

Robust Estimation of the Vector Autoregressive Model by a Least Trimmed Squares procedure

Christophe Croux¹ and Kristel Joossens¹

Faculty of Business and Economics , Katholieke Universiteit Leuven
Naamsestraat 69, B-3000 Leuven, Belgium, *christophe.croux@econ.kuleuven.be*

Abstract. The vector autoregressive model is very popular for modeling multiple time series. Estimation of its parameters is typically done by a least squares procedure. However, this estimation method is unreliable when outliers are present in the data, and therefore we propose to estimate the vector autoregressive model by using a multivariate least trimmed squares estimator. We also show how the order of the autoregressive model can be determined in a robust way. The robust procedure is illustrated on a real data set.

Keywords: Robustness, Multivariate Time Series, Outliers, Trimming, Vector Autoregressive Models

1 Introduction

The use of autoregressive models for predicting and modelling univariate time series is standard and well known. In many applications, one does not observe a single time series, but several series, possibly interacting with each other. For these multiple time series the vector autoregressive model became very popular, and is described in standard textbooks on time series (e.g. Brockwell and Davis 2003, Chapter 7). In this paper we propose a robust procedure to estimate vector autoregressive models and to select their order.

Let $\{y_t \mid t \in \mathbb{Z}\}$ be a p -dimensional stationary time series. The vector autoregressive model of order k , denoted by VAR(k), is given by

$$y_t = \mathcal{B}'_0 + \mathcal{B}'_1 y_{t-1} + \dots + \mathcal{B}'_k y_{t-k} + \varepsilon_t, \quad (1)$$

with y_t a p -dimensional vector, the intercept parameter \mathcal{B}'_0 a vector in \mathbb{R}^p and the slope parameters $\mathcal{B}_1, \dots, \mathcal{B}_k$ being matrices in $\mathbb{R}^{p \times p}$. Throughout the paper M' will stand for the transpose of a matrix M . The p -dimensional error terms ε_t are supposed to be independently and identically distributed with a density of the form

$$f_{\varepsilon_t}(u) = \frac{g(u' \Sigma^{-1} u)}{(\det \Sigma)^{1/2}}, \quad (2)$$

with Σ a positive definite matrix, called the *scatter matrix* and g a positive function. If the second moment of ε_t exists, Σ will be (proportional

to) the covariance matrix of the error terms. Existence of a second moment, however, will not be required for the robust estimator. We focus on the unrestricted VAR(k) model, where no restrictions are put on the parameters $\mathcal{B}_0, \mathcal{B}_1, \dots, \mathcal{B}_k$.

Suppose that the multivariate time series y_t is observed for $t = 1, \dots, T$. The vector autoregressive model (1) can be rewritten as a multivariate regression model

$$y_t = \mathcal{B}'x_t + \varepsilon_t, \quad (3)$$

for $t = k+1, \dots, T$ and with $x_t = (1, y'_{t-1}, \dots, y'_{t-k})' \in \mathbb{R}^q$, where $q = pk + 1$. The matrix $\mathcal{B} = (\mathcal{B}'_0, \mathcal{B}'_1, \dots, \mathcal{B}'_k)' \in \mathbb{R}^{q \times p}$ contains all unknown regression coefficients. In the language of regression, $X = (x_{k+1}, \dots, x_T)' \in \mathbb{R}^{n \times q}$ is the matrix containing the values of the explanatory variables and $Y = (y_{k+1}, \dots, y_T)' \in \mathbb{R}^{n \times p}$ the matrix of responses, where $n = T - k$. The classical least squares estimator for the regression parameter \mathcal{B} in (3) is given by the well known formula

$$\hat{\mathcal{B}}_{\text{OLS}} = (X'X)^{-1}X'Y,$$

and the scatter matrix Σ is estimated by

$$\hat{\Sigma}_{\text{OLS}} = \frac{1}{n-p}(Y - X\hat{\mathcal{B}}_{\text{OLS}})'(Y - X\hat{\mathcal{B}}_{\text{OLS}}). \quad (4)$$

In applied time series research, one is aware of the fact that outliers can seriously affect parameter estimates, model specification and forecasts based on the selected model. Outliers in time series can be of different nature, the most well known types being additive outliers and innovational outliers. With respect to the autoregressive model (1), an observation y_t is an additive outlier if only its own value has been affected by contamination. On the other hand, an outlier is said to be innovational if the error term ε_t in (1) is contaminated. Innovational outliers will therefore have an effect on the next observations as well, due to the dynamic structure in the series. Additive outliers have an isolated effect on the time series, but they still may seriously affect the parameter estimates.

Several procedures to detect different types of outliers for univariate time series have been proposed. For a detailed treatment of robust univariate time series analysis we refer to Maronna, Martin and Yohai (2006, Chapter 8). While most previous studies focus on a single series, this paper deals with robust analysis of multivariate time series.

A common practice for handling outliers in a multivariate process is to first apply univariate techniques to the component series in order to remove the outliers, followed by treating the adjusted series as outlier-free and model them jointly. But this procedure encounters several difficulties. First, in a multivariate process, contamination in one component may be caused by

an outlier in the other components. Secondly, a multivariate outlier cannot always be detected by looking at the component series separately, since it can be an outlier for the correlation structure only. Therefore it is better to cope with outliers in a multivariate framework. Tsay, Peña and Pankratz (2000) discuss the problem of multivariate outliers in detail.

The aim of this paper is to propose a robust estimation procedure for the vector autoregressive model, the most popular model for multiple time series analysis. Not much work has been done for the robust estimation of multivariate time series. Franses, Kloek and Lucas (1999) used Generalized M-estimators, which are known to have low robustness in higher dimensions. Another approach was taken by García Ben, Martínez and Yohai (1999), using so-called *Residual Autocovariance* (RA)-estimators, being an affine equivariant version of the estimators of Li and Hui (1989). García Ben et al. (1999) showed, by means of a simulation study, that the RA-estimators are resistant to outliers. Using an appropriate starting value, the RA-estimators are iteratively computed as solutions of certain estimating equations.

Our proposal for obtaining a resistant estimator for the VAR model is to replace the multivariate least squares estimator for (3) by a highly robust estimator. We will use the *Multivariate Least Trimmed Squares* (MLTS) estimator, discussed in Agulló, Croux and Van Aelst (2008). This estimator is defined by minimizing a trimmed sum of squared Mahalanobis distances, and can be computed by a fast algorithm. The procedure also provides a natural estimator for the scatter matrix of the residuals, which can then be used for model selection criteria. This estimator is reviewed in Section 2. The robustness of the estimator is studied by means of several simulation experiments in Section 3, where a comparison with the RA-estimators is also made. In Section 4 it is explained how to select the autoregressive order of the model in a robust way. The robust VAR methodology is applied on real data sets in Section 5, while Section 6 concludes.

2 The multivariate least trimmed squares estimator

The unknown parameters of the VAR(k) will be estimated via the multivariate regression model (3). For this the Multivariate Least Trimmed Squares estimator (MLTS), based on the idea of the Minimum Covariance Determinant estimator (Rousseeuw and Van Driessen 1999), is used. The MLTS selects the subset of h observations having the property that the determinant of the covariance matrix of its residuals from a least squares fit, solely based on this subset, is minimal.

Consider the data set $Z = \{(x_t, y_t), t = k + 1, \dots, T\} \subset \mathbb{R}^{p+q}$. Let $\mathcal{H} = \{H \subset \{k + 1, \dots, T\} \mid \#H = h\}$ be the collection of all subsets of size h . For any subset $H \in \mathcal{H}$, let $\hat{B}_{OLS}(H)$ be the classical least squares fit based on the observations of the subset:

$$\hat{B}_{OLS}(H) = (X'_H X_H)^{-1} X'_H Y_H,$$

where X_H and Y_H are submatrices of X and Y , consisting of the rows of X , respectively Y , having an index in H . The corresponding scatter matrix estimator computed from this subset is then

$$\hat{\Sigma}_{\text{OLS}}(H) = \frac{1}{h-p} (Y_H - X_H \hat{\mathcal{B}}_{\text{OLS}}(H))' (Y_H - X_H \hat{\mathcal{B}}_{\text{OLS}}(H)).$$

The MLTS estimator is now defined as

$$\hat{\mathcal{B}}_{\text{MLTS}}(Z) = \hat{\mathcal{B}}_{\text{OLS}}(\hat{H}) \quad \text{where} \quad \hat{H} = \underset{H \in \mathcal{H}}{\operatorname{argmin}} \det \hat{\Sigma}_{\text{OLS}}(H), \quad (5)$$

and the associated estimator of the scatter matrix of the error terms is given by

$$\hat{\Sigma}_{\text{MLTS}}(H) = c_\alpha \hat{\Sigma}_{\text{OLS}}(\hat{H}). \quad (6)$$

In definition (6), c_α is a correction factor to obtain consistent estimation of Σ at the model distribution (2) of the error terms, and α the trimming proportion for the MLTS estimator, i.e. $\alpha \approx 1 - h/n$. In the case of multivariate normal error terms it has been shown (e.g. Croux and Haesbroeck 1999) that $c_\alpha = (1 - \alpha)/F_{\chi_{p+2}^2}(q_\alpha)$. Here $F_{\chi_q^2}$ is the cumulative distribution function of a χ^2 distribution with q degrees of freedom, and $q_\alpha = \chi_{q,1-\alpha}^2$ is the upper α -quantile of this distribution.

Equivalent characterizations of the MLTS estimator are given by Agulló, Croux and Van Aelst (2008). They prove that any $\hat{\mathcal{B}} \in \mathbb{R}^{p \times q}$ minimizing the sum of the h smallest squared Mahalanobis distances of its residuals (subject to $\det \Sigma = 1$) is a solution of (5). In mathematical terms,

$$\hat{\mathcal{B}}_{\text{MLTS}} = \underset{\mathcal{B}, \Sigma; |\Sigma|=1}{\operatorname{argmin}} \sum_{s=1}^h d_{s:n}^2(\mathcal{B}, \Sigma).$$

Here $d_{1:n}(\mathcal{B}, \Sigma) \leq \dots \leq d_{n:n}(\mathcal{B}, \Sigma)$ is the ordered sequence of the residual Mahalanobis distances

$$d_s(\mathcal{B}, \Sigma) = ((y_t - \mathcal{B}'x_t)' \Sigma^{-1} (y_t - \mathcal{B}'x_t))^{1/2}, \quad (7)$$

for $\mathcal{B} \in \mathbb{R}^{p \times q}$. We see that the MLTS-estimator minimizes the sum of the h smallest squared distances of its residuals, and is therefore the multivariate extension of the Least Trimmed Squares (LTS) estimator of Rousseeuw (1984).

Since the efficiency of the MLTS estimator is rather low, the reweighted version is used in this paper, to improve the performance of MLTS. The Reweighted Multivariate Least Trimmed Squares (RMLTS) estimates are defined as

$$\hat{\mathcal{B}}_{\text{RMLTS}} = \hat{\mathcal{B}}_{\text{OLS}}(J) \quad \text{and} \quad \hat{\Sigma}_{\text{RMLTS}} = c_\delta \hat{\Sigma}_{\text{OLS}}(J), \quad (8)$$

where $J = \{j \in \{1, \dots, n\} \mid d_j^2(\hat{\mathcal{B}}_{\text{MLTS}}, \hat{\Sigma}_{\text{MLTS}}) \leq q_\delta\}$ and $q_\delta = \chi_{q,1-\delta}^2$. The idea is that outliers have large residuals with respect to the initial robust MLTS estimator, resulting in a large residual Mahalanobis distance

$d_j^2(\hat{\mathcal{B}}_{\text{MLTS}}, \hat{\Sigma}_{\text{MLTS}})$. If the latter is above the critical value q_δ , then the observation is flagged as an outlier. The final RMLTS is then based on those observations not having been detected as outliers. In this paper, we set $\delta = 0.01$ and take as trimming proportion for the initial MLTS estimator $\alpha = 25\%$.

3 Simulation experiments

In order to study the robustness of the estimators, we perform a simulation study comparing the OLS estimator with the robust RMLTS and the RA estimators. As in García Ben et al. (1999), RA estimators are computed as iteratively reweighted maximum likelihood estimates, with a Tukey Biweight weight function (tuned to have a 95% relative asymptotic efficiency for Gaussian innovations). Since this weight function is re-descending, it is important to use a robust starting value to ensure convergence to the “right” solution. In our implementation, the RMLTS was used as starting value.

We generate bivariate time series according to the VAR(2) model

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} .10 \\ .02 \end{pmatrix} + \begin{pmatrix} .40 & .03 \\ .04 & .20 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} .100 & .005 \\ .010 & .080 \end{pmatrix} \begin{pmatrix} y_{1,t-2} \\ y_{2,t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}, \quad (9)$$

where $\varepsilon_t \sim N_2(0, \Sigma)$ with

$$\Sigma = \begin{pmatrix} 1 & .2 \\ .2 & 1 \end{pmatrix}. \quad (10)$$

The aim is to look at the effect of the outliers on the parameter estimates. There are 10 regression parameters to be estimated, and to summarize the performance of the estimators, we calculate the total Bias and total Mean Squared Error (MSE). The former is computed as

$$\text{Bias} = \sqrt{\sum_{i=1}^q \sum_{j=1}^p \left(\frac{1}{\text{nsim}} \sum_{s=1}^{\text{nsim}} \hat{\mathcal{B}}_{ij}^s - \mathcal{B}_{ij} \right)^2} \approx \|E[\hat{\mathcal{B}} - \mathcal{B}]\|,$$

where $\hat{\mathcal{B}}^s$, for $s = 1, \dots, \text{nsim}$, is the estimate obtained from the s -th generated series, \mathcal{B} is the true parameter value and $\text{nsim} = 1000$ the number of simulations. The MSE is given by

$$\text{MSE} = \sum_{i=1}^q \sum_{j=1}^p \left[\frac{1}{\text{nsim}} \sum_{s=1}^{\text{nsim}} (\hat{\mathcal{B}}_{ij}^s - \mathcal{B}_{ij})^2 \right].$$

After generating series of length $T = 500$, according to model (9), m outliers will be introduced. The classical and robust estimators are used to estimate this VAR(2) model for the uncontaminated series ($m = 0$), and for the contaminated ones ($m > 0$), where several types of outliers are considered. Below

we look at the effect of additive, innovational, and correlation outliers on the different estimators. Note that other types of contamination do exist, like level shifts and patches of outliers.

Additive outliers are introduced by randomly selecting m bivariate observations, and contaminating them by adding the value 10 to all the components of the selected observations. We consider different contamination levels, ranging from one single outlier up to 5% of additive outliers, i.e. $m = 25$. The Bias and MSE for the OLS, RA and RMLTS estimator are given in Table 1, as a function of the number m of additive outliers.

Both Bias and MSE grow with the number of outliers, the increase being much faster for the non robust OLS. Using the robust estimators instead of OLS leads to a very small loss in efficiency when no outliers are present. When even only one outlier is present, the RA and RMLTS are already more efficient, and this decrease in MSE becomes very substantial for larger amounts of outliers. Comparing the robust procedures, RMLTS performs slightly better as RA in this simulation setting.

Table 1. Simulated Bias and Mean Squared Error for the OLS, and the robust RA and RMLTS estimator of a bivariate VAR(2) model, in presence of m additive outliers in a series of length 500.

m	OLS		RA		RMLTS	
	Bias	MSE	Bias	MSE	Bias	MSE
0	0.00	0.020	0.00	0.022	0.00	0.022
1	0.08	0.030	0.02	0.023	0.02	0.023
2	0.14	0.045	0.03	0.025	0.03	0.024
3	0.18	0.063	0.05	0.028	0.04	0.026
4	0.22	0.079	0.06	0.031	0.04	0.027
5	0.25	0.096	0.07	0.035	0.05	0.029
10	0.38	0.193	0.14	0.061	0.07	0.039
15	0.51	0.319	0.21	0.086	0.11	0.057
20	0.64	0.478	0.25	0.101	0.17	0.080
25	0.76	0.659	0.29	0.115	0.25	0.104

Innovational outliers are generated by first randomly selecting m innovation terms ε_t in (9). Then add the value 10 to the first component of the innovations, yielding the contaminated innovations series ε_t^C . Bivariate series are then simulated according to (9), but with ε_t replaced by ε_t^C . The Bias and MSE when estimating the uncontaminated ($m = 0$) and contaminated series are given in Table 2, for the classical as well as the robust estimation procedures.

The Bias and MSE for OLS grow for an increasing number of outliers, although at a smaller rate than for contamination with additive outliers. For the robust estimator we see a small decrease of the MSE, implying that the

Table 2. Simulated Bias and Mean Squared error for the OLS, and the robust RA and RMLTS estimator of a bivariate VAR(2) model, in presence of m innovational outliers in a series of length 500.

m	OLS		RA		RMLTS	
	Bias	MSE	Bias	MSE	Bias	MSE
0	0.00	0.021	0.00	0.022	0.00	0.022
1	0.02	0.022	0.00	0.021	0.00	0.021
2	0.04	0.023	0.01	0.020	0.01	0.020
3	0.06	0.025	0.01	0.019	0.01	0.019
4	0.08	0.029	0.01	0.018	0.01	0.018
5	0.10	0.033	0.01	0.018	0.01	0.018
10	0.20	0.068	0.01	0.017	0.01	0.017
15	0.30	0.123	0.01	0.016	0.01	0.016
20	0.40	0.198	0.01	0.016	0.01	0.016
25	0.49	0.289	0.01	0.017	0.01	0.016

robust procedure is precise in presence than in absence of innovational outliers! This is due to the fact that an innovational outlier in the time series results in a single vertical outlier, but also in several good leverage points when estimating the autoregressive model. The robust method can cope with the vertical outlier and takes profit of the good leverage points to decrease the MSE. The OLS estimator gets biased due to the vertical outliers, but the presence of the good leverage points explains why the effect of innovational outliers is less strong than for additive outliers. Finally, note that the difference between the two robust approaches is not significant here, showing again that RMLTS and RA perform very similarly. Hence, the RA method does neither improves, neither deteriorates the initial RMLTS estimate.

Correlation outliers are generated as innovational outliers, but instead of (10), we take

$$\Sigma = \begin{pmatrix} 1 & .9 \\ .9 & 1 \end{pmatrix} \quad (11)$$

and place the innovation outliers all at the same position $(2, -2)'$. By placing the outliers in this way, they are only outlying for the correlation structure, and not with respect to the marginal distributions of the innovations. This type of outliers strongly influences results of a (robust) univariate analysis. To illustrate this, we will estimate the VAR model (9) equation by equation, applying twice a univariate reweighted least trimmed squares estimator (RLTS) instead of the RMLTS. Bias and MSE when estimating the uncontaminated and contaminated series by OLS, the univariate RLTS and the multivariate RMLTS, are given in Table 3.

When no outliers are present, there is hardly any difference between the different estimation procedures: the robust procedures show only a marginal loss in MSE. From Table 3 one can see that the univariate RLTS yields a

Table 3. Simulated Bias and Mean Squared error for the OLS, robust univariate (RLTS) and multivariate (RMLTS) estimators of a bivariate VAR(2) model in presence of m correlation outliers in a series of length 500.

m	OLS		RLTS		RMLTS	
	Bias	MSE	Bias	MSE	Bias	MSE
0	0.01	0.084	0.01	0.098	0.01	0.093
1	0.01	0.074	0.01	0.088	0.01	0.083
2	0.02	0.069	0.02	0.083	0.01	0.076
3	0.02	0.056	0.02	0.074	0.01	0.069
4	0.02	0.054	0.03	0.067	0.01	0.062
5	0.03	0.046	0.03	0.065	0.01	0.059
10	0.06	0.046	0.06	0.054	0.01	0.044
15	0.08	0.043	0.08	0.049	0.01	0.037
20	0.11	0.044	0.11	0.048	0.01	0.032
25	0.14	0.049	0.14	0.053	0.01	0.030

comparable Bias as for OLS, growing for an increasing number of correlation outliers. On the other hand, the multivariate RMLTS approach offers protection against the correlation outliers, remaining almost without bias. As for the previous simulation scheme, the MSE tends to decrease with the number of outliers (because the latter introduce good leverage points). We conclude from this simulation experiment that a fully multivariate robust approach is necessary when estimating a VAR model.

4 Determining the autoregressive order

To select the order k of a vector autoregressive model, information criteria are computed for several values of k and an optimal order is selected by minimizing the criterion. Most information criteria are in terms of the value of the log likelihood l_k of the VAR(k) model. Using the model assumption (2) for the distribution of the error terms, we get

$$l_k = \sum_{t=k+1}^T g(\varepsilon_t' \Sigma^{-1} \varepsilon_t) - \frac{n}{2} \log \det \Sigma,$$

with $n = T - k$. When error terms are multivariate normal the above leads to

$$l_k = -\frac{n}{2} \log \det \Sigma - \frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{t=k+1}^T \varepsilon_t' \Sigma^{-1} \varepsilon_t. \quad (12)$$

The log likelihood will depend on the autoregressive order via the estimate of the covariance matrix of the residuals. For the ordinary least squares esti-

mator we have

$$\hat{\Sigma}_{\text{OLS}} = \frac{1}{n-p} \sum_{t=k+1}^T \hat{\varepsilon}_t(k) \hat{\varepsilon}_t'(k),$$

where the $\hat{\varepsilon}_t(k)$ are the residuals corresponding with the estimated VAR(k) model. Using trace properties, the last term in (12) equals the constant $-(n-p)p/2$ for the OLS estimator. To prevent that outliers might affect the optimal selection of the information criteria, we estimate Σ by the RMLTS estimator:

$$\hat{\Sigma}_{\text{RMLTS}} = \frac{c_\delta}{m(k) - p} \sum_{t \in J(k)} \hat{\varepsilon}_t(k) \hat{\varepsilon}_t'(k),$$

with $J(k)$ as in (8) and $m(k)$ the number of elements in $J(k)$. The last term in (12) equals now $-(m(k) - p)p/(2c_\delta)$.

The most popular information criteria to select the order of the autoregressive model are of the form

$$\frac{-2}{n} l_k + h(n) \frac{(kp+1)p}{n},$$

where $(kp+1)p$ is the number of unknown parameters, which penalizes for model complexity, and where $h(n)$ can take different forms. We will consider the following three criteria: the popular Akaike information criterion, corresponding to $h(n) = 2$, the Hannan-Quinn criterion, corresponding to $h(n) = 2 \log(\log(n))$ and the Schwarz criterion, also called the Bayesian Information Criterion, for which $h(n) = \log(n)$.

5 Example

As an example, we consider the bivariate time series of *maturity rates* (Tsay 2002, p. 324–325). The first series “GS1” is the 1-year Treasury constant maturity rate, and the second series “N3” is the 3-year Treasury constant maturity rate. The data are monthly and sampled from April 1953 to January 2001. As in the book of Tsay (2002), we work with the log-transformed version of both series. We consider the series as stationary. From the plot of the series (Figure 1), it can be seen that there might be some outliers around the years 1954 and 1958.

In Table 4 different lag length criteria, as discussed in Section 4, are presented, once based on the OLS estimator, and once based on the RMLTS. The information criteria clearly depend on the chosen estimator. For example, when using the AIC the classical method suggests a VAR(8) model while the robust indicates a VAR(6) model. On the other hand the Schwarz criterion selects an optimal order 3 for both estimators. Since it is well known that the latter criterion yields a consistent estimate of the optimal order we continue the analysis with $k = 3$.



Fig. 1. Time plot of the “maturity rate” series. The solid line represents the 1-Year Treasury constant maturity rate and the dashed line the 3-Year Treasury constant maturity rate, both in logs.

Table 4. Lag length criteria using the OLS and RMLTS estimator for the “maturity rate” series.

k	1	2	3	4	5	6	7	8
Based on OLS estimation								
AIC	-7.35	-7.58	-7.61	-7.62	-7.61	-7.60	-7.62	-7.62
HQ	-7.33	-7.55	-7.57	-7.57	-7.54	-7.53	-7.53	-7.52
SC	-7.306	-7.50	-7.51	-7.48	-7.44	-7.40	-7.39	-7.36
Based on RMLTS estimation								
AIC	-7.43	-7.62	-7.67	-7.69	-7.69	-7.74	-7.69	-7.71
HQ	-7.42	-7.59	-7.63	-7.64	-7.62	-7.67	-7.60	-7.61
SC	-7.39	-7.55	-7.57	-7.564	-7.52	-7.55	-7.46	-7.45

After estimating the VAR(3) model with the robust RMLTS estimator, the corresponding robust residual distances $d_t(\hat{\mathcal{B}}_{\text{RMLTS}}, \hat{\Sigma}_{\text{RMLTS}})$ are computed as in (7), for $t = k + 1, \dots, T$. Figure 2 displays these distances with respect to the time index, and high residual distances indicate outlying observations. It is important to compute these distances based on the robust RMLTS, in order to avoid the well-known masking effect. Furthermore, it is common to compare these distances with a critical value from the chi-square distribution with p degrees of freedom, and we took $\chi_{p,0.99}$. Figure 2 reveals that several suspectable high residuals are detected, in particular around the years 1954 and 1958. But there are also a couple of other, less extreme outliers, which are more difficult to retrieve from the time series plot in Fig-

ure 1. Due to the presence of outliers, it is appropriate to make use of robust methods for further analysis of this data set.

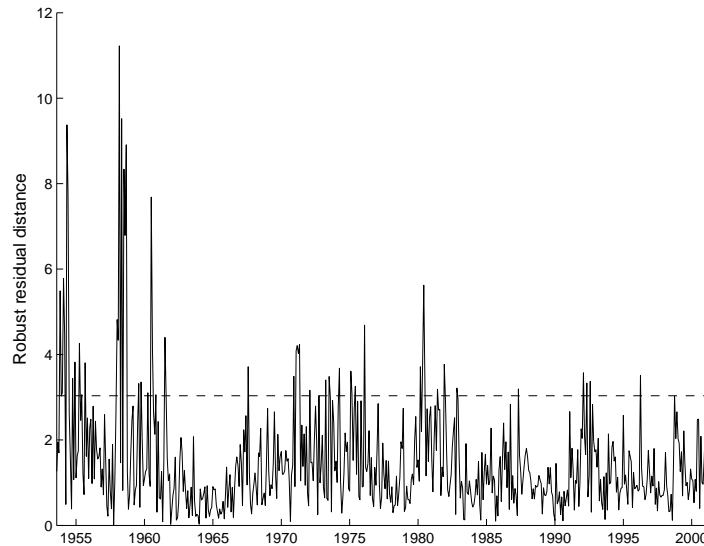


Fig. 2. Robust residual distances for the “maturity rate” series, based on RMLTS estimator of a VAR(3) model. The dashed line represents the critical value at the 1% level.

6 Conclusions

For multivariate time series correlation outliers can be present, which are not necessarily visible in plots of the single univariate series. Development of robust procedures for multiple time series analysis is therefore even more important than for univariate time series analysis.

In this paper we have shown how robust multivariate regression estimators can be used to estimate Vector Autoregressive models. We use the reweighted multivariate least trimmed squares estimator, but other robust multivariate regression estimators could be used as well. Software to compute the MLTS estimator is available at <http://www.econ.kuleuven.be/christophe.croux/public>.

The estimation of VAR models as multivariate regression models has one major disadvantage. A fraction ε of outliers in the original series can produce up to $k\varepsilon$ outliers for the regression model (1), due to the fact that k delayed versions of the time series are used as explanatory variables. Hence, if a robust regression estimator has a breakdown point of, for example, $1/2$, this reduces to $1/(2k)$ when estimating the VAR(k) model. To solve this problem of propagation of outliers, it has been proposed to first filter the series with

a robust filter, and then to apply a robust estimator on the robustly filtered data (see Bianco et al. 2001, Maronna et al. 2006). Other types of robust filters were proposed by Davies et al. (2004) and Fried et al. (2006). However, while robust filters are available for univariate series, multivariate versions have not been developed yet, up to our best knowledge, and we leave this for future research.

In the simulation experiments the RMLTS estimators have been compared with the residual autocovariance (RA) estimators of García Ben et al. (1999). The RA estimates are computed iteratively, and we propose to use the RMLTS as a starting value for computing the RA estimators. It turned out that both robust estimators behave then similarly. If there are no outliers in the data set present, the robust estimators perform almost as good as the classical estimator. But if there are outliers, bias and MSE only remain under control when using the robust estimator.

References

- AGULLÓ, J., CROUX, C., VAN AELST, S. (2008): The multivariate least trimmed squares estimator. *Journal of Multivariate Analysis* 99(3), 311–318.
- BIANCO, A. M., GARCA BEN, M., MARTNEZ, E. J., YOHAI, V. J. (2001): Outlier detection in regression models with ARIMA errors using robust estimates. *Journal of Forecasting* 20, 565–579.
- BROCKWELL, P. J., DAVIS, R. A. (2003): Introduction to Time Series and Forecasting. Wiley, New York.
- CROUX, C., HAESBROECK, G. (1999): Influence function and efficiency of the MCD-scatter matrix estimator. *Journal of Multivariate Analysis* 71, 161–190.
- DAVIES, P. L., FRIED, R., AND GATHER, U. (2004): Robust signal extraction for on-line monitoring data. *Journal of Statistical Planning and Inference* 122, 65–78.
- FRANSES, H. P., KLOEK, T., LUCAS, A. (1999): Outlier robust analysis of long-run marketing effects for weekly scanning data. *Journal of Econometrics* 89, 293–315.
- FRIED, R., BERNHOLT, T., GATHER, U. (2006): Repeated median and hybrid filters. *Computational Statistics and Data Analysis*, 50, 2313–2338.
- GARCA BEN, M., MARTNEZ, E. J., YOHAI, V. J. (1999): Robust estimation in vector autoregressive moving average models. *Journal of Time Series Analysis* 20, 381–399.
- LI, W. K., HUI, Y. V. (1989): Robust multiple time series modelling. *Biometrika* 76, 309–315.
- MARONNA, R. A., MARTIN, R. D., YOHAI, V. Y. (2006): *Robust Statistics: Theory and Practice*. Wiley, New York.
- ROUSSEEUW, P. J. (1984): Least median of squares regression. *Journal of the American Statistical Association* 79, 871–880.
- ROUSSEEUW, P. J., VAN DRIESSEN, K. (1999): A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41, 212–223.
- TSAY, R. S. (2002): *Analysis of Financial Time Series*. Wiley, New York.
- TSAY, R. S., PEÑA, D., PANKRATZ, A. E. (2000): Outliers in multivariate time series. *Biometrika* 87, 789–804.