{T}}\boldsymbol{\theta}\}.$$

This estimation procedure is invariant to the location of the process, and therefore ϕ_0 is not directly estimable via the proposed procedure. Hettmansperger & McKean (1998 Section 3.5.2) give a solution to this problem, discussing a median-based estimate of the intercept in the classical iid linear regression model. As we shall see, this procedure is also applicable to any AR(p) model that includes a location parameter.

The estimates obtained from minimizing (4) are referred to as weighted Wilcoxon (WW) estimates. For example, when $b_{ij} = 1$, (4) is equivalent (up to a constant) to Jaeckel's (1972) dispersion function with Wilcoxon scores (Hettmansperger & McKean, 1978). WW-estimates have been studied extensively in the context of linear regression (see e.g. Sievers, 1983; Naranjo & Hettmansperger, 1994; Hettmansperger & McKean, 1998 Chapter 5; Chang *et al.*, 1999). Depending on the type of weights employed, WW-estimates possess a continuous totally bounded influence function and have a positive breakdown point, possibly 50%. Given these properties, along with the fact that outliers in an AR(p) subsequently produce leverage points, WW-estimation in an AR(p) context provides a robust alternative to some of the existing estimation procedures.

This paper explores the use of WW-estimates in the context of autoregression. In Section 2 we define three basic weighting schemes that have appeared in the literature and briefly discuss their implementation. Section 3 compares various types of WW-estimates using several time series that are known to contain outliers. Section 4 presents asymptotic uniform linearity and the asymptotic distribution of the estimate for the three sub-classes of WW-estimates. Section 5 discusses applications of the theory to hypothesis testing and a model selection procedure. In Section 6 we present the results of a Monte Carlo study for autoregressive models. For a variety of different estimates we consider situations involving innovative and additive outlier models under several error distributions. To our knowledge, there exists no other study that compares such an assortment of general robust regression techniques for autoregressive time series estimation. Section 7 gives concluding remarks.

2. Three basic weighting schemes

The class of WW-estimates is essentially indexed by the type of weighting scheme employed. This section discusses three sub-classes: non-random weights, Mallows weights and Schweppe weights. All three sub-classes have been discussed in the linear regression literature in one form or another. We now discuss them in the context of autoregression.

2.1. Non-random weights

In general, we define non-random weights as any set of weights that do not explicitly depend on any observed values of the time series and are predetermined by the investigator. The simplest example is given by $b_{ij} = 1$ for all $i \neq j$ and 0 otherwise. In this case (4) reduces to

$$D(\boldsymbol{\phi}) = 2\sum_{i=1}^{n}\Big(R\big(\varepsilon_i(\boldsymbol{\phi})\big) - \frac{n+1}{2}\Big)\varepsilon_i(\boldsymbol{\phi}),$$

where $R(\varepsilon_i(\boldsymbol{\phi}))$ denotes the rank of $\varepsilon_i(\boldsymbol{\phi})$ among $\{\varepsilon_j(\boldsymbol{\phi})\}$. This corresponds to Jaeckel's (1972) dispersion function with Wilcoxon scores; hence the name WW-estimates.

Consider next a time series which can be divided into two or more groups based on *a priori* knowledge: for example, some catastrophic event, political change, or new law. This, of course, is often the case in time series analysis. In this situation it may or may not make sense to make between-group comparisons. If we let $b_{ij} = 1$ when all the elements of $(\boldsymbol{Y}_{i-1}, X_i, \boldsymbol{Y}_{j-1}, X_j)$ belong to the same group and 0 otherwise, then only within-group comparisons are used for estimating model parameters. We have found this type of weighting scheme useful when analysing time series that contain level shifts, gross outliers or missing values; see Section 3.1 (i.e. the Belgium phone call data) for an example.

2.2. Mallows weights

We define Mallows weights (Mallows, 1975) as any weighting scheme that depends on the observed values of the design points. Thus, in the context of WW-estimation, Mallows weights are of the form $b_{ij} = b(\boldsymbol{Y}_{i-1}, \boldsymbol{Y}_{j-1})$ for some weight function b. As a simple example, consider the AR(1) model and let the weights be defined by $b_{ij} = |X_{j-1} - X_{i-1}|^{-1}$. For this particular weighting scheme it can be shown that the dispersion function (see (4)) and its (negative) gradient are given by

$$D(\phi) = \sum_{1\le i<j\le n}\Big|\frac{X_j - X_i}{X_{j-1} - X_{i-1}} - \phi\Big| \qquad \text{and} \qquad S(\phi) = \sum_{1\le i<j\le n}\operatorname{sgn}\Big(\frac{X_j - X_i}{X_{j-1} - X_{i-1}} - \phi\Big),$$

respectively. Recall that the estimate of ϕ is an approximate solution of $S(\phi) = 0$. Hence, the estimate, say $\hat{\phi}_{\mathrm{PS}}$, corresponds to the median of the pairwise slopes; that is,

$$\hat{\phi}_{\mathrm{PS}} = \operatorname*{med}_{1 \le i < j \le n} \left\{ \frac{X_j - X_i}{X_{j-1} - X_{i-1}} \right\}. \tag{5}$$

In the simple linear regression context this estimate was first considered by Theil (1950) and later by Sen (1968). It is also similar in nature to the estimate proposed by Boldin (1994) who used $\mathrm{med}\{X_i / X_{i-1}\}$ to estimate ϕ.

Next, consider a factored weighting scheme of the form $b_{ij} = h_i h_j$ where

$$h_i = h(\boldsymbol{Y}_{i-1}) = \min\left\{1, \left(\frac{c}{d_i(\boldsymbol{Y}_{i-1})}\right)^{k/2}\right\}. \tag{6}$$

Here, $d_i^2(\boldsymbol{Y}_{i-1})$ denotes the squared Mahalanobis distance calculated with some (robust) measure of location and dispersion while c and k represent tuning constants. For example, the measures of location and dispersion, say $\hat{\boldsymbol{\mu}}_n$ and $\hat{\boldsymbol{\Sigma}}_n$, could be the minimum volume ellipsoid (MVE) estimates proposed by Rousseeuw & van Zomeren (1991) or the fast minimum covariance determinant (FAST-MCD) estimates proposed by Rousseeuw & Van Driessen (1999a); which are $n^{1/2}$ consistent. In the examples and simulations that follow we use the FAST-MCD estimates and set $c = \mathrm{med}\{d_i\} + 3\mathrm{MAD}\{d_i\}$ and $k = 4$. These weights have been studied extensively in the context of linear regression (see e.g. Naranjo & Hettmansperger, 1994; Naranjo *et al.*, 1994; McKean, Naranjo & Sheather, 1996a; Hettmansperger & McKean, 1998 Chapter 5). We follow the convention in the literature and refer to this particular WW-estimate as a GR-estimate.

2.3. Schweppe weights

We define Schweppe weights (see e.g. Handschin *et al.*, 1975; Rousseeuw & Leroy, 1987 p. 13; Coakley & Hettmansperger, 1993; and Chang *et al.*, 1999) as any weighting scheme that depends on both the design points and the response. This class of weights also yields robust estimates but typically has higher efficiency than, say, Mallows weighting schemes. As an example consider the following weights,

$$b_{ij} = \psi\left(\left|\frac{b}{a_i a_j}\right|\right), \qquad \text{where} \quad a_i = \frac{e_i(\hat{\boldsymbol{\phi}}_0)}{\hat{\sigma}_n \psi(c_{0.95}(\chi_p^2)/d_i^2(\boldsymbol{Y}_{i-1}))}, \tag{7}$$

where d_i is defined in (6), and $\psi(t) = 1,\ t,\ -1$ when $t \ge 1,\ -1 < t < 1,\ t \le -1$. The tuning constant, b, is set at $(\mathrm{med}\{a_i\} + 3\mathrm{MAD}\{a_i\})^2$ and

$$\hat{\sigma}_n = \mathrm{MAD} = 1.483\ \mathrm{med}_i\left\{|e_i(\hat{\boldsymbol{\phi}}_0) - \mathrm{med}_j\{e_j(\hat{\boldsymbol{\phi}}_0)\}|\right\}.$$

Lastly, $e_i(\hat{\boldsymbol{\phi}}_0)$ denotes the ith residual based on an initial estimate. In the sections that follow, $\hat{\boldsymbol{\phi}}_0$ is taken to be the fast least trimmed squares estimator proposed by Rousseeuw & Van Driessen (1999b), which is $n^{1/2}$ consistent. The weights in (7) were used by Chang *et al.* (1999) and Hettmansperger & McKean (1998 Section 5.8.1) in the linear regression context and correspond to the so-called HBR-estimate. Note that these weights fall into the category of Schweppe weights because residual information is incorporated.

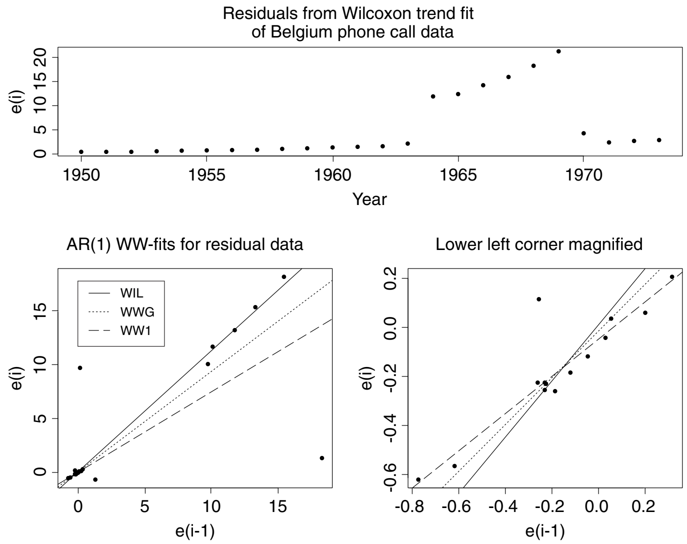

Figure 1. Time series plot and lag one scatter-plot with fits for the Belgium phone call data

Finally, note that $D(\boldsymbol{\phi})$ is a convex function, so it is easy to compute the WW-estimate once the weights (regardless of which type) are obtained. For instance, once the weights are calculated we can use any L_1 regression routine with $b_{ij}(X_j - X_i)$ and $b_{ij}(\boldsymbol{Y}_{j-1} - \boldsymbol{Y}_{i-1})$ playing the role of the response variables and design points, respectively, to calculate the estimates. This is analogous to weighted least squares regression. For example, the S-PLUS functions `cov.mve`, `cov.mcd`, `ltsreg`, `lmsreg` and `l1fit` can be used to compute the weights and obtain the WW-estimates. Alternatively, simple Gauss–Newton procedures (see Kapenga, McKean & Vidmar, 1998) can be used to minimize $D(\boldsymbol{\phi})$.

3. Examples

3.1. Belgium phone call data

Consider the Belgium phone call data given by Rousseeuw & Leroy (1987 p. 26 Table 2). The data portray the yearly total number (in tens of millions) of international phone calls made between 1950 and 1973. A time series plot (see Rousseeuw & Leroy, 1987 p. 25 Figure 3) of the data reveals an upward trend and a group of outliers. The outliers (years 1964–69) are the result of a new unit of measurement, namely the total number of minutes as opposed to calls. Additionally, the data for years 1963 and 1970 are suspect because these observations are sums of total number of calls and minutes due to mid-year changes.

We proceed by fitting a simple linear regression to the total calls versus year data using the Wilcoxon estimate. The fit is given by $-7.133 + 0.145(\text{year})$ and Figure 1 gives the corresponding residuals. We initially fit an AR(4) model to the residual data using the WW1 estimate discussed below, and then apply the model selection procedure discussed in Section 5 to identify an AR(1) model for the series. Note that this model is consistent with the partial autocorrelation function (PACF), which uses just the first 13 observations in the series. That is, the only significant lag in the PACF is the first lag. We therefore illustrate the use of non-random weighting schemes by fitting an AR(1) to the residual data.

There are at least three different ways to proceed. To begin, let $b_{ij} = 1$ for all (i, j) and recall that this is essentially the Wilcoxon (WIL) estimate on the residual data. Here, all comparisons are given an equal weight. In the light of the above outlier discussion it may make sense to divide the data into three groups. We let group I correspond to years 1950–62 and 1971–73, group II correspond to years 1964–69 and group III correspond to the years 1963 and 1970. Now let $b_{ij} = 1$ when all the elements of $(X_{i-1}, X_i, X_{j-1}, X_j)$ are from the same group (only I or II in this example) and 0 otherwise. We denote this estimate by WWG as only within-group comparisons are used to calculate the WW-estimate. Alternatively, we could decide to keep group I as is, collapse groups II and III, and allow only within group I comparisons. Here, the WW-estimate depends only on the good (group I) data so we denote it by WW1. Note that this is a somewhat conservative approach to what essentially amounts to deleting outliers.

Figure 1 also contains a lag one scatter-plot of the data along with the three fits under consideration. Note that WW1 fits the bulk of the data in the lower left corner of the plot and ignores the group II data whereas the WIL fit is biased towards group II. WWG could be viewed as a weighted average of the two groups. Depending on the goals of the analysis or subject matter expertise we could now decide on which fit is best. For example, the last three values of the residual series are –0.762, –0.607 and –0.552. The one-step-ahead forecasts for the WIL, WWG and WW1 estimates are –0.623, –0.538 and –0.466, respectively. Thus, if forecasting were the primary objective here, we might choose WW1 because its one-step-ahead predictor is consistent with the increasing trend of the last three values of the series.

3.2. Residential extensions data

The following is an example of a monthly time series (RESX), which originated at Bell Canada. A description of the data, along with the actual dataset, can be found in Rousseeuw & Leroy (1987 pp. 278–280). This dataset contains the number of telephone installations in a particular area and has two outliers. The first outlier occurs in November 1972 and is the result of a bargain month in which telephone installations were free. The second outlier occurs in December 1972 and is a consequence of the first because some of November's orders were not completed.

Several authors (e.g. Brubacher, 1974; Martin, 1980; Rousseeuw & Leroy, 1987 Section 7.2) have analysed the seasonal differenced series given by

$$X_i = \mathrm{RESX}_{i+12} - \mathrm{RESX}_i \qquad (i = 1, 2, \dots, 77).$$

We begin with a discussion of model selection. As an application of the theory presented in Sections 4 and 5 we propose a robust backward elimination method using Wald-type tests on a full model estimate of the AR(4) (see Section 5 for a discussion). Briefly, the procedure involves testing the sequence of hypotheses given in Table 1 until one is rejected at level α. Then, the order of the process is estimated by $4 - s + 1$ where s denotes the position of the rejected null hypothesis in the sequence. This method was applied for both the WIL and GR fits of the AR(4). Table 1 presents the results of the procedure. If we let $\alpha = 0.05$, then the results based on the GR indicate an AR(2) fits the data whereas the WIL results indicate an AR(3). Note that the order estimate based on the GR agrees with that of Martin (1980) who used a robust version of Akaike's AIC function to estimate the order of an autoregressive time series.

TABLE 1

Model selection results for the residential extensions data

Hypotheses	WIL $\hat{W}^2$	WIL P-value	GR $\hat{W}^2$	GR P-value
$H_0^{(1)}: \phi_4 = 0$	0.29	0.590	0.05	0.817
$H_0^{(2)}: \phi_3 = \phi_4 = 0$	7.79	0.020	2.48	0.290
$H_0^{(3)}: \phi_2 = \phi_3 = \phi_4 = 0$	58.75	0.000	12.18	0.007
$H_0^{(4)}: \phi_1 = \phi_2 = \phi_3 = \phi_4 = 0$	370.46	0.000	64.90	0.000

TABLE 2

Parameter estimates for the residential extensions data

Estimate	$\hat{\phi}_{n1}$	se $(\hat{\phi}_{n1})$	$\hat{\phi}_{n2}$	se $(\hat{\phi}_{n2})$
BRUB	0.530	na	0.370	na
LS	0.473	0.116	–0.166	0.116
RLS	0.412	na	0.501	na
LMS	0.393	na	0.674	na
LTS	0.340	na	0.576	na
M	0.503	na	–0.155	na
GM	0.413	na	0.333	na
WIL	0.503	0.026	–0.151	0.026
GR	0.369	0.089	0.388	0.083
HBR	0.403	0.074	0.303	0.076
WW1	0.384	0.108	0.328	0.109
THEIL	0.423	0.048	0.294	0.045

na: not available

Next we consider estimation of the AR(2) parameters. To begin, we fit the WIL and WW1-estimates defined in the previous example. Here, the WW1-estimate corresponds to two groups: group II pertains to the outliers for November and December 1972 while group I represents the remaining observations. Note that this is similar to the interpolation estimates of Brubacher (1974), denoted by BRUB, in that the outlier positions must be specified in advance. We also fit the GR and HBR-estimates which correspond to the weighting schemes given in (6) and (7), respectively. Our last WW-estimate corresponds to $b_{ij} = \|\boldsymbol{Y}_{j-1} - \boldsymbol{Y}_{i-1}\|^{-1}$. We call this the THEIL-estimate because it reduces to (5) in the case of the AR(1). In addition to the WW-estimates we also include the least squares (LS), least median squares (LMS), LMS-based re-weighted least squares (RLS) and least trimmed squares (LTS) estimates discussed in Rousseeuw & Leroy (1987 Section 1.2) for the sake of comparison. The M and GM-estimates of Martin (1980) are also given.

Table 2 shows the parameter estimates and standard errors for the AR(2) model. The WW-estimates (with the exception of the WIL-estimate) are in agreement with other well-known robust estimates. Furthermore, note the negative signs on the LS, M, and WIL-estimates of ϕ_2. This difference between the WIL-estimate and the other WW-estimates can provide valuable insight into the type of outliers present. See Section 6 for a discussion of the different types of outliers. For example, as our simulations in Section 6 show, the WIL-estimate is highly efficient under an innovation outlier model. However, when additive outliers are present the WIL-estimate is not robust and differs from the other WW-estimates; in particular, the GR and HBR-estimates. We are currently exploring the TDBETAS diagnostic proposed by McKean,

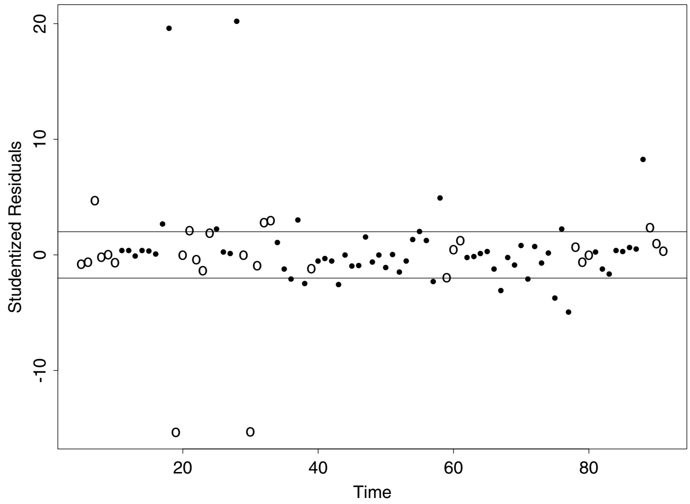

Figure 2. HBR (studentized) residuals for the AR(4) fit of the interest rates data

Naranjo & Sheather (1996b) for these types of comparisons. Furthermore, a plot similar to the one in Figure 2 indicated the outliers were of the additive variety. This example illustrates the well-known fact that not all robust estimates are appropriate for AR estimation when additive outliers are present (e.g. Denby & Martin, 1979).

3.3. Interest rates data

A quick inspection of a time series plot often reveals the atypical observations in the series. This was certainly true for the first two examples. Consider next the interest rates data found in Kunsch (1984a). The series consists of 91 monthly interest rates of an Austrian bank and is listed and plotted in Kunsch (1984b p. 259). Kunsch discovered that the series has three large outliers corresponding to months 18, 28 and 29. Upon further inspection, Kunsch uncovered five additional peculiarities corresponding to months 4, 7, 58, 77 and 88 which are not apparent from an inspection of the time series plot.

In our analysis we initially fitted an AR(6) model using the GR and HBR-estimates, and then applied the model selection procedure discussed in Section 5 to identify an appropriate model. Both procedures indicated an AR(4) model for the series. Figure 2 plots the (regression-based) studentized residuals (see e.g. Hettmansperger & McKean, 1998 Section 5.9; Chang *et al.*, 1999) for the HBR-fit of the AR(4) model. Because we are plotting the studentized residuals $(\hat{\varepsilon}_i/\mathrm{se}(\hat{\varepsilon}_i))$, which are adjusted for both location and scale in design space, a ± 2 cut-off point can be used to identify potential outliers. Additionally, for each residual we have incorporated information on the design point; the previous four lags in this example. A solid circle indicates a good design point in the sense that it is not outlying with respect to the others. Outliers were determined using a FAST-MCD-based version of Mahalanobis distance and a $c_{0.95}(\chi_4^2)$ cut-off value. Open circles indicate bad design points.

In simulated examples we have found this type of plot to be useful in identifying the types of outliers present. See Section 6 for a discussion of the different types of outliers. For example, residual outliers followed by open circles inside the ± 2 cut-off point are indicative of innovation outliers while those followed by open circles outside this region correspond to additive outliers. Figure 2 shows that the HBR-procedure identifies three additive outliers at

TABLE 3

Parameter estimates for the interest rates data

Estimate	$\hat{\phi}_{n1}$	se ($\hat{\phi}_{n1}$)	$\hat{\phi}_{n2}$	se ($\hat{\phi}_{n2}$)	$\hat{\phi}_{n3}$	se ($\hat{\phi}_{n3}$)	$\hat{\phi}_{n4}$	se ($\hat{\phi}_{n4}$)
LS	0.766	0.110	–0.130	0.134	0.264	0.135	–0.060	0.109
WIL	0.973	0.036	–0.065	0.044	0.113	0.044	–0.093	0.036
GR	0.955	0.061	0.080	0.084	0.055	0.085	–0.147	0.062
HBR	1.027	0.040	–0.046	0.050	0.064	0.051	–0.107	0.039

TABLE 4

Parameter estimates for the US exports data

Parameter	WIL	GR	HBR	LTS	LMS
$\hat{\phi}_{n1}$	–0.310	–0.404	–0.328	–0.525	–0.286
se ($\hat{\phi}_{n1}$)	0.051	0.060	0.051	na	na
$\hat{\phi}_{n2}$	–0.272	–0.270	–0.245	–0.355	–0.001
se ($\hat{\phi}_{n2}$)	0.054	0.066	0.056	na	na
$\hat{\phi}_{n3}$	–0.062	–0.033	–0.032	0.153	–0.208
se ($\hat{\phi}_{n3}$)	0.058	0.070	0.061	na	na
$\hat{\phi}_{n4}$	–0.074	–0.025	–0.064	–0.172	–0.401
se ($\hat{\phi}_{n4}$)	0.058	0.070	0.061	na	na
$\hat{\phi}_{n5}$	0.112	0.181	0.127	0.262	–0.134
se ($\hat{\phi}_{n5}$)	0.058	0.069	0.061	na	na
$\hat{\phi}_{n6}$	0.069	0.138	0.086	0.166	–0.103
se ($\hat{\phi}_{n6}$)	0.054	0.064	0.056	na	na
$\hat{\phi}_{n7}$	0.121	0.174	0.125	0.122	0.061
se ($\hat{\phi}_{n7}$)	0.050	0.059	0.051	na	na

na: not available

months 18, 28 and 29. The HBR-procedure also indicates innovation outliers at months 7, 58, 77 and 88, which is in agreement with the findings of Kunsch (1984a).

The estimates considered here were the LS-estimate and the three WW-estimates (WIL, GR and HBR) discussed in the previous example. Table 3 gives the estimates and standard errors for the AR(4) model. In this example there appear to be differences among the three WW-estimates. For example, the sign on $\hat{\phi}_{n2}$ is positive for the GR-estimate and negative for the WIL- and HBR-estimates. However, in all three cases this coefficient is not significantly different from zero. Furthermore, the lag three coefficient is significant for the WIL-estimate and non-significant for the GR and HBR-estimates. Thus, in terms of statistical significance, the GR- and HBR-estimates are in agreement. As pointed out in the Residential Extensions data example, this discrepancy between the WIL-estimate and the GR/HBR-estimates indicates the presence of additive outliers. This is consistent with our findings in Figure 2.

3.4. US exports data

As a final example consider the US Exports series discussed by Bruce & Martin (1989) and Garel & Hallin (1999), who kindly provided us with the data. The data are a monthly time series that represents differenced logarithms of US exports to Latin America from 1966 to 1983. The above authors uncovered four patches of outliers (1/69–2/69, 9/71–11/71, 12/76–2/77 and 1/78–2/78) and two isolated outliers (10/68 and 10/70) in the series. See either Bruce & Martin (1989) or Garel & Hallin (1999) for the details.

We initially fitted an AR(8) model using the WIL-, GR- and HBR-estimates, and then applied the model selection procedure discussed in Section 5 to identify an appropriate model for

TABLE 5

Spearman's rank correlogram for the US exports residual processes

Estimate	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8	Lag9	Lag10
WIL	–0.116	0.058	0.049	0.069	0.028	0.030	0.014	0.053	0.059	–0.116
GR	–0.031	–0.006	0.000	0.068	–0.005	–0.019	–0.000	0.084	0.013	–0.149
HBR	–0.099	0.022	0.022	0.090	0.017	0.003	0.014	0.067	0.042	–0.131
LMS	–0.019	–0.171	0.245*	0.270*	0.062	0.087	0.029	0.186*	0.138	–0.170
LTS	–0.011	0.005	–0.109	0.204*	–0.082	–0.035	0.084	0.101	–0.028	–0.153

*Significant at 1%

the series. The GR- and HBR-based procedures identified an AR(7) while the WIL-based procedure identified an AR(4). The AR(7) is consistent with the findings of Garel & Hallin (1999) who also found a significant lag seven correlation in the original series. The estimates, along with the least median squares (LMS) and the least trimmed squares (LTS) estimates discussed in Rousseeuw & Leroy (1987 Section 1.2), are given in Table 4. The most prominent feature of Table 4 is that the signs on $\hat{\phi}_{n3}$, $\hat{\phi}_{n5}$ and $\hat{\phi}_{n6}$ for the LTS- and LMS-estimates do not always agree with those on the three WW-estimates. This presents an interesting dilemma; which estimates are the more reliable? To help answer this question we consider the Spearman-based correlograms of the residuals given in Table 5. If, in fact, the AR(7) is the correct model, then we would expect the residuals to behave like white noise. This is not the case for the LMS- and LTS-fits because both exhibit significant correlations (see e.g. Hallin, Ingenbleek & Puri, 1985 p. 1171). In contrast, the three correlograms corresponding to the WW-estimates indicate white noise values.

4. Asymptotic theory

In this section we summarize the asymptotics associated with each of the three weighting schemes discussed in Section 2. In particular, the asymptotic distribution of the estimate, $\hat{\boldsymbol{\phi}}_n$, which is defined as any value that minimizes (4), is given for each case. We also discuss the asymptotic joint distribution of $(\hat{\phi}_0, \hat{\boldsymbol{\phi}}_n)$, where $\hat{\phi}_0 = \operatorname{med}\{X_i - \boldsymbol{Y}_{i-1}^{\mathsf{T}}\hat{\boldsymbol{\phi}}_n\}$.

We denote the true parameter vector (excluding the location term) for the AR(p) by $\boldsymbol{\phi}_0$ and let $\boldsymbol{\Delta} \in \mathbb{R}^p$. We parallel traditional rank-based methods of proof by defining the following functions of $\boldsymbol{\Delta}$,

$$D_n(\boldsymbol{\Delta}) = \frac{1}{n} D\Big(\boldsymbol{\phi}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\Big), \qquad \boldsymbol{S}_n(\boldsymbol{\Delta}) = -\frac{\partial}{\partial \boldsymbol{\Delta}} D_n(\boldsymbol{\Delta}) = \frac{1}{n^{3/2}} \boldsymbol{S}\Big(\boldsymbol{\phi}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\Big),$$

and

$$Q_n(\boldsymbol{\Delta}) = D_n(\mathbf{0}) - \boldsymbol{S}_n^{\mathsf{T}}(\mathbf{0})\boldsymbol{\Delta} + \boldsymbol{\Delta}^{\mathsf{T}}\boldsymbol{C}_F\boldsymbol{\Delta},$$

where D represents the dispersion function in (4), $\boldsymbol{S}$ is the negative of its gradient and $\boldsymbol{C}_F$ is a $p \times p$ matrix which depends on the error distribution F and the weighting scheme under consideration.

For each of the three cases we present the asymptotic uniform linearity (AUL) and asymptotic uniform quadraticity (AUQ) results as well as the asymptotic distribution of $\boldsymbol{S}_n(\mathbf{0})$. It is well known (see e.g. Jaeckel, 1972) that these results are sufficient for the asymptotic normality of $\hat{\boldsymbol{\phi}}_n$. In all cases,

$$\text{AUL: } \sup_{\|\boldsymbol{\Delta}\| \le c} \|\boldsymbol{S}_n(\boldsymbol{\Delta}) - \boldsymbol{S}_n(\mathbf{0}) + 2\boldsymbol{C}_F\boldsymbol{\Delta}\| = o_p(1) \qquad \text{for all } c > 0,$$

and

$$\text{AUQ: } \sup_{\|\boldsymbol{\Delta}\| \le c} |D_n(\boldsymbol{\Delta}) - Q_n(\boldsymbol{\Delta})| = o_p(1) \qquad \text{for all } c > 0.$$

Heiler & Willers (1988) have shown that AUL and AUQ are equivalent in the context of linear regression. In their proof, linearity and convexity are the only two characteristics of the regression model and dispersion function that are exploited. The $\mathrm{AR}(p)$ is linear in its parameters and our dispersion function is convex, so the equivalence of the above results in the context of the $\mathrm{AR}(p)$ is also true.

4.1. Non-random weights

We begin by summarizing the results for the non-random weights case. Let $b_{i\bullet} = (1/n)\sum_{j=1}^{n} b_{ij}$ and define the following matrices,

$$\boldsymbol{C}_A = \operatorname*{plim}_{n\to\infty} \frac{1}{n^2}\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}(\boldsymbol{Y}_{j-1}-\boldsymbol{Y}_{i-1})(\boldsymbol{Y}_{j-1}-\boldsymbol{Y}_{i-1})^{\mathsf{T}} \tag{8}$$

and
$$\boldsymbol{\Sigma}_A = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} b_{i\bullet}^2\, \boldsymbol{Y}_{i-1}\boldsymbol{Y}_{i-1}^{\mathsf{T}}\,. \tag{9}$$

Then the main results of this section are summarized in the following theorem.

Theorem 1. *Under suitable regularity conditions (e.g. those found in Terpstra, 1997 Section 2.2) we have*

(a) AUL *and* AUQ *hold with* $\boldsymbol{C}_F = \tau\boldsymbol{C}_A$ *and* $\tau = \mathrm{E}(f(\varepsilon_1))$,

(b) $\boldsymbol{S}_n(\boldsymbol{0}) \xrightarrow{\mathrm{d}} \mathrm{N}\big(\boldsymbol{0}, \tfrac{1}{3}\boldsymbol{\Sigma}_A\big)$,

(c) $\sqrt{n}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) \xrightarrow{\mathrm{d}} \mathrm{N}\big(\boldsymbol{0}, \tfrac{1}{12\tau^2}\boldsymbol{C}_A^{-1}\boldsymbol{\Sigma}_A\boldsymbol{C}_A^{-1}\big)$.

Detailed proofs of the results of this theorem can be found in Terpstra (1997 Chapter 2) in the case of the AR(1). However, the methods used there easily extend to the general $\mathrm{AR}(p)$ upon exploiting the geometric absolute regularity of the process in conjunction with martingale theory. We omit the details for the sake of brevity. Finally, note that consistent estimates of $\boldsymbol{C}_A$ and $\boldsymbol{\Sigma}_A$ are given by (8) and (9), respectively. Discussions of a consistent estimator for τ can be found in Koul (1992 Section 7.3c) and Koul, Sievers & McKean (1987).

4.2. Mallows weights

For Mallows-based weighting schemes we let $b_{ij} = b(\boldsymbol{Y}_{i-1}, \boldsymbol{Y}_{j-1})$, where $b(\boldsymbol{u}_1, \boldsymbol{u}_2)$ is a function from $\mathbb{R}^p \times \mathbb{R}^p$ to $\mathbb{R}$. Note that in instances such as GR-estimates the weight function may also depend on random measures of location and scatter as well as stochastic tuning constants. In such cases it may be more appropriate to write $b(\boldsymbol{Y}_{i-1}, \boldsymbol{Y}_{j-1}; \hat{\boldsymbol{\theta}})$ for $b(\boldsymbol{Y}_{i-1}, \boldsymbol{Y}_{j-1})$ where $\hat{\boldsymbol{\theta}}$ denotes a vector of estimated nuisance parameters. However, we assume throughout that $\hat{\boldsymbol{\theta}}$ can be replaced by its non-stochastic counterpart without affecting the asymptotic results. Now let G denote the marginal distribution for $\boldsymbol{Y}_{i-1}$ and define the following matrices,

$$\boldsymbol{C}_B = \tfrac{1}{2}\int\!\!\int (\boldsymbol{Y}_1 - \boldsymbol{Y}_0) b(\boldsymbol{Y}_0, \boldsymbol{Y}_1)(\boldsymbol{Y}_1 - \boldsymbol{Y}_0)^{\mathsf{T}}\, dG(\boldsymbol{Y}_0)\, dG(\boldsymbol{Y}_1)$$

and
$$\boldsymbol{\Sigma}_B = \int \boldsymbol{\Lambda}(\boldsymbol{Y}_0)\boldsymbol{\Lambda}^{\mathsf{T}}(\boldsymbol{Y}_0)\, dG(\boldsymbol{Y}_0), \text{ where } \boldsymbol{\Lambda}(\boldsymbol{Y}_0) = \int (\boldsymbol{Y}_1 - \boldsymbol{Y}_0) b(\boldsymbol{Y}_0, \boldsymbol{Y}_1)\, dG(\boldsymbol{Y}_1).$$

The main results of this section are summarized in Theorem 2.

Theorem 2. *Under suitable regularity conditions (e.g. those found in Terpstra & Rao, 2001) we have*

(a) AUL *and* AUQ *hold with* $\boldsymbol{C}_F = \tau \boldsymbol{C}_B$ *and* $\tau = \mathrm{E}(f(\varepsilon_1))$,

(b) $\boldsymbol{S}_n(\mathbf{0}) \xrightarrow{\mathrm{d}} \mathrm{N}\big(\mathbf{0}, \frac{1}{3}\boldsymbol{\Sigma}_B\big)$,

(c) $\sqrt{n}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) \xrightarrow{\mathrm{d}} \mathrm{N}\big(\mathbf{0}, \frac{1}{12\tau^2}\boldsymbol{C}_B^{-1}\boldsymbol{\Sigma}_B\boldsymbol{C}_B^{-1}\big)$.

The proofs of these results follow from the theory of U-statistics and absolutely regular processes. The details are omitted for the sake of brevity, but can be found in Terpstra & Rao (2001).

For Wald-type inference, consistent estimates of τ, $\boldsymbol{\Sigma}_B$, and $\boldsymbol{C}_B^{-1}$ are needed. A consistent estimate of τ was given in the non-random weights section. For $\boldsymbol{\Sigma}_B$ and $\boldsymbol{C}_B^{-1}$ note that $\boldsymbol{\Sigma}_B = \boldsymbol{\Sigma}_B(G)$ and $\boldsymbol{C}_B = \boldsymbol{C}_B(G)$ are actually functionals defined on an appropriate functional space, $\mathcal{G}$. Thus, natural estimates of $\boldsymbol{\Sigma}_B$ and $\boldsymbol{C}_B$ are the von Mises' functionals given by $\boldsymbol{\Sigma}_B(G_n)$ and $\boldsymbol{C}_B(G_n)$, where G_n assigns a probability of $1/n$ to each $\boldsymbol{Y}_{t-1}$, $t = 1, 2, \ldots, n$. It follows from Yoshihara (1976 Theorem 1 p.243), for example, that $\boldsymbol{\Sigma}_B(G_n)$ and $\boldsymbol{C}_B(G_n)$ are consistent estimates of $\boldsymbol{\Sigma}_B$ and $\boldsymbol{C}_B$, respectively. If we let $\boldsymbol{X}^\top = [\boldsymbol{Y}_0, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n-1}]$ and $\boldsymbol{W}$ denote an $n \times n$ matrix whose elements are given by

$$w_{ij} = \begin{cases} -\dfrac{1}{n} b_{ij} & i \neq j, \\ \dfrac{1}{n} \displaystyle\sum_{i \neq k} b_{ik} & i = j, \end{cases}$$

then we can see that $\boldsymbol{C}_B(G_n) = (1/n)\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X}$ and $\boldsymbol{\Sigma}_B(G_n) = (1/n)\boldsymbol{X}^\top \boldsymbol{W}^2\boldsymbol{X}$. A consistent estimate of $\boldsymbol{C}_B^{-1}$ is also obtained because the inverse of a matrix is a continuous mapping of its elements. These estimates are consistent for $\boldsymbol{C}_A$ and $\boldsymbol{\Sigma}_A$ as well.

4.3. Schweppe weights

Lastly, we summarize the results for Schweppe-type weighting schemes. Let $b_{ij} = b(\boldsymbol{Y}_{i-1}, \hat{\varepsilon}_i, \boldsymbol{Y}_{j-1}, \hat{\varepsilon}_j)$, where $b(\boldsymbol{u}_1, v_1, \boldsymbol{u}_2, v_2)$ is a function from $\mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}$ to $\mathbb{R}$ and $\hat{\varepsilon}_i$ is the ith residual from some initial fit. Similar to Mallows weights, the weight function may also depend on stochastic nuisance parameters. Let

$$\tau_F(\boldsymbol{Y}_0, \boldsymbol{Y}_1) = \int_{-\infty}^{\infty} b(\boldsymbol{Y}_0, \varepsilon, \boldsymbol{Y}_1, \varepsilon) f(\varepsilon)\, dF(\varepsilon)$$

$$A_F(\boldsymbol{Y}_0, \boldsymbol{Y}_1, \boldsymbol{Y}_2) = \int_{-\infty}^{\infty} B(\boldsymbol{Y}_0, \boldsymbol{Y}_1; \varepsilon) B(\boldsymbol{Y}_0, \boldsymbol{Y}_2; \varepsilon)\, dF(\varepsilon)$$

where
$$B(\boldsymbol{u}_1, \boldsymbol{u}_2; \varepsilon) = \int_{-\infty}^{\infty} \mathrm{sgn}(s - \varepsilon) b(\boldsymbol{u}_1, \varepsilon, \boldsymbol{u}_2, s)\, dF(s),$$

and define the following matrices,

$$\boldsymbol{C}_F = \tfrac{1}{2} \iint (\boldsymbol{Y}_1 - \boldsymbol{Y}_0)\tau_F(\boldsymbol{Y}_0, \boldsymbol{Y}_1)(\boldsymbol{Y}_1 - \boldsymbol{Y}_0)^\top\, dG(\boldsymbol{Y}_0)\, dG(\boldsymbol{Y}_1) \tag{10}$$

and
$$\boldsymbol{\Sigma}_F = \iiint (\boldsymbol{Y}_1 - \boldsymbol{Y}_0) A_F(\boldsymbol{Y}_0, \boldsymbol{Y}_1, \boldsymbol{Y}_2)(\boldsymbol{Y}_2 - \boldsymbol{Y}_0)^\top\, dG(\boldsymbol{Y}_0)\, dG(\boldsymbol{Y}_1)\, dG(\boldsymbol{Y}_2)\,. \tag{11}$$

The main results of this section are summarized in Theorem 3.

Theorem 3. *Under suitable regularity conditions (e.g. those in Terpstra, McKean & Naranjo, 2000) we have*

(a) AUL *and* AUQ *hold where* $\boldsymbol{C}_F$ *is defined in (10),*

(b) $\boldsymbol{S}_n(\boldsymbol{0}) \xrightarrow{\mathrm{d}} \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_F)$ *where* $\boldsymbol{\Sigma}_F$ *is defined in (11),*

(c) $\sqrt{n}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) \xrightarrow{\mathrm{d}} \mathrm{N}\big(\boldsymbol{0}, \tfrac{1}{4}\boldsymbol{C}_F^{-1}\boldsymbol{\Sigma}_F\boldsymbol{C}_F^{-1}\big)$.

The proofs of these results resemble proofs of the results based on Mallows weighting schemes. Terpstra *et al.* (2000) give the details and also Method of Moment-type estimates for $\boldsymbol{C}_F$ and $\boldsymbol{\Sigma}_F$. These estimates are similar in nature to those given in Chang *et al.* (1999) and Hettmansperger & McKean (1998 p. 323).

As previously mentioned, WW-estimates are invariant to the location of the process, and therefore ϕ_0 is not directly estimable via the proposed procedure. A solution to this problem is given in Hettmansperger & McKean (1998 Section 3.5.2), where a median-based estimate of the intercept in the classical iid linear regression model is discussed. However, this procedure can also be applied to the AR(p) model. More specifically, define the initial residuals as follows,

$$\hat{\varepsilon}_i = X_i - \boldsymbol{Y}_{i-1}^{\mathsf{T}}\hat{\boldsymbol{\phi}}_n \qquad (i = 1, 2, \ldots, n).$$

Then, a natural robust estimate of ϕ_0 is $\hat{\phi}_0 = \mathrm{med}_i\{\hat{\varepsilon}_i\}$.

Theorem 4 gives the asymptotic joint distribution of $\hat{\boldsymbol{\phi}} = (\hat{\phi}_0, \hat{\boldsymbol{\phi}}_n)$. The method of proof is analogous to Hettmansperger & McKean (1998 Section 3.5.2). However, the means by which these analogues are established is slightly different due to the stochastic nature of the AR(p) model. The details are omitted for the sake of brevity.

Theorem 4. *In addition to the regularity conditions of Theorems 1, 2 or 3, assume* $F(0) = \frac{1}{2}$ *and*

$$\sqrt{n}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Omega}(\boldsymbol{Y}_{i-1}, \varepsilon_i) + \boldsymbol{o}_p(1), \tag{12}$$

where $\boldsymbol{\Omega}\colon \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}^p$, $\mathrm{E}(\|\boldsymbol{\Omega}(\boldsymbol{Y}, \varepsilon)\|^2) < \infty$ *and* $\mathrm{E}(\boldsymbol{\Omega}(\boldsymbol{Y}_{t-1}, \varepsilon_t) \mid \boldsymbol{Y}_0, \varepsilon_1, \ldots, \varepsilon_{t-1}) = \boldsymbol{0}$. *Let* $\kappa(\boldsymbol{Y}_0, \varepsilon_1) = \mathrm{sgn}(\varepsilon_1) - \tau_s\mu\boldsymbol{1}^{\mathsf{T}}\boldsymbol{\Omega}(\boldsymbol{Y}_0, \varepsilon_1)$, *where* $\mu = \mathrm{E}(X_1)$; *then* $\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{\mathrm{d}} \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Phi})$, *where*

$$\boldsymbol{\Phi} = \begin{bmatrix} \tau_s^{-2}\mathrm{E}\big(\kappa^2(\boldsymbol{Y}_0, \varepsilon_1)\big) & \tau_s^{-1}\mathrm{E}\big(\kappa(\boldsymbol{Y}_0, \varepsilon_1)\boldsymbol{\Omega}^{\mathsf{T}}(\boldsymbol{Y}_0, \varepsilon_1)\big) \\ \tau_s^{-1}\mathrm{E}\big(\kappa(\boldsymbol{Y}_0, \varepsilon_1)\boldsymbol{\Omega}(\boldsymbol{Y}_0, \varepsilon_1)\big) & \mathrm{E}\big(\boldsymbol{\Omega}(\boldsymbol{Y}_0, \varepsilon_1)\boldsymbol{\Omega}^{\mathsf{T}}(\boldsymbol{Y}_0, \varepsilon_1)\big) \end{bmatrix}$$

and $\tau_s = 2f(0)$.

Note that (12) is *a priori* satisfied because $\boldsymbol{\Omega}$ essentially represents the projection of $\boldsymbol{S}_n(\boldsymbol{0})$ onto the sequence $\{\boldsymbol{Z}_t\}$ where $\boldsymbol{Z}_t = (\varepsilon_t, \boldsymbol{Y}_{t-1})$. For example, for Mallows-type weighting schemes and Schweppe-type weighting schemes we have, respectively,

$$\boldsymbol{\Omega}(\boldsymbol{Y}_{i-1}, \varepsilon_i) = \frac{1}{2\tau}\boldsymbol{C}_B^{-1}\big(2F(\varepsilon_i) - 1\big)\int b(\boldsymbol{Y}_{i-1}, \boldsymbol{Y}_1)(\boldsymbol{Y}_1 - \boldsymbol{Y}_{i-1})\,dG(\boldsymbol{Y}_1),$$

$$\boldsymbol{\Omega}(\boldsymbol{Y}_{i-1}, \varepsilon_i) = \tfrac{1}{2}\boldsymbol{C}_F^{-1}\int B(\boldsymbol{Y}_{i-1}, Y_1; \varepsilon_i)(\boldsymbol{Y}_1 - \boldsymbol{Y}_{i-1})\,dG(\boldsymbol{Y}_1).$$

Hettmansperger & McKean (1998 Section 1.5.5) give a consistent estimate of τ_s.

5. Wald-type inference

In this section we briefly discuss a Wald-type statistic for testing

$$H_0\colon \boldsymbol{A}\phi = \boldsymbol{\eta} \quad \text{versus} \quad H_1\colon \boldsymbol{A}\phi \neq \boldsymbol{\eta}\,, \tag{13}$$

where $\boldsymbol{\phi}$ denotes the true autoregressive parameter vector, $\boldsymbol{A}$ is a specified $q \times p$ matrix with rank q and $\boldsymbol{\eta}$ is a specified $q \times 1$ vector, typically the zero vector. To begin, let $\hat{\boldsymbol{\Sigma}}$ denote a consistent estimate of one of the variance–covariance matrices given in part (c) of Theorems 1, 2 or 3. Then, an approximate α-level test of (13) can be performed using the following Wald-type test statistic,

$$\hat{W}^2 = n(\boldsymbol{A}\hat{\boldsymbol{\phi}} - \boldsymbol{\eta})^{\mathsf{T}}(\boldsymbol{A}\hat{\boldsymbol{\Sigma}}\boldsymbol{A}^{\mathsf{T}})^{-1}(\boldsymbol{A}\hat{\boldsymbol{\phi}} - \boldsymbol{\eta})\,, \tag{14}$$

where $\hat{\boldsymbol{\phi}}$ denotes the corresponding estimate from Theorems 1, 2 or 3. The following theorem is a direct consequence of the theory presented in Section 4. For local behaviour of $\hat{W}^2$ we define the following sequence of alternative hypotheses: $H_n\colon \boldsymbol{A}\phi = \boldsymbol{\eta}_n$ where $\boldsymbol{\eta}_n = \boldsymbol{\eta} + n^{-1/2}\boldsymbol{\xi}$ and $\boldsymbol{\xi} \neq \mathbf{0}$.

Theorem 5. *Under the assumptions of Theorems 1, 2 or 3 we have*

(a) *under* H_0, $\hat{W}^2 \xrightarrow{\mathrm{d}} \chi^2_q$,

(b) *under* H_n, $\hat{W}^2 \xrightarrow{\mathrm{d}} \chi^2_q(\eta^*)$, *i.e. the non-central chi-squared distribution with* q *degrees of freedom and non-centrality parameter* $\eta^* = \boldsymbol{\xi}^{\mathsf{T}}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathsf{T}})^{-1}\boldsymbol{\xi}$,

(c) *the test: reject* H_0 *if* $\hat{W}^2 > c_{1-\alpha}(\chi^2_q)$ *is consistent under* H_1.

As an application of Theorem 5, consider a time series that can be described by an AR(p) model. We assume that p represents a maximum order for the process. In practice, we fit an AR(p) model with large p followed by a careful residual analysis to assess the adequacy of this model. Now let $0 \leq p_1 < p$, $p_2 = p - p_1$ and $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$, where $\boldsymbol{\phi}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{\phi}_2 \in \mathbb{R}^{p_2}$. Furthermore, let $\boldsymbol{A} = [\mathbf{0}\ \boldsymbol{I}]$, where $\mathbf{0}$ is the $p_2 \times p_1$ zero matrix and $\boldsymbol{I}$ is the $p_2 \times p_2$ identity matrix. Then an approximate α-level test of $H_0\colon \boldsymbol{\phi}_2 = \mathbf{0}$ versus $H_1\colon \boldsymbol{\phi}_2 \neq \mathbf{0}$ is given as: reject H_0 if $\hat{W}^2 > c_{1-\alpha}(\chi^2_q)$ where $\hat{W}^2$ is given in (14).

As above, let p represent a maximum order for the process. To determine a value for p_1 consider the following sequence of hypotheses,

$$H_0^{(i)}\colon \boldsymbol{\phi}_{2(i)} = \mathbf{0} \quad \text{versus} \quad H_1^{(i)}\colon \boldsymbol{\phi}_{2(i)} \neq \mathbf{0} \qquad (i = 1, 2, \ldots, p)\,,$$

where $\boldsymbol{\phi}_{2(i)} = (\phi_{p-i+1}, \phi_{p-i+2}, \ldots, \phi_p)$. Next, test $H_0^{(1)}$, $H_0^{(2)}$, … in succession with the Wald statistic until the first $H_0^{(i)}$ is rejected. Denote this rejected null hypothesis as $H_0^{(s)}$. There was insufficient evidence to reject the first $s-1$ null hypotheses, so assume the last $s-1$ autoregressive parameters are zero. However, because $H_0^{(s)}$ was rejected the last s autoregressive parameters are not all zero. Thus, $\hat{p}_1 = p - s + 1$ provides an estimate of the true order of the process. See Lütkepohl (1993 Section 4.2) for a formal discussion and some simulation results pertaining to this procedure. This is the method we used to identify the models for the examples appearing in Section 3.

6. A Monte Carlo study

We study the behaviour of several estimates, in particular the WW-estimates, via Monte Carlo in this section. For simplicity and to save computation time we consider only the AR(1). Thus, we define and denote the core process as

$$X_i = \phi_0 + \phi_1 X_{i-1} + \varepsilon_i \qquad (i = 1, 2, \ldots, n),$$

where $\phi_1 \in (-1, 1)$, X_0 is an observable random variable independent of $\{\varepsilon_i\}$, and the ε_i are iid according to some distribution F. Also, we define and denote the observed process as

$$X_i^* = X_i + \nu_i \qquad (i = 1, 2, \ldots, n),$$

where the ν_i are iid according to the mixture distribution, $(1-\gamma)\delta_0 + \gamma M$. Here, γ denotes the proportion of contamination, δ_0 is a point mass at zero, and M is the contaminating distribution function. Note that when $\gamma = 0$ the observed process reduces to the core process and we get Fox's (1972) Type II or Innovation Outlier (IO) model. As discussed by Rousseeuw & Leroy (1987 p. 275), this is the model that yields good leverage points in the sense that they have relatively little impact on estimates. When $\gamma > 0$ we get Fox's (1972) Type I or Additive Outlier (AO) model. This model produces bad leverage points which can have a significant impact on estimates, even some robust estimates (see e.g. Rousseeuw & Leroy, 1987 p. 275).

The following Monte Carlo study simulates, computes and compares 16 estimates (divided into four groups) of ϕ_1 under various combinations of IO and AO models and several values of F, γ and M. The estimates (and group numbers) are listed below:

L2(1)	Least squares estimate;
L1(1)	Least absolute deviation estimate;
WIL(1)	WW-estimate with $b_{ij} = 1$ for all (i, j);
LMS(2)	Least median squares estimate;
FLTS(2)	Fast least trimmed squares estimate;
RL2LMS(3)	LMS-based re-weighted least squares estimate;
RL2LTS(3)	FLTS-based re-weighted least squares estimate;
BOLD(3)	Weighted L1 with $w_i = \lvert X_{i-1}\rvert^{-1}$;
THEIL(3)	WW-estimate with $b_{ij} = \lvert X_{j-1} - X_{i-1}\rvert^{-1}$;
GM(4)	Default S-PLUS GM-estimate;
RL1GR(4)	Weighted L1 with $w_i = h_i$ as defined in (6);
RL1HBR(4)	Weighted L1 with $w_i = \psi(\lvert\sqrt{b}/a_i\rvert)$ as defined in (7);
GR(4)	WW-estimate as defined in (6);
GRB(4)	WW-estimate with $b_{ij} = b_i b_j$ and $b_i = 4b(\frac{1}{4}d_i)/d_i$ where d_i is given in (6) and b denotes Tukey's redescending bi-square function;
HBR(4)	WW-estimate as defined in (7);
HBR2(4)	WW-estimate with $b_{ij} = b_i b_j$, where

$$b_i = 1 - I\big(e_i^2(\hat{\boldsymbol{\phi}}_0) > \hat{\sigma}_n^2 c_{0.95}(\chi_1^2)\big) I\big(d_i^2 > c_{0.95}(\chi_p^2)\big)(1 - h_i).$$

The quantities defined in HBR2 are the same as those appearing in the GR and HBR-estimates.

The estimates in Group 1 are all sensitive to outliers appearing in design space, and consequently should not be completely robust for AR estimation. Group 2 contains the two (regression) high breakdown estimates discussed in Rousseeuw & Leroy (1987 Section 1.2). Recall that these objective functions are based on the ordered squared residuals. The estimates appearing in Group 3 are essentially weighted versions of the estimates in Group 1. For RL2LMS and RL2LTS we used the hard rejection weights given in Rousseeuw & Leroy (1987 p.17). Furthermore, recall that BOLD and THEIL essentially correspond to $\mathrm{med}\{X_i/X_{i-1}\}$ and (5), respectively. The estimates in Group 4 are also weighted versions of the estimates in Group 1. However, the weights used here are designed to identify and downweight leverage points. Recall that the GR-based schemes downweight all leverage points while the HBR-based weights attempt to distinguish between good and bad leverage points.

In all our simulations we generated 1000 realizations of size 100, set the mean of the process to 10, and let $\phi_1 \in \{-0.8, -0.4, 0, 0.4, 0.8\}$. With the exception of FLTS, the estimates were computed with S-PLUS using the algorithms and weighting schemes discussed in Sections 2 and 3. For comparison, we calculated the empirical ARE based on empirical MSEs relative to least squares. Our primary interest lies in the behaviour of the autoregressive parameter and not the intercept, so only the results for ϕ_1 are reported.

6.1. Isolated IO simulations

For each of the five values of ϕ_1 six error distributions were considered: the standard normal, Laplace, logistic, two contaminated normals and a centred Gamma (2,1/3) distribution. The standard normal, Laplace and logistic were chosen because these are the respective distributions where the L2, L1 and WIL-estimates are optimal (see e.g. Hettmansperger, 1984). For the contaminated normal distributions we chose $0.75\,\mathrm{N}(0,1) + 0.25\,\mathrm{N}(0,100)$ and $0.70\,\mathrm{N}(-1.5,1) + 0.30\,\mathrm{N}(3.5,1)$. The first distribution has heavy tails while the second distribution is bi-modal and skewed to the right. Note that the centred Gamma $(2, 1/3)$ distribution is also skewed to the right.

The results are summarized in Table 6. First, observe that the GR, GRB, and HBR-estimates are comparable to L2, L1, and WIL when the latter estimates are optimal. Furthermore, note that the WIL, HBR, and HBR2-estimates are consistent with one another and typically lie towards the top of the ARE scale while the estimates in Group 2 (LMS and FLTS), as well as the BOLD-estimate, are located towards the bottom. The one exception appears in the first contaminated normal model where the re-weighted least squares estimates seem to dominate. Furthermore, note that the GM-estimate is among the worst (lower third) estimates in four of the six models and finishes second to last in the second contaminated normal model. Another notable observation comes from the last two models where the WW-estimates yield high AREs for skewed distributions. Finally, note that in most cases the Schweppe-type weighting schemes produce slightly higher efficiencies than the Mallows-based schemes.

6.2. Isolated AO simulations

For isolated AO simulations we let ν_i be iid according to the following mixture distribution, $(1-\gamma)\delta_0 + \gamma M$, where M denotes some contaminating distribution function. Furthermore, it is assumed that the ε_i are independent of the ν_i. For each of the five values of ϕ_1 we ran simulations for the models given in Table 7. Note that these models include both symmetric and skewed distributions for ε_i and ν_i.

Table 8 summarizes the results. Recall that the AO model produces bad leverage points and that the estimates in Group 1 are not robust for these types of outliers. Hence, as expected, Group 1 contains the least efficient estimates when $\phi_1 \neq 0$. However, note that the WIL- and

TABLE 6

AREs relative to LS for IO simulations

Autoregressive parameter, ϕ_1						Autoregressive parameter, ϕ_1				
–0.8	–0.4	0	0.4	0.8		–0.8	–0.4	0	0.4	0.8
N(0, 1)						Laplace				
1.00	1.00	1.00	1.00	1.00	L2	1.00	1.00	1.00	1.00	1.00
0.66	0.65	0.66	0.62	0.69	L1	1.45	1.33	1.45	1.54	1.58
0.96	0.95	0.97	0.95	0.97	WIL	1.44	1.33	1.41	1.41	1.41
0.18	0.18	0.17	0.17	0.21	LMS	0.51	0.45	0.45	0.50	0.59
0.16	0.15	0.15	0.15	0.18	FLTS	0.60	0.56	0.51	0.57	0.70
0.77	0.74	0.77	0.76	0.84	RL2LMS	1.19	1.12	1.08	1.20	1.28
0.77	0.73	0.78	0.78	0.84	RL2LTS	1.28	1.19	1.18	1.29	1.38
0.41	0.42	0.45	0.36	0.44	BOLD	0.73	0.76	0.80	0.74	0.92
0.96	0.88	0.89	0.89	0.95	THEIL	1.38	1.22	1.20	1.27	1.28
0.88	0.89	0.84	0.92	0.91	GM	1.17	0.93	0.77	0.92	1.13
0.64	0.62	0.62	0.61	0.67	RL1GR	1.34	1.19	1.16	1.24	1.38
0.62	0.61	0.63	0.58	0.64	RL1HBR	1.39	1.24	1.15	1.36	1.47
0.94	0.90	0.91	0.91	0.95	GR	1.37	1.16	1.09	1.16	1.28
0.89	0.88	0.82	0.91	0.90	GRB	1.27	0.99	0.83	0.98	1.22
0.96	0.94	0.97	0.95	0.97	HBR	1.46	1.35	1.44	1.45	1.43
0.91	0.89	0.93	0.88	0.91	HBR2	1.49	1.36	1.38	1.46	1.44
Logistic						$0.75\,N(0, 1) + 0.25\,N(0, 100)$				
1.00	1.00	1.00	1.00	1.00	L2	1.00	1.00	1.00	1.00	1.00
0.88	0.81	0.80	0.83	0.88	L1	8.16	7.21	7.43	7.78	10.31
1.10	1.08	1.09	1.07	1.10	WIL	7.86	6.72	6.56	6.87	8.22
0.25	0.22	0.24	0.23	0.29	LMS	3.41	3.02	3.32	3.33	5.26
0.23	0.21	0.20	0.21	0.25	FLTS	3.58	2.59	3.05	3.08	5.04
0.85	0.83	0.83	0.84	0.89	RL2LMS	11.61	9.27	9.22	9.37	16.09
0.87	0.85	0.84	0.85	0.89	RL2LTS	11.99	8.36	8.50	9.67	15.90
0.53	0.46	0.54	0.42	0.58	BOLD	2.75	2.40	1.50	2.39	4.62
1.07	0.97	0.96	0.95	1.02	THEIL	6.80	5.21	3.67	5.61	7.45
1.02	0.91	0.83	0.92	1.02	GM	5.21	1.66	0.40	1.90	6.99
0.80	0.71	0.71	0.73	0.80	RL1GR	6.24	3.19	1.26	3.27	7.92
0.80	0.72	0.69	0.71	0.79	RL1HBR	7.43	3.99	3.42	4.54	10.94
1.06	0.95	0.95	0.93	1.00	GR	6.27	3.15	1.27	3.43	7.02
1.03	0.93	0.84	0.92	1.03	GRB	4.89	1.24	0.32	1.38	5.86
1.09	1.08	1.08	1.06	1.09	HBR	9.37	9.19	8.60	9.83	11.35
1.05	1.01	1.00	0.96	0.99	HBR2	9.51	8.17	7.70	8.44	12.15
$0.70\,N(-1.5, 1) + 0.30\,N(3.5, 1)$						Centred Gamma $(2, \frac{1}{3})$				
1.00	1.00	1.00	1.00	1.00	L2	1.00	1.00	1.00	1.00	1.00
1.08	1.01	1.09	1.19	1.29	L1	0.75	0.88	0.82	0.85	0.91
1.80	1.85	1.85	1.89	1.89	WIL	1.38	1.44	1.41	1.45	1.48
1.31	1.12	1.20	1.24	1.98	LMS	0.33	0.39	0.34	0.35	0.46
1.37	1.15	1.31	1.40	1.96	FLTS	0.30	0.33	0.30	0.29	0.43
2.06	1.78	1.94	2.12	2.69	RL2LMS	1.17	1.29	1.24	1.21	1.50
1.20	1.15	1.22	1.28	1.42	RL2LTS	1.13	1.27	1.13	1.17	1.45
0.20	0.61	0.81	0.39	0.26	BOLD	0.38	0.52	0.49	0.40	0.44
1.94	1.77	1.65	1.80	1.87	THEIL	1.32	1.34	1.23	1.27	1.38
0.88	0.87	0.86	0.84	0.86	GM	1.06	0.91	0.78	0.86	1.08
1.13	0.99	1.01	1.12	1.22	RL1GR	0.71	0.74	0.61	0.67	0.80
1.14	1.12	0.95	1.13	1.28	RL1HBR	0.75	0.85	0.65	0.68	0.86
1.90	1.80	1.58	1.77	1.83	GR	1.33	1.26	1.10	1.17	1.33
1.87	1.83	1.76	1.76	1.85	GRB	1.25	1.00	0.89	0.98	1.27
1.76	1.79	1.74	1.84	1.87	HBR	1.38	1.46	1.40	1.44	1.48
2.03	1.96	0.87	1.43	1.81	HBR2	1.43	1.54	1.19	1.17	1.36

TABLE 7

AO models used in the Monte Carlo study

Model	Distribution for ε_i	Distribution for ν_i
1	$N(0, 1-\phi_1^2)$	$0.95\delta_0 + 0.05\,N(0, 100)$
2	$N(0, 1-\phi_1^2)$	$0.90\delta_0 + 0.10\,N(0, 100)$
3	$0.70N(-1.5, 1) + 0.30N(3.5, 1)$	$0.85\delta_0 + 0.15(0.70\,N(-10, 100) + 0.30\,N(20, 25))$
4	Centred Gamma $(2, \frac{1}{3})$	$0.85\delta_0 + 0.15(0.80\,N(20, 25) + 0.20\,N(-10, 49))$
5	$0.95\,N(0, 1-\phi_1^2) + 0.05\,N(0, 100)$	$0.85\delta_0 + 0.15(0.80\,N(20, 25) + 0.20\,N(-10, 49))$
6	$0.70\,N(-1.5, 1) + 0.30\,N(3.5, 1)$	$0.85\delta_0 + 0.15(0.90\,N(0, 1) + 0.10\,N(0, 100))$

HBR-estimates are highly efficient when $\phi_1 = 0$. This can best be explained by the fact that the AO model reduces to an AR(0) when $\phi_1 = 0$. That is, a true AR model is being observed in this case. Furthermore, note that for the estimates outside Group 1 the ARE is an increasing function of $|\phi_1| \neq 0$. This is not surprising because it is well known that the lag one correlation of the core process is equal to ϕ_1. A common trait among the robust estimates in Groups 2–4 is that they all try to base the estimate on a good cluster of data. However, as ϕ_1 increases, this cluster becomes tighter and tighter, hence improving the estimate. Unlike the situation in the IO study where the GM-estimate was among the worst estimators, the GM-estimate has the highest ARE in 14 of the 24 cases where $\phi_1 \neq 0$. However, notice that it has poor ARE at Model 6 as well as all the situations where $\phi_1 = 0$. Thus, it would have poor power relative to other estimators to detect alternatives close to zero. The GRB and/or GR-estimates finish among the top three estimators in 18 of the 24 cases and are generally close (in ARE) to the GM-estimate in the cases where it dominates. Hence, contrary to the IO model, Mallows-based weighting schemes seem to be more efficient than Schweppe-type weighting schemes when an AO model holds. As a final observation note that the class of WW-estimates performs very well under Model 6. Outside $\phi_1 = 0.8$, the three best estimators are all WW-estimates.

A large-scale simulation study which covers all possible IO/AO combinations is virtually impossible. Instead, we have focused on isolated IOs and AOs for various error distributions and degrees of contamination. Our simulations have shown that there is no overall favourite among the estimates considered. However, when considered as a family, the WW-estimates performed very well across all the simulations. In every case at least one of the weighting schemes was among the top performing estimates.

7. Conclusion

This paper explores the use of a family of weighted Wilcoxon estimates for autoregressive time series estimation. In Section 2 we identified three sub-classes of the so-called WW-estimates. These sub-classes contain the well-known WIL-, GR- and HBR-estimates found in Hettmansperger & McKean (1998 Chapter 5). The WW-estimates are computed using an L_1 routine once the weights are obtained. Furthermore, the weights themselves can be obtained using routines in S-PLUS. In Section 3 we investigated the performance of these WW-estimates on a variety of real-world examples, demonstrating that important information on a dataset can be obtained readily by using several weighting schemes on a problem. These examples illustrated how the non-random weights sub-class could be used to control the effects of level shifts, gross outliers and or missing values. In Section 4 asymptotic linearity properties and the asymptotic normality of the three sub-classes of WW-estimates were presented. This allows

TABLE 8

AREs relative to LS for AO simulations

Autoregressive parameter, ϕ_1						Autoregressive parameter, ϕ_1				
–0.8	–0.4	0	0.4	0.8		–0.8	–0.4	0	0.4	0.8
		Model 1						Model 2		
1.00	1.00	1.00	1.00	1.00	L2	1.00	1.00	1.00	1.00	1.00
1.87	1.05	1.70	1.14	1.79	L1	1.19	1.04	3.35	1.09	1.19
1.38	1.03	2.24	1.11	1.37	WIL	1.07	1.03	4.21	1.08	1.09
13.77	1.44	0.22	1.48	13.28	LMS	16.68	1.45	0.47	1.49	13.31
12.19	1.28	0.19	1.37	11.43	FLTS	15.71	1.39	0.41	1.43	11.67
32.84	2.07	0.93	2.08	28.65	RL2LMS	27.81	1.67	1.72	1.65	18.04
32.92	1.99	0.96	2.01	22.43	RL2LTS	24.99	1.67	1.92	1.62	14.32
22.53	3.39	0.45	3.32	17.92	BOLD	15.37	3.42	0.64	3.06	12.62
18.56	3.98	1.14	3.64	14.23	THEIL	7.44	2.80	1.73	2.54	5.83
67.05	8.82	0.75	9.45	53.71	GM	60.68	9.68	0.57	9.01	47.52
40.37	5.40	0.49	5.70	33.23	RL1GR	30.89	5.89	0.55	5.35	24.33
34.71	2.44	0.58	2.53	27.38	RL1HBR	29.43	2.04	0.85	1.97	18.11
44.98	6.99	0.70	6.60	34.57	GR	24.65	6.28	0.71	5.35	18.37
60.16	8.10	0.65	8.33	47.00	GRB	53.47	7.74	0.45	7.23	39.30
17.76	2.01	1.56	1.96	13.15	HBR	6.67	1.56	2.62	1.53	5.46
33.12	2.20	0.99	2.13	21.15	HBR2	21.04	1.78	1.62	1.69	12.82
		Model 3						Model 4		
1.00	1.00	1.00	1.00	1.00	L2	1.00	1.00	1.00	1.00	1.00
1.35	1.03	4.24	1.21	1.26	L1	1.11	0.93	1.71	1.14	1.53
1.13	1.06	5.01	1.14	1.12	WIL	1.19	1.04	2.60	1.08	1.15
9.22	2.09	4.51	2.42	7.03	LMS	7.78	1.08	1.13	1.34	8.18
21.84	3.22	6.76	3.36	15.03	FLTS	9.05	1.02	1.22	1.36	8.27
7.26	1.69	5.21	1.87	5.92	RL2LMS	5.48	1.24	3.13	1.24	4.65
8.83	1.88	4.95	1.93	7.55	RL2LTS	3.75	1.17	3.01	1.23	3.81
5.66	3.09	1.64	2.79	4.08	BOLD	2.01	1.61	0.83	1.83	3.56
3.77	2.36	3.56	2.39	3.21	THEIL	2.39	1.77	1.99	1.66	2.08
10.03	6.17	0.63	6.64	9.01	GM	2.09	2.78	0.59	2.68	2.33
9.03	4.12	1.54	4.80	6.73	RL1GR	2.16	1.81	0.78	2.24	3.39
11.51	2.86	5.58	3.04	8.45	RL1HBR	2.72	1.40	1.19	1.17	3.90
5.16	3.86	1.86	4.22	4.42	GR	2.30	2.33	1.25	2.19	2.15
3.86	1.69	5.33	1.78	3.58	GRB	1.49	1.27	2.83	1.35	1.68
1.80	1.34	6.93	1.45	1.72	HBR	1.21	1.04	2.80	1.19	1.42
6.33	2.25	7.49	2.18	5.40	HBR2	2.75	1.57	2.17	1.05	2.53
		Model 5						Model 6		
1.00	1.00	1.00	1.00	1.00	L2	1.00	1.00	1.00	1.00	1.00
0.96	0.91	15.84	1.05	1.07	L1	2.28	1.20	1.18	1.66	2.38
0.99	0.95	19.01	1.06	0.99	WIL	2.66	1.65	1.89	1.74	2.50
65.32	1.23	10.13	1.35	70.12	LMS	3.92	1.79	1.26	2.24	5.03
74.61	1.25	4.88	1.41	55.17	FLTS	4.42	1.97	1.42	2.45	5.85
96.13	1.26	32.18	1.41	94.53	RL2LMS	4.27	2.15	1.61	2.41	4.82
103.20	1.32	26.86	1.48	67.55	RL2LTS	3.38	1.65	1.23	1.78	3.69
7.78	2.23	2.46	2.58	16.76	BOLD	0.75	1.01	0.80	0.78	0.93
5.64	1.67	9.18	1.75	4.64	THEIL	4.31	2.52	1.66	2.67	3.76
33.71	10.31	0.53	11.26	66.21	GM	2.86	1.54	0.86	1.63	2.57
42.92	5.83	0.92	6.43	49.21	RL1GR	3.02	1.69	0.98	2.14	2.97
90.58	1.40	6.40	1.55	68.17	RL1HBR	3.30	2.11	1.07	2.02	3.15
29.04	5.33	1.11	5.34	24.93	GR	4.66	2.70	1.50	2.90	3.97
34.22	3.61	3.74	3.46	69.92	GRB	3.82	1.93	1.71	2.07	3.60
11.32	1.25	19.79	1.44	24.05	HBR	3.50	2.10	1.85	2.30	3.28
88.52	1.36	16.12	1.41	58.58	HBR2	4.96	3.15	1.07	2.14	3.99

for asymptotic inference once consistent estimates of the unknown quantities are calculated. For example, a Wald-type statistic was discussed in Section 5 and applied to a model selection procedure which was used in the examples of Section 3. Lastly, the Monte Carlo study in Section 6 supported the use of WW-estimates for autoregressive estimation across a variety of outlier situations, including IO and AO models. In general, there is no overall favourite among the different estimates and various outlier models considered. In all these situations, though, several members of the family of WW-estimates were either at the top or near the top in ARE. We think this family of weighted Wilcoxon estimates offers the user an easily computable, robust family of estimates for autoregressive models. Besides robust estimates, the use of the various weighting schemes offers the user informative diagnostic procedures for analysing time series datasets.

References

BOLDIN, M.V. (1994). On median estimates and tests in autoregressive models. *Math. Methods Statist.* **3**, 114–129.

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

BRUBACHER, S.R. (1974). *Time Series Outlier Detection and Modeling with Interpolation.* Technical report. Bell Laboratories.

BRUCE, A.G. & MARTIN, D. (1989). Leave-k-out diagnostics for time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **51**, 363–424.

BUSTOS, O.H. (1982). General M-estimates for contaminated pth-order autoregressive processes: consistency and asymptotic normality. *Z. Wahrsch. Verw. Gebiete* **59**, 491–504.

CHANG, W.H., MCKEAN, J.W., NARANJO, J.D. & SHEATHER, S.J. (1999). High-breakdown rank regression. *J. Amer. Statist. Assoc.* **94**, 205–219.

COAKLEY, C.W. & HETTMANSPERGER, T.P. (1993). A bounded influence, high breakdown, efficient regression estimator. *J. Amer. Statist. Assoc.* **88**, 872–880.

DENBY, L. & MARTIN, R.D. (1979). Robust estimation of the first-order autoregressive parameter. *J. Amer. Statist. Assoc.* **74**, 140–146.

DOUKHAN, P. (1994). Mixing: properties and examples. In *Lecture Notes in Statistics*, eds S. Fienberg *et al.*, Vol. 85. New York: Springer-Verlag.

FERRETTI, N.E., KELMANSKY, D.M. & YOHAI, V.J. (1991). Estimators based on ranks for ARMA Models. *Comm. Statist. Theory Methods* **20**, 3879–3907.

FOX, A.J. (1972). Outliers in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **34**, 350–363.

GAREL, B. & HALLIN, M. (1999). Rank-based autoregressive order identification. *J. Amer. Statist. Assoc.* **94**, 1357–1371.

HALLIN, M., INGENBLEEK, J-F. & PURI, M.L. (1985). Linear serial rank tests for randomness against ARMA alternatives. *Ann. Statist.* **13**, 1156–1181.

HANDSCHIN, E., KOHLAS, J., FIECHTER, A. & SCHWEPPE, F. (1975). Bad data analysis for power system state estimation. *IEEE Trans. Power Apparatus and Systems* **2**, 329–337.

HANNAN, E.J. (1970). *Multiple Time Series.* New York: John Wiley and Sons.

HEILER, S. & WILLERS, R. (1988). Asymptotic normality of R-estimates in the linear model. *Statistics* **19**, 173–184.

HETTMANSPERGER, T.P. (1984). *Statistical Inference Based on Ranks.* New York: John Wiley and Sons.

HETTMANSPERGER, T.P. & MCKEAN, J.W. (1978). Statistical inference based on ranks. *Psychometrika* **43**, 69–79.

HETTMANSPERGER, T.P. & MCKEAN, J.W. (1998). *Robust Nonparametric Statistical Methods.* Great Britain: Arnold.

JAECKEL, L.A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43**, 1449–1458.

KAPENGA, J.A., MCKEAN, J.W. & VIDMAR, T.J. (1988). *RGLM: Users Manual.* New Orleans: Amer. Statist. Assoc. Short Course on Robust Statistical Procedures for the Analysis of Linear and Nonlinear Models.

KOUL, H.L. (1992). Weighted empiricals and linear models. In *Lecture Notes — Monograph Series*, ed. R.J. Serfling, Vol. 21. Hayward, California: Institute of Mathematical Statistics.

KOUL, H.L. & OSSIANDER, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Ann. Statist.* **22**, 540–562.

KOUL, H.L. & SALEH, A.K.M.E. (1993). R-estimation of the parameters of autoregressive [AR(p)] models. *Ann. Statist.* **21**, 534–551.

KOUL, H.L., SIEVERS, G.L. & MCKEAN, J.W. (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scand. J. Statist.* **14**, 131–141.

KUNSCH, H. (1984a). Infinitesimal robustness for autoregressive processes. *Ann. Statist.* **12**, 843–863.

KUNSCH, H. (1984b). Programs for robust estimation in autoregressive model fitting. *Proc. Inst. Statist. Math.* **32**, 241–260. (In Japanese)

LÜTKEPOHL, H. (1993). *Introduction to Multiple Time Series Analysis*, Second edn. Springer-Verlag.

MALLOWS, C.L. (1975). On Some Topics in Robustness. Unpublished Memorandum. Murray Hill, NJ: Bell Telephone Laboratories.

MARTIN, R.D. (1980). Robust estimation of autoregressive models (with discussion). In *Directions in Time Series*, eds D.R. Brillinger & G.C. Tiao, pp. 228–254. Hayward, CA: Institute of Mathematical Statistics.

MARTIN, R.D. & YOHAI, V.J. (1991). Bias robust estimation of autoregression parameters. In *Directions in Robust Statistics and Diagnostics Part I*, eds W. Stahel & S. Weisberg, Vol. 33, pp. 233–246. New York: Springer-Verlag.

MCKEAN, J.W., NARANJO, J.D. & SHEATHER, S.J. (1996a). An efficient and high breakdown procedure for model criticism. *Comm. Statist. Theory Methods* **25**, 2575–2595.

MCKEAN, J.W., NARANJO, J.D. & SHEATHER, S.J. (1996b). Diagnostics to detect differences in robust fits of linear models. *Comput. Statist.* **11**, 223–243.

NARANJO, J.D. & HETTMANSPERGER, T.P. (1994). Bounded influence rank regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **56**, 209–220.

NARANJO, J.D., MCKEAN, J.W., SHEATHER, S.J. & HETTMANSPERGER, T.P. (1994). The use and interpretation of rank-based residuals. *J. Nonparametr. Statist.* **3**, 323–341.

ROUSSEEUW, P.J. & LEROY, A.M. (1987). *Robust Regression and Outlier Detection.* New York: John Wiley and Sons.

ROUSSEEUW, P.J. & VAN DRIESSEN, K. (1999a). A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41**, 212–223.

ROUSSEEUW, P.J. & VAN DRIESSEN, K. (1999b). *Computing LTS Regression for Large Data Sets.* Technical report. University of Antwerp.

ROUSSEEUW, P.J. & VAN ZOMEREN, B.C. (1991). Robust distances: simulations and cutoff values. In *Directions in Robust Statistics and Diagnostics Part II*, eds W. Stahel & S. Weisberg, Vol. 34, pp. 195–203. New York: Springer-Verlag.

SEN, P.K. (1968). Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Statist. Assoc.* **63**, 1379–1389.

SIEVERS, G.L. (1983). A weighted dispersion function for estimation in linear models. *Comm. Statist. Theory Methods* **12**, 1161–1179.

TERPSTRA, J.T. (1997). A Robust Estimate for an Autoregressive Time Series. PhD Dissertation, Western Michigan University.

TERPSTRA, J.T., MCKEAN, J.W. & NARANJO, J.D. (2000). Highly efficient weighted Wilcoxon estimates for autoregression. *Statistics* **35**, 45–80.

TERPSTRA, J.T. & RAO, M.B. (2001). Generalized rank estimates for an autoregressive time series: a U-statistic approach. *Stat. Inference Stoch. Process.* **4**, 155–179.

THEIL, H. (1950). A rank-invariant method of linear and polynomial regression analysis (parts 1–3). *Ned. Akad. Wet. Proc. Ser. A* **53**, 386–392, 521–525, 1397–1412.

YOSHIHARA, K. (1976). Limiting behavior of U-statistics for stationary absolutely regular processes. *Z. Wahrsch. Verw. Gebiete* **35**, 237–252.