MULTIVARIATE ROBUST FILTERING

Robert Kunst

Forschungsbericht/
Research Memorandum No. 226

Jänner 1986

Zusammenfassung

Die vorliegende Arbeit präsentiert eine Erweiterung einer Idee von Kleiner, Martin&Thomson(1979) auf multivariate autoregressive Prozesse. An einigen Beispielen mit Daten aus dem Gebiet der österreichischen Wirtschaft werden die Eigenschaften der Verfahren vorgeführt.

Ausgehend von der Annahme, daß die Beobachtungen zwar einem autoregressiven Vektor-Prozeß entsprechen, jedoch durch zusätzliche additive Störungen verunreinigt sind, wird versucht, diese Störungen zu eliminieren, um den "wahren" Prozeß und sein Bildungsgesetz zurückzugewinnen. Dies geschieht durch den in Abschnitt 2 dargelegten multivariaten robusten Filter.

In Abschnitt 3 wird das Basisexperiment vorgestellt. Quartalsdaten für Geldmenge und Brutto-Inlandsprodukt werden, unterschiedlich bereinigt, dem Verfahren unterzogen. Abschnitt 4 widmet sich kleinen Kontrollexperimenten mit alternativen Methoden. Abschnitt 5 beschäftigt sich damit zu untersuchen, inwieweit künstlich erzeugte Ausreißer erkannt werden und den Iterationsprozeß beeinflussen.

In Abschnitt 6 wird gezeigt, daß die Resultate nicht zuletzt von der Weite des Bandes abhängen, welches "guten Daten" a priori zugestanden wird. Abschnitt 7 ist einem Experiment mit Aktiendaten gewidmet, wobei die Grenzen der Anwendbarkeit des Verfahrens deutlich werden, wenn die Ausreißer einem anderen als dem Basismodell folgen. In Abschnitt 8 wird eine weitere Abweichung in der Struktur des robusten Filters ausgeleuchtet, nämlich die mangels mathematischer Beweise eher willkürliche Wahl der Gewichtsfunktion. Abschnitt 9 präsentiert Vorschläge zum Messen der Eigenschaft einzelner Beobachtungen, Ausreißer zu sein.

Abstract

This paper presents an extension of an idea by Kleiner, Martin & Thomson(1979) to multivariate autoregressive processes. The properties of the procedures are reported by some examples with economic data.

Starting from the assumption that observations obey an autoregressive vector process but are contaminated by additive disturbances, it is endeavoured to eliminate the disturbances to regain the "true" process and its law of generation. This is done iteratively by the multivariate robust filter set forth in chapter 2.

Chapter 3 presents the basic experiment. The procedure is applied to quarterly data for the Austrian monetary base and gross domestic product after stationarizing the sample in two different ways. Chapter 4 is devoted to small experiments with alternative methods. Chapter 5 investigates into how the results are affected by artificially generated outliers and whether these outliers are safely detected.

In chapter 6 it is shown that the results depend on the width of the band within which data are still judged to be "good". Chapter 7 reports an experiment with stock price data where the outliers are supposed to obey a law different from the basic assumption of robust filtering. Consequently, the results are poor. In chapter 8 the effects of changing the weighting function in the filter are illustrated. Finally, chapter 9 summarizes suggestions for measuring the "outlier-ness" of single observations based on the application of the filtering procedure.

1. Introduction

This paper deals with a bivariate version of the robust filtering method introduced by Kleiner, Martin&Thomson(1979). Such procedures pretend to serve a triple purpose. First, to detect outliers in a time series sample. Second, to adjust the outliers to estimates of the true non-outlier observations. Third, to estimate the process which generates the true uncontaminated data. It is plain to see that such a procedure must rely on specific assumptions. The chosen class of time series models, i.e. finite-order autoregressive processes, plays a role and the concept of true uncontaminated data disturbed by noise has to be accepted.

The bulk of this paper concentrates on the first purpose of the filter. Contrary to e.g. the concept of Kitagawa&Akaike(1982) the procedure makes strictly binary decisions depending on one or few critical parameters but without a priori specification of the possible number of outliers. Thus, it is rather to be compared to standard regression diagnostics. Consequently, the terms "type I" and "type II errors" in this paper always refer to the null hypothesis that the given observation is a non-outlier.

Exactly retrieving the original process - even if such a one exists - is unlikely. The second purpose should only be seen as a means for the third purpose, that is identifying the true behaviour of the data-generating process. To this end, several procedures are available, like the GM also introduced in Kleiner et al.(1979) or the FLS by Heathcote&Welsh(1983). It is primarily the cumbersome minimization procedures involved with such methods which makes them less attractive than the multifunctional and easy-to-use robust filtering. A comparison of performance would require extensive simulations which are definitely beyond the scope of this paper.

2. The procedure

The procedure used throughout this paper is based on the univariate robust filtering algorithm developed by Thomson(1977) and presented to the public by Kleiner et al.(1979). Kleiner et al. postulate a so-called additive outliers model (AO) that somehow resembles an errors-in-variables model as it is common in econometrics. Let (y_t^*) be a stationary process and let us restrict our attention to autoregressive processes for the ease of notation. Then Kleiner's model reads

$$(2.1) \qquad \begin{aligned} y_t^* &= \Sigma\, a_j y_{t-j}^* + e_t \\ y_t &= y_t^* + f_t \end{aligned}$$

Only (y_t) is observed but y_t^*, e_t, f_t are unknown. (e_t) and (f_t) represent uncorrelated disturbances. The distribution of (f_t) is usually specified to be a mixed one so that for 90-95% of the sample observed y_t and true stationary y_t^* are equal. It is unreasonable to be too specific about distribution assumptions since robustness against deviations in this direction is one of the features required from the procedures to be introduced. Now, in order to do away with the obvious biases in the estimation of a_j, which result from the application of standard procedures on the basis of the y_t to AO, Kleiner et al. suggest the following procedure:

a) non-robust estimates of the a_j are provided as starting values, e.g. by OLS, using the observed y_t instead of the unknown y_t^* data.

b) the robust filter

$$y_t = \Sigma a_j y_{t-j} + cs\Phi((y_t - \Sigma a_j y_{t-j})/cs)$$

is applied to obtain the y_t. Kleiner et al. provide suggestions for the specification of c,s and Φ. Step b generates new data that are supposed to be closer to the y_t^* than the observed y_t.

c) a procedure similar to step a is applied to the y_t to generate new estimates for a_j. Steps b and c are repeated until the coefficient estimators converge.

Although the procedure has been motivated by plausibility arguments and some simulations seem to exist (documented not only in Kleiner et al., but also in Martin(1981)) a rigid proof for its convergence and a mathematical analysis of its properties are still missing.

There are several directions into which the filter may be generalized. First, s is meant to be an estimate of scales which is usually calculated from the data. It could be replaced by a matrix S to account for the cross-correlations between the regressors y_{t-j}. This approach is compatible with robust regression techniques but, keeping in mind that the procedure is ad-hoc in principle, seems to be too rigid and cumbersome, especially as the "tuning constant" c remains undefined. Second, Φ could be estimated from the data, adaptively in the sense of Hogg(1974). However, as some of the comments on his paper remarked, simple weighting functions that are robust over a wide class of distributions are preferable even in the regression context. This holds even more for robust filtering as the remaining parts of the procedure are rather ad hoc. To overcome this ad-hoc-ness the filter could be made more sophisticated, as it was done by Martin(1979) who introduced the ACM filter. Again, a rigid proof is missing and, according to the author's experience, all that can be gained is more sophistication.[1]

Few economic time series (e.g interest rates) are optimally represented by univariate models. Most series are endogenous in the sense of Granger causation, which means that other series incorporate information about their behaviour additionally to univariate history. This fact suggests the most useful extension of robust filtering, namely to multivariate time series. Let us retain the AO model but now with matrix coefficients and vector processes. Then

$$(2.2) \qquad y_t = \Sigma a_j y_{t-j} + cs\Phi((y_t - \Sigma a_j y_{t-j})/cs)$$

could still be used for updating vector y_t, provided we know about s and the function Φ which now maps R^n into R^n. By neglecting cross-influences Φ may be assumed to be a diagonal functional matrix so that (2.2) may be written

$$(2.3) \qquad \begin{aligned} y_{1t} &= \Sigma a_{11j} y_{1t-j} + \ldots + \Sigma a_{1nj} y_{nt-j} + cs_1\Phi((y_{1t} - \\ &\qquad\qquad \Sigma\Sigma a_{1ij} y_{it-j})/cs_1) \\ &\ldots \\ y_{nt} &= \Sigma a_{n1j} y_{1t-j} + \ldots + \Sigma a_{nnj} y_{nt-j} + cs_n\Phi((y_{nt} - \\ &\qquad\qquad \Sigma\Sigma a_{nij} y_{it-j})/cs_n) \end{aligned}$$

where Φ now denotes a function mapping R^n into R. This is more restrictive than (2.2) since the component functions are assumed to be structurally the same, but it is also more general as s_i are allowed to vary between components which is compatible with calculating s_i from the individual residuals, again neglecting cross-terms. In order not to discard existing cross-influences completely it should be kept in mind to calculate the t-th values from the freshly corrected series up to t-1 and not to run the filter component after component. If some of the components are exogenous in the sense of Granger, updating could be forfeited for them so that the filter easily generalizes to ARX models. However, in that case it does not adjust for additive disturbances in the exogenous variables.

For the sake of simplicity, a bivariate version of (2.3) will be used throughout this paper. However, only the generalization from univariate to bivariate models is problematic, so that no important case is excluded from the investigation. This bivariate robust filter will be denoted as BRF whereas its univariate counterpart will be labeled URF. Before reporting about empirical findings, the specifications for c,s, and Φ have to be fixed.

Kleiner et al.(1979) define s as "the positive square root of a robust variance estimate based on the prediction residuals". In a different part of their paper such an estimate is provided, namely

$$(2.4) \qquad s=\text{median}(|r_t|)/0.6745$$

which means in words the median of the absolute values of the residuals, corrected by a constant to make sure that s is just the standard deviation in the case of normally distributed r_t. Not all researchers of the URF algorithm used an estimate based on the residuals. Martin(1981)'s version uses an s based on the data which, as the scale of the data is higher, avoids the problems with c yet to be mentioned.

Kleiner defines c to be a "constant chosen to obtain high efficiencies for both" the undisturbed and the outliers model if the percentage of outliers in the data is "not too large". This, of course, is not a definition. Empirically, it can be found that the obvious choice c=1.0, together with a standard Φ and the above-mentioned s, leads to an excessive amount of outliers corrected (about 25-30%). This may even be proved theoretically for the Gaussian model. c=2 usually gives about 5% outliers which looks reasonable. Therefore, c=2 was chosen as the basic specification. The effects of changing the tuning constant are reported in chapter 6.

There are several suggestions for the selection of Φ. Kleiner et al. use a "redescending" function which stays close to the y=x line around zero and asymptotically softly approaches zero. Martin prefers the so-called Hampel function which is not differentiable and drops to zero sharply. In this paper the Huber function

$$(2.5) \qquad \Phi(x) = \begin{matrix} 1 & x > 1 \\ x & |x| < 1 \end{matrix}$$

$$-1 \qquad x < -1$$

is retained. A filter using this function leaves well predictable time series values exactly where they are (which the redescending function does not) and depends on one constant only which can be completely tuned by c (which is opposed to the Hampel function which depends on two or three constants). Huber's Φ does not descend to zero which means that the information a big outlier contains about the sign of the underlying disturbance is still useful for model prediction. For a discussion of this feature see chapter 8.

3. Application to data

To gain insight into the properties of the multivariate filter it is applied to Austrian economic data. The chosen series are the gross domestic product (GDP) and the monetary base (MB), quarterly data from 1964.1 to 1985.2 in both cases. The investigation of money/income-relationships has a long tradition, noteworthy publications are the causality tests for US data by Sims(1972) and for Canadian data by Hsiao(1979), each of them using slightly different techniques. For Austria, the monetary base represents "money" better than e.g. M1, a fact that is well-known to Austrian investigators.

With these two series, bivariate models are specified. Straightforward modelling is inappropriate, as raw data of GDP and MB are non-stationary. To take out the trend, the data were logged and differenced once. With these data,

$$X_t^1 = \log(X_t / X_{t-1})$$

from 1964.2 to 1985.2, the first model was built. It has already been mentioned that the BRF needs the number of lags as input. The (transformed) GDP needs 4 autoregressive lags and MB 8 lags according to Akaike's AIC if modelled univariately. Specification of off-diagonal lags by AIC is less reliable and might furthermore be more sensitive to outlier elimination, so off-diagonal lags were overall set to 8. This specification forms the basis for the application of BRF.

By looking at the spectra and cross-spectra of the series, strong seasonal patterns become visible, so that this first model is still inappropriate. After taking fourth differences to generate

$$X_t^2 = X_t^1 - X_{t-4}^1 = \log(X_t/X_{t-1}) - \log(X_{t-4}/X_{t-5})$$

both series are approximately stationary. In this case, AIC detects 8 lags for both variables. The off-diagonals were again set to 8. Now, with an appropriate and an inappropriate specification at hand, the outcomes can be compared.

If the inappropriate model is used, BRF with c=2 detects 4 outliers in the MB and 7 outliers in the GDP series. This means that MB is smoother than GDP containing a lower percentage of outliers, i.e. less than 5 percent versus about 9 percent. More detailed results are given in table 1.

If the appropriate model is used, BRF with c=2 corrects 3 MB observations and 3 GDP observations. This result is satisfactory insofar as the amount of outliers has diminished relative to the inappropriate model. The points in time of the outliers are worth mentioning: 1971.4, 1974.4, and 1983.1 for money and 1972.1, 1972.4, and 1974.4 for income. Only two outliers coincide between the de-seasonalized and the non-de-seasonalized version. The full results are displayed in table 2.

A Granger-like procedure has been applied to both models, that is, the joint significance of the eight lags of each of the two variables in an OLS regression with the caused variable as regressand has been calculated by the prob values of the F test. For the results see table 3.

The diagonal elements of table 3 are only given for the sake of completeness. Let us focus our attention to the outcome of the appropriate model first. Causality clearly runs from GDP to MB, the causal direction becoming stronger and the feedback becoming weaker after correcting the data by BRF. The latter effect shows that the BRF is not fooled by insignificant parts of the lag polynomial.

As a means of diagnostic check, the correlation of the residuals from the appropriate model has been calculated. Correlation is very low, decreasing from -.06 to -.04 by BRF correction. This means that no "instantaneous causality" is present. The overall fit of the MB regressions (r^2c) is marginally improved by BRF from .386 to .418, whereas in the GDP regressions it falls from .345 to .338. This indicates that the BRF does not simply adapt the data to the given model as might be argued. Compare the S values in table 1 which even rise for the GDP equation.

At the other hand, the results from using the inappropriate model (also listed in table 3) show tendencies in the filtered data toward generating feedback. It is almost certain that by adverse specification of the lag structure a significant feedback can be produced. These results may suffice for concluding that the model class on which the BRF is based must be carefully selected and if this class requires stationary data these data must be provided.

Table 1: BRF APPLIED TO THE INAPPROPRIATE MODEL

# of observations: 85
# of outliers in MB series: 4 (1973.2,1983.1,1983.4,1984.4)
# of outliers in GDP series: 7 (1972.4,1974.2,1976.2,1978.1,1980.2,1983.1,1983.3)
# of iterations: 5

AR polynomials fitted to data: (x=MB,y=GDP)

dependent variable	before BRF		after BRF	
	x	y	x	y
x_{-1}	-.11	.14	-.14	.12
x_{-2}	-.21	-.71	-.22	-.82
x_{-3}	-.00	.11	-.00	.09
x_{-4}	.09	.17	.10	.24
x_{-5}	-.18	-.28	-.21	-.32
x_{-6}	-.11	.04	-.12	.05
x_{-7}	-.11	-.05	-.09	-.02
x_{-8}	-.10	-.03	-.12	-.08
y_{-1}	-.02	.05	-.00	.07
y_{-2}	.10	.18	.11	.22
y_{-3}	-.07	-.07	-.06	-.07
y_{-4}	.26	.87	.27	.87
y_{-5}	.06		.06	
y_{-6}	-.07		-.07	
y_{-7}	.07		.06	
y_{-8}	-.00		-.00	
S	.017	.016	.017	.018

S is the robust standard deviation of residuals as defined in (2.4.)

Table 2: BRF APPLIED TO THE APPROPRIATE MODEL

# of observations:	81	
# of outliers in MB series:	3	(1971.4,1974.4,1983.1)
# of outliers in GDP series:	3	(1972.1,1972.4,1974.4)
# of iterations:	3	

AR polynomials fitted to data (x=MB,y=GDP)

dependent variable	before BRF x	before BRF y	after BRF x	after BRF y
x_{-1}	-.14	-.02	-.13	-.02
x_{-2}	.05	-.08	.06	-.10
x_{-3}	-.04	.07	-.05	.05
x_{-4}	-.47	.09	-.48	.08
x_{-5}	-.17	-.02	-.18	-.02
x_{-6}	.08	.02	.10	.01
x_{-7}	-.00	.02	.00	.01
x_{-8}	-.35	.11	-.35	.09
y_{-1}	-.42	-.31	-.46	-.31
y_{-2}	.04	-.07	.01	-.08
y_{-3}	-.14	.09	-.16	.09
y_{-4}	-.02	-.45	-.01	-.45
y_{-5}	-.05	-.10	-.06	-.11
y_{-6}	-.08	-.17	-.09	-.17
y_{-7}	.02	-.08	.01	-.09
y_{-8}	-.19	-.32	-.21	-.31
S	.016	.013	.016	.013

Table 3: GRANGER CAUSALITY IN THE BIVARIATE MODELS (PROB VALUES)

a) inappropriate model

	before BRF		after BRF	
	cause			
	MB	GDP	MB	GDP
effect				
MB	69.94	.04	59.70	.03
GDP	36.13	.00	14.89	.00

b) appropriate model

	before BRF		after BRF	
	cause			
	MB	GDP	MB	GDP
effect				
MB	.67	1.75	.43	.64
GDP	66.78	.09	77.57	.12

4. Alternative methods

Putting aside the problems of stochastic regressors which are an inherent component of time series models, usual test statistics for OLS routines can be applied to vector autoregressive models. One of the most common tests for outliers is simply calculating the t-value for a dummy variable taking a single observation out of the data set. This is a bivariate procedure which should be able to separate bivariate outliers from spurious univariate irregularities. This t-test was applied to the correctly transformed GDP series, modeled with 8 GDP lags and 8 MB lags. The following table gives the observations with the highest t-statistics:

1972.1	2.50	1976.1	1.89
1974.4	2.50	1979.1	1.77
1975.2	2.25	1972.2	1.76
1972.4	2.12	1978.1	1.62

To reach a 5%-level of significance, values over 1.99 are needed. Out of the four values satisfying this condition, three are BRF outliers but one (1975.2) is not. This means that the results of both tests are similar but differ.

Of course, extensive simulations or thorough mathematical calculations would be needed to decide upon the better procedure of the two. However, it may be suspected that the t-test makes mistakes as it uses lagged values which are eventually contaminated, so that both type I and type II errors might appear. This might have happened in the above example with the 1972.2 and 1975.2 observations that closely follow 1972.1 and 1974.4, respectively. The BRF routine, at the other hand, uses corrected data for its investigation and thus avoids spurious patches of outliers.

Another procedure suggests itself for a comparison with BRF, namely the univariate version URF. Even before comparing the two on the basis of a data set, we presume that at least some URF outliers in one series are not BRF outliers since their behaviour can be explained by fluctuations in the other series which, in their place, might or might not be BRF outliers. If URF is applied to the (correctly transformed) GDP series, it detects 6 outliers, that is the observations 1972.1, 1972.4, 1974.4, 1975.2, 1978.1 and 1979.1. The values of 1975.2, already known by its high t-value (2.25), 1979.1 and 1978.1, sixth and eighth in the list of t-statistics, now joined the ranks. It may be summarized that URF and BRF results differ and that BRF is theoretically more appropriate in a multivariate framework. This conclusion is corroborated by running URF over the MB data. URF corrects 4 data (1973.2,1975.1,1978.2,1983.1), only one of which is a BRF outlier. Since the MB series seems to depend strongly on the GDP series, the more marked change in this case is plausible. The magnitude of the number of BRF iterations (3) exactly equals that of the two URFs.

5. Robustness of BRF towards artificial outliers

There is a certain disadvantage with empirical data since we do not really know whether outliers be present. Neither do we know about the true model class nor the true parameters. However, as the small bivariate time series model seems to fit the money/income data quite well, it can be used as a starting point for an investigation into how BRF reacts if data are changed without having to perform costly Monte Carlo simulations.

To this end, a single observation of MB (1976.2), rather in the middle of the sample and without any indication of being an outlier, was made into an outlier by adding values of 1.0, 2.0 etc. which are quite big relative to the scale of the data. The following table reports the increase in the number of iterations and the change in the number of outliers detected by BRF (including the artificial observation). Let us ignore the last column for the moment.

value added	0.0	1.0	2.0	3.0	4.0	5.0	10.0	2/-2
#iterations	3	8	8	8	10	8	10	10
#outliers MB	3	4	5	5	4	5	5	4
#outliers GDP	3	4	4	4	4	4	4	4

It can be seen that the number of iterations increases with the size of the outlier only marginally once the outlier is definitely too large to be ignored. The number of outliers remains approximately the same. With the GDP series, all modifications report just the same outliers, namely the original BRF outliers plus the 1979.1 observation which obviously represents a borderline case. With the MB series, the artificial observation is always detected and only one outlier (1978.2) appears that is not familiar to us from the original data. This might be a spurious effect but 1978.2 represents a URF outlier, anyway. The original MB outliers 1974.4 and 1983.1 persist while the 1971.4 irregularity is lost in some specifications.

To gain insight into the effects of "patchy" outliers, two consecutive observations (1976.2 and 1976.3) were changed by the same amount but with alternating signs. The picture remains roughly the same: both outliers are detected, the GDP outliers are unaffected, but the increase in the number of iterations is marked and the non-artificial outliers detected are the original 1974.4 and the questionable 1978.2. The loss of the 1983.1 outlier is remarkable as this observation represents the "most outlying" one (see chapter 9).

However, the original purpose of the BRF was ,at least, twofold: to correct the data and to correct the estimated AR process. So what happened to the AR pattern ? With one outlier, the original AR model (the one fitted to the filtered data) is more or less reproduced, with the difference in the coefficients below the threshold given for the convergence of the program (0.01). With two consecutive outliers, BRF faces slightly more problems with retrieving the original model. The largest deviations and the most marked changes during the iterations are found in the Φ_{12} polynomial which describes the dependence of the disturbed MB series on the undisturbed GDP series. The Φ_{11} coefficients quickly converge.

It may be concluded that BRF generally safely detects outliers and shows only little tendencies of generating new and spurious ones. BRF is satisfactorily robust in this direction and certainly more so than any method not using the prediction/correction framework.

6. Changing the tuning constant

As mentioned above, by the tuning constant c the amount of spurious outliers detected by BRF can be determined if the model is uncontaminated. This means that the approximate probability of type I errors can be fixed, provided the residual distribution or at least some of its fractiles are known. If the assumed residual distribution is misspecified, this probability is incorrect. If the model is misspecified, too, the residuals will be bad approximations for the true innovations. Since we assume certain misspecifications if we run a BRF procedure, empirical simulations are a more useful method for deciding upon c than theoretical ratiocinations.

When Kleiner et al.(1979) suggested selecting c to obtain high power for the procedure they obviously mean that several c's should be tried and the "optimal" c be fixed ex post. It seems more desirable to fix c ex ante in order to receive a fixed index for the data contamination, e.g. the percentage of data corrected. If this index is higher than some bounds the data can be said to be "moderately" or "highly" contaminated which might lead to a revision of the hypothesized time series model.

The outcome of the BRF procedure with c=2.0 applied to the data set MB/GDP, both appropriately transformed, has already been reported. 3 outliers in a sample of 81 with the first 8 observations kept constant gives an outlier share of 3/73 which is about 4%. If c is gradually increased, the outliers disappear one by one. This phenomenon could even be used for generating a descriptive statistic for the "significance" of an outlier, namely the value of c where the outlier disappears (see chapter 9). If c is decreased more data are corrected. If c is decreased below one, the iteration process sometimes refuses to converge. The following list compares the results from c=2 to those from c=1.5:

c	#outliers		#iter.	max.abs.change			
	MB	GDP		Φ_{11}	Φ_{22}	Φ_{12}	Φ_{21}
1.5	12	10	7	.053	.046	.127	.043
2.0	3	3	3	.023	.008	.039	.017

With c=1.5 the share of corrected data approaches 15%, which makes c-values less than 1.5 look rather unreasonable. The set of c=1.5 outliers contains the c=2 outliers. Their distribution across time is maybe interesting. For MB, one outlier is situated within the 1960s (1967.3), four in the early 1970s and seven from 1981 onward. This cumulation of MB outliers in the 1980s indicates that the money series has become more irregular lately. For GDP, the outliers are rather equally distributed between 1969 and 1980, three values within the years 1974 and 1975, during the oil crisis. Only one outlier is found outside this interval, namely 1984.2. The last three columns in the list provide the maximal change in the coefficients of the lag polynomials. Obviously, it is the cross-terms which cause the main impediment for BRF convergence. The value of .127 corresponds to a change in the first-order parameter of the Φ_{12} polynomial from -.423 to -.550. That polynomial reflects the influences running from the GDP to the MB series.

Again, the example is no substitute for extensive Monte Carlo simulations. However, the interval [1.5,2] looks the most interesting for the tuning constant. Of course, the selection of the tuning constant somehow relies on the Huber weighting function. For experiments with different functions see chapter 8.

7. Innovations outliers

The usefulness of the filter critically depends on the basic assumptions for the outliers model. As long as the AO specifications are retained - including near-normal distribution of the true innovations and a mixed law for the disturbance process - BRF will work. The power of the procedure will increase with thicker tails of the disturbance distribution. At the other hand, if the true innovations distribution becomes worse the probability of type I errors increases. More good data will be erroneously corrected.

An outliers model which incorporates deviations within the innovations so that they feed back into the succeeding data is called an IO (innovations outliers) model. From theoretical work it is well known that usual autoregressive procedures are consistent under these circumstances and even show faster convergence than with normal errors. This makes the application of filtering methods such as BRF not only useless but even dangerous. The faster convergence relies on the additional information inherent in the "heavy tails" of the law which is lost completely. Moreover, it is unknown whether BRF autoregressive estimates are consistent at all in the basic specification and this is even more doubtful in the IO case.

Data which are known in econometrics to follow some kind of IO models are e.g. stock price data (see Fama(1965)). To see what happens if robust filtering is incorrectly applied to IO data, the univariate version URF was run over daily data of stock prices of the Austrian Steyr company, logged and differenced once according to theory. From the 248 data, 96 were zero which means that the price remained unchanged at that day. The filter was run with 2 lags only and corrected not a single one of the zero observations. The results are summarized in table 4.

Table 4: URF WITH P=2 APPLIED TO STEYR STOCK PRICE DATA

# of observations:	248
# of zero observations:	96
# of corrected data:	42
# of iterations:	4

AR(2) fitted to original data: $x_t = .089x_{t-1} + .118x_{t-2} + e_t$ (S=.0056)

AR(2) fitted to filtered data: $x_t = .169x_{t-1} + .190x_{t-2} + e_t$ (S=.0051)

correction patches with more than 2 observations:

obs. #24-26 (3 obs.)
obs. #48-52 (5 obs.)
obs. #106-108 (3 obs.)
obs. #196-198 (3 obs.)
obs. #231-233 (3 obs.)

# of corrected data in patches: 17

8. Changing the weighting function

Up to this point, it was taken as granted that Huber's weighting function offers itself as the basis for robust procedures. It joins the advantages of the quadratic loss function being rather smooth around zero and of the LAD function giving less weight to extreme observations. Huber(1964) has shown that this function has nice properties as regards to robustness representing a minimax solution over a wide class of probability densities. Huber's Φ may be derived from the likelihood function for a mixed law, namely a convolution of a Gaussian and a Laplace distribution. Many researchers, however, among them Martin(1981), prefer different functions in the robust filtering context for several reasons.

The Huber function, if used in outlier correction, replaces large deviations from the forecasted value by the forecasted value plus the scale estimate. This means that even an arbitrarily large observation is regarded as containing information about the sign even though the value itself is obviously wrong. "Redescending" functions like the Hampel or Andrews function would replace gross outliers by just the forecasted value. Moreover, using redescending functions can be motivated by concepts like the breakdown point or the influence function (see Hampel(1974)). For the sake of an experiment, the Huber function in the basic specification was replaced by a Hampel function with the following design:

$$\Phi(x) = \begin{array}{ll} 0 & |x| > 3c/2 \\ x & |x| < c \\ c-2(x-c) & c < x < 3c/2 \end{array}$$

$$-c-2(x+c) \quad -3c/2 < x < -c$$

The results of this exercise are more or less discouraging. With c=2 the procedure refuses to converge. With c around 3 it reports no outliers. The effects of choosing c to be 2.3 in the Hampel weighting function are documented in table 5. The Hampel-BRF outliers form a subset of the Huber-BRF outliers. For a direct comparison, c=2.3 looks too high. However, reducing c puts up difficulties. Apart from the consequent lengthy iteration process and the strong adjustment of lag polynomials, it seems that the BRF then loses track of the data in the way described by Martin(1981) who used an URF and purported that the observed effect were impossible with his ACM version of the filter. Our experience suggests that URF or BRF and Hampel's Φ is a combination not to be used. Of course, the Hampel function may be modified in the tail areas but we do not see any reason why to replace a well-behaved function with a doubtful one which seems hyper-sensitive against small changes in the function design.

The theoretical reason for selecting a certain weighting function should rely on a priori assumptions about the shape of the laws determining the innovations and the outliers. Even with Gaussian innovations, large residuals could result from correct data in a correct model. If the disturbance density is heavier-tailed than the normal law, this probability decreases with increasing the size of the residuals. With Huber's Φ, a marginal belief in the correctness of the data remains if the residuals are taken to infinity, whereas with Hampel's Φ this belief reaches level zero for some finite value (3c/2 in the above example). This is compatible with bounded innovations only. But if the true innovations were bounded - a highly non-standard assumption - the least-squares estimation involved with the BRF is no longer optimal. These ratiocinations are ad hoc and demand for explicit proofs. Our conjecture is that under the standard AO assumptions the optimal weighting function would necessarily be unbounded but with an asymptotically sublinear behaviour. It is likely that optimization with respect to parameter estimation and with respect to outlier correction or detection will imply different "optimal" functions. The preliminary conclusion is that more "robust" functions (like the Hampel redescending) make the BRF less robust.

Table 5: APPLICATION OF HAMPEL'S Φ TO "APPROPRIATE" MODEL

# of observations:	81	
# of outliers in MB series:	2	(1974.4,1983.1)
# of outliers in GDP series:	1	(1974.4)
# of iterations: 3		

max. abs. change in polynomial coefficients

in Φ_{11}:	.037
in Φ_{12}:	.028
in Φ_{22}:	.035
in Φ_{21}:	.017

9. Measuring the size of outliers

As mentioned above, BRF relies on binary decisions in analogy to classical testing in the spirit of Neyman and Pearson. The design is set before observations are divided into good observations and outliers. This philosophy is necessary for the filtering procedure and must be retained but several measures for the significance of an outlier may be suggested.

First, if Huber's Φ is used, the c-value may be approximated where a specified observation is treated as an outlier whereas it passes the filter for marginally greater values. Provided that BRF is modified so that it is forced to always converge, this value is finite and calculable for all but for the first observations in the sample. Since the procedure relies on iteratively estimated parameters, this measure is non-trivial. To receive the interesting results at least, the BRF would have to be applied about 20 times, changing the tuning constant in steps of 0.1 from 1 to 3.

Second, the absolute difference of $|y_t - y_t^c|$ may be used as an indicator for outlier significance. This is computationally much easier but it only makes sense for observations corrected by the procedure. In our basic model differences give the following ranking:

MB	1983.1	.0125	(2.6)	GDP	1972.4	.0057	(2.2)
GDP	1974.4	.0098	(2.4)	GDP	1972.1	.0038	(2.2)
MB	1974.4	.0092	(2.5)	MB	1971.4	.0026	(2.1)

One would expect the observations ranked higher to be less influenced by alterations of the basic BRF specification if the measure were correct. This is approximately true, since only the MB 1971.4 is lost in some cases (one exception with the patchy outliers). Still, the accurateness of this naive measure remains doubtful. The numbers in brackets in the above table are the approximate c-values which imply a similar ordering in the example.

It may be contended that outlier measures should be constructed analogous to the concept of Cook(1977). However, Cook's statistics is interpreted as measuring the influence of one observation on the parameters of the identified structure. Since, in the case of BRF, the structure is identified iteratively the c of the first suggestion looks like a good equivalent to Cook's measure. Moreover, it is easily extended to different model specifications like e.g. ARMA or trended data models.

10. Summary and conclusions

Although the reported experiments are a useful tool to gain insight into the properties of the BRF, they can neither be regarded as definite proofs nor do they cover all directions. For instance, save for the IO case, the paper did not investigate into changes of the model assumptions or into changes of the autoregressive order. Therefore, conclusions have to be drawn with care only.

It seems that transformation of the data or, alternatively, adapting the model class covered by BRF specification to reach an agreement between data and model class is important. Or, to put it into other words, outliers must be the only deviation from the specified model class. Otherwise robust filtering will correct data at random. This goes for IO data as well as for trends or structural breaks or seasonal cycles.

Contrary to suggestions in the literature, redescending weighting functions, especially with narrow linear areas around zero, are inappropriate for robust filtering. It is suggested that the theoretically best functions will even be strictly increasing.

Generally, robust filtering represents a safe and sound method for outlier correction in the univariate as well as multivariate case. In econometrics, where forecasting is a primary goal, multivariate filtering methods could help substantially in the model identification process and consequently in prediction on the basis of the identified model and of the corrected data. Since the implementation of the procedure is easy, it may be said that it should be used in economic time series forecasts whenever outliers are likely to be present in the data.

References

Cook, R.D.(1977): Detection of influential observations in linear regression; Technometrics 19,15-18.

Fama, E.(1965):The Behavior of Stock Market Prices; Journal of Business 38,34-105.

Hampel, F.R.(1974): The influence curve and its role in robust estimation; JASA 69,383-393.

Heathcote, C.R. & A.H. Welsh(1983): The Robust Estimation of Autoregressive Processes by Functional Least Squares; J.Appl.Prob. 20,737-753.

Hogg, R.V.(1974): Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory; JASA 69, 348,909-923.

Hsiao, C.(1979): Autoregressive Modeling of Canadian Money and Income Data; JASA 74,367,553-560.

Huber, P.J.(1964): Robust Estimation of a Location Parameter; Annals of Mathematical Statistics 35,73-101.

Kitagawa, G. & H. Akaike(1982): A Quasi Bayesian Approach to Outliers Detection; Ann. Inst. of Statistical Mathematics 34,95-104.

Kleiner, B., R.D. Martin and D.J. Thomson(1979): Robust Estimation of Power Spectra; J.Royal Stat. Soc. B 41,3,313-351.

Martin, R.D. (1979): Approximate Conditional Mean Type Smoothers and Interpolators; in: Th. GASSER & M. ROSENBLATT(ed.): Smoothing Techniques for Curve Estimation, Springer, New York.

Martin, R.D. (1981): Robust Methods for Time Series; in: D.F.FINDLEY(ed.): Applied Time Series Analysis II, Academic Press, New York.

Sims, C.A.(1972): Money, income, and causality; American Economic Review 62,540-552.

Thomson, D.J.(1977): Spectrum Estimation Techniques for Characterization and Development of WT4 Waveguide-I; Bell Systems Technical Journal 56,1769-1815.

Note to members of the Institute for Advanced Studies

The procedures described in this paper have been implemented by the author on the Sperry-Univac computer of the institute. The univariate version of the filter may be called by executing ISA*RF.URF and the bivariate version by executing ISA*RF.BRF. The programs are interactive and ask for further specification elements via terminal input. The data, however, have to be supplied on system file 8. Further information about the routines is given on ISA*RF.INFO. Comments about the performance of the programs are welcome.