Computational Statistics and Data Analysis 65 (2013) 68–79

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Robust estimation for vector autoregressive models

Nora Muler [a,*], Víctor J. Yohai [b]

[a] Department of Mathematics, University Torcuato di Tella, Argentina

[b] Department of Mathematics, School of Exact and Natural Sciences, University of Buenos Aires and CONICET, Argentina

ARTICLE INFO

Article history:
Received 18 May 2011
Received in revised form 4 February 2012
Accepted 14 February 2012
Available online 27 February 2012

Keywords:
Robust estimators
BMM-estimator
VAR models

ABSTRACT

A new class of robust estimators for VAR models is introduced. These estimators are an extension to the multivariate case of the MM-estimators based on a bounded innovation propagation AR model. They have a filtering mechanism that avoids the propagation of the effect of one outlier to the residuals of the subsequent periods. Besides, they are consistent and have the same asymptotic normal distribution as regular MM-estimators for VAR models. A Monte Carlo study shows that these estimators compare favorable with respect to other robust ones.

1. Introduction

Two main approaches have been proposed to handle outliers when estimating a vector ARMA (VARMA) model. One approach is to use diagnostics tools to detect different types of outliers (like additive, innovation, etc.) and then estimating the ARMA coefficients and the size of the outliers by maximum likelihood. We can mention the work of Galeano et al. (2006), who use projection pursuit methods to develop a method for detecting outliers in vector ARMA models. However, this type of procedure may fail when there are multiple outliers that mask each other.

A second approach is to use robust estimators. The advantage of robust estimators is that they are not much influenced by outliers even in the presence of masking. We can cite the work of García Ben et al. (1999) which generalize to VARMA models the RA-estimators for univariate ARMA models proposed by Bustos and Yohai (1986). We can also cite García Ben et al. (2001) who propose estimators based on a τ-scale for VAR models, and Croux and Joossens (2008) who define estimators based on a multivariate version of the least trimmed squares (LTS) criterion.

Most estimators for VAR models are based on residuals, and in the case of a regular VAR(p) model, the effect of one additive outlier at time t may affect not only the residual corresponding to period t, but also the residuals of p subsequent periods. This mechanism of propagation of the outlier effect is described for the case of univariate ARMA models in Chapter 8 of Maronna et al. (2006).

Two related approaches have been proposed to overcome this problem in the case of the univariate ARMA models. The first approach is to use a robust filter to compute the residuals and the second is to embed the class of ARMA models in a larger class that contains models where the propagation of the effect of the outliers is considerably reduced. For univariate ARMA models, robust estimators based on a robust filter have been proposed among others by Martin et al. (1983) and by Bianco et al. (1996). The second approach was introduced by Muler et al. (2009) for univariate ARMA models. A similar idea was used by Muler and Yohai (2008) for univariate GARCH models.

Boudt and Croux (2010) extended the results of Muler and Yohai (2008) to vector GARCH models. Croux et al. (2010) proposed a robust method for estimating the parameters of multivariate exponential smoothing models based on data

* Correspondence to: Departamento de Matemáticas, Universidad Torcuato di Tella, Miñones 2177, Buenos Aires, Argentina.
E-mail address: nmuler@utdt.edu (N. Muler).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.02.011

cleaning. However, there is no proposal for VARMA models aimed at reducing the propagation of the effect of outliers. In this paper we extend the procedure proposed by Muler et al. (2009) to VAR models. An extensive Monte Carlo study shows that this procedure compares favorably with other robust procedures.

The paper is organized as follows. In Section 2 we introduce the bounded influence propagation VAR models. In Section 3 we define the proposed estimators based on these models. In Section 4 we study the asymptotic distribution of the proposed estimators. In Section 5 we discuss some issues related to the computation of the proposed estimators. In Section 6 we present the results of the Monte Carlo study, where we compare several estimators with the proposed robust estimators. In Section 7 we discuss the performance of the different estimators in an example with real data. In Section 8 we make some concluding remarks. The Appendix contains a proof.

2. Bounded innovation propagation models

Let $\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})$, $t = 1, \ldots, T$ be observations of a stationary order p vector autoregressive (VAR(p)) model. Then, we may write $\Phi(B)(\mathbf{y}_t - \mu) = \mathbf{u}_t$, where $\Phi(z)$ is an operator given by

$$\Phi(z) = I_m - \sum_{r=1}^{p} \phi_r z^r$$

and ϕ_r, $1 \leq r \leq p$ are $m \times m$ matrices. We assume that $\mathbf{u}_t$ is a multivariate white noise process of dimension m with probability distribution density $g(\mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u})$ where Σ is a $m \times m$ positive definite matrix. In order that the process $\mathbf{y}_t$ be stationary, the roots of $\det(\Phi(z))$ should lie outside the unit circle. Observe that when the first moment of $\mathbf{u}_t$ exists we have that $E(\mathbf{y}_t) = \mu$.

Let us define

$$\Lambda(B) = \Phi^{-1}(B) = I_m + \sum_{r=1}^{\infty} \lambda_r B^r, \tag{1}$$

where λ_r are $m \times m$ matrices. Then we have the following infinity order vector moving average (VMA(∞)) representation of the VAR process

$$\mathbf{y}_t = \mu + \mathbf{u}_t + \sum_{r=1}^{\infty} \lambda_r \mathbf{u}_{t-r}.$$

To control the propagation of the effect of one outlier on the subsequent residuals, we will use a class of models similar to the one proposed in Muler et al. (2009) for the univariate case. So, we define the family of bounded innovation propagation VAR (BIP–VAR) models by processes $\mathbf{y}_t$ satisfying

$$\mathbf{y}_t = \mu + \mathbf{u}_t + \sum_{r=1}^{\infty} w\left(\frac{M(\mathbf{u}_{t-r}, \Sigma)}{k}\right) \lambda_r \mathbf{u}_{t-r},$$

where $\mathbf{u}_t$ is as in the VAR model, $M(\mathbf{u}, \Sigma)$ is the Mahalanobis distance $M(\mathbf{u}, \Sigma) = (\mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u})^{1/2}$, the λ_r's are defined as in (1), k is a positive parameter and $w : (0, \infty) \to (0, \infty)$ is a non-increasing continuous weight function which penalizes large $\mathbf{u}_t$'s. In fact we will assume that there exists $K > 0$ such that for all $x > 0$, $w(x) \leq K/x$. We will also assume that there exists k_0 such that

$$w(x) = 1 \quad \text{if } 0 \leq x \leq k_0, \tag{2}$$

that is, the function w gives constant weight to innovations that are not contaminated by outliers. An example of a w function is given in Section 6. The parameter k controls the degree of influence of large innovations over the process $\mathbf{y}_t$. In this model the lag effect of a large innovation in period t has a bounded effect on y_{t+j} for any $j \geq 0$ and this effect will almost disappear in a few periods. When k increases, the influence of a large value of $\mathbf{u}_{t-i}$ on $\mathbf{y}_t$ will increase too. Note that when $k \to \infty$ we have that $w(M(\mathbf{u}_{t-r}, \Sigma)/k) \to 1$ and therefore this model approaches the regular VAR(p) model.

It is immediate to show that the proposed model can also be written as

$$\mathbf{y}_t = \mu + \mathbf{u}_t - w\left(\frac{M(\mathbf{u}_t, \Sigma)}{k}\right)\mathbf{u}_t + \Phi^{-1}(B)\left(w\left(\frac{M(\mathbf{u}_t, \Sigma)}{k}\right)\mathbf{u}_t\right)$$

and so

$$\Phi(B)(\mathbf{y}_t - \mu) = \Phi(B)\left(\mathbf{u}_t - w\left(\frac{M(\mathbf{u}_t, \Sigma)}{k}\right)\mathbf{u}_t\right) + w\left(\frac{M(\mathbf{u}_t, \Sigma)}{k}\right)\mathbf{u}_t.$$

Therefore,

$$\mathbf{y}_t - \sum_{r=1}^{p} \phi_r \mathbf{y}_{t-r} = \mu - \sum_{r=1}^{p} \phi_r \mu + \mathbf{u}_t - \sum_{r=1}^{p} \phi_r \left(\mathbf{u}_{t-r} - w\left(\frac{M(\mathbf{u}_{t-r}, \Sigma)}{k}\right)\mathbf{u}_{t-r}\right)$$

and then

$$\mathbf{y}_t = \mu - \sum_{r=1}^{p} \phi_r \mu + \sum_{r=1}^{p} \phi_r \left(\mathbf{y}_{t-r} - \mathbf{u}_{t-r} + w\left(\frac{M(\mathbf{u}_{t-r},\, \Sigma)}{k} \right) \mathbf{u}_{t-r} \right) + \mathbf{u}_t. \tag{3}$$

Hence we can write the innovations residuals of the BIP–VAR(p) model as

$$\mathbf{u}_t = \mathbf{y}_t - \mu + \sum_{r=1}^{p} \phi_r \mu - \sum_{r=1}^{p} \phi_r \left(\mathbf{y}_{t-r} - \mathbf{u}_{t-r} + w\left(\frac{M(\mathbf{u}_{t-r},\, \Sigma)}{k} \right) \mathbf{u}_{t-r} \right). \tag{4}$$

3. Bounded MM-estimators for VAR models

Huber (1981) defined the class of scale M-estimators. Given an univariate sample $\mathbf{v} = (v_1, \ldots, v_n)$, an M-estimator of scale $S(\mathbf{v})$ is defined by the value s satisfying

$$\frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{v_i}{s} \right) = \kappa, \tag{5}$$

where κ is generally chosen so that $\kappa = E_{H_0}(\rho(v))$, where H_0 is the nominal distribution of the v_i's. This choice implies that $s(v_1, \ldots, v_n)$ converges to 1 when $v_1, \ldots, v_n$ is a stationary and ergodic stochastic process with marginal distribution H_0. It will be assumed that the function ρ satisfies the following properties:

A The function ρ satisfies: (i) ρ is even, (ii) $0 \le v \le b^*$ implies $\rho(v) \le \rho(v^*)$, (iii) $\sup_v \rho(v) = 1$ and (iv) ρ is twice differentiable.

We can define two breakdown points: one is defined as the minimum fraction of inliers that takes this scale estimator to 0 and the other as the minimum fraction of outliers required to take the M-scale to infinity. Huber (1981) proves that the breakdown point to infinity of a scale M-estimator is $\epsilon^*_\infty = \kappa$ and the breakdown point to zero is $\epsilon^*_0 = 1 - \kappa$. Then, the breakdown point of this scale estimator is given by

$$\epsilon^* = \min(\kappa,\ 1 - \kappa). \tag{6}$$

Therefore, choosing ρ such that $\kappa = 0.5$ we obtain $\epsilon^* = 0.5$, which is the highest breakdown point for an equivariant scale estimator (see Huber (1981)).

Suppose that $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$, is a random sample satisfying the linear regression model $y_i = \theta' \mathbf{x}_i + \varepsilon_i$, where $y_i \in R$ is the response, $\mathbf{x}_i \in R^m$ are the regressor vectors and $\theta \in R^m$ the vector of regression coefficients. Rousseeuw and Yohai (1984) introduced the class of S-estimators which are defined as

$$\widehat{\theta} = \arg\min s(y_1 - \theta' \mathbf{x}_1, \ldots, y_n - \theta' \mathbf{x}_n),$$

where s is an M-scale. Rousseeuw and Yohai (1984) proved that the breakdown point of regression S-estimators is also given by (6). However, Hössjer (1992) showed that regression S-estimators cannot combine simultaneously a high breakdown point with a high efficiency under Gaussian errors. Yohai (1987) introduced the class of MM-estimators which have simultaneously these two properties. The MM-estimators are computed in two steps. In a first step an S-estimator $\widehat{\theta}_1$ with breakdown point 0.5 but with possibly low Gaussian efficiency is computed. In the second stage the MM-estimator is obtained as an M-estimator that uses as scale $s(y_1 - \widehat{\theta}_1' \mathbf{x}_1, \ldots, y_n - \widehat{\theta}_1' \mathbf{x}_n)$. Yohai and Zamar (1988) introduced another class of estimators that combine high breakdown point and high normal efficiency properties: the class of τ-estimators.

S-, τ-and MM-regression estimators were extended to multivariate regression models. Bilodeau and Duchesne (2000) and Van Aelst and Willems (2005), extended S-estimators, García Ben et al. (2006) the τ-estimators and Kudraszow and Maronna (2010) the MM-estimators. These three classes of estimators for multivariate regression can be adapted to fit a VAR(p) model to a given a vector series $\mathbf{y}_t$. For this purpose is enough to take as regressors all the components of $\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-p}$. However these estimators do not include any mechanism to avoid the propagation of the effect of outliers when computing residuals.

The MM-estimators for VAR models are derived from the MM-estimators for multivariate regression. Let $\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})$, $t = 1, \ldots, T$ be a set of observations which follow a VAR(p) model. Put $\mathbf{\Phi} = (\phi_1, \ldots, \phi_p)$, and $\beta = (\mu, \mathbf{\Phi})$. Let ρ_1 and ρ_2 be two functions satisfying property **A** and $\rho_2 \le \rho_1$. For a given β, let $\overline{\mathbf{u}}_t(\beta)$, $t = p+1, \ldots, T$ be the innovation residuals of the VAR(p) model defined by

$$\overline{\mathbf{u}}_t(\beta) = \mathbf{y}_t - \mu - \sum_{r=1}^{p} \phi_r (\mathbf{y}_{t-r} - \mu). \tag{7}$$

We will assume that

$$E(\rho_1(z^{1/2})) = 0.5, \tag{8}$$

where z has a chi-squared distribution with m degrees of freedom. Then the MM-estimators for the VAR(p) model are defined by the following two steps:

Step 1. S-estimators $\left(\widetilde{\beta}_T^S, \widetilde{\Sigma}_T^S\right)$ of (β, Σ) are computed by

$$\left(\widetilde{\beta}_T^S, \widetilde{\Sigma}_T^S\right) = \arg\min_{\beta,\Sigma} \det(\Sigma)$$

subject to

$$s(M(\overline{\mathbf{u}}_{p+1}(\beta), \Sigma), \ldots, M(\overline{\mathbf{u}}_T(\beta), \Sigma)) = 1, \tag{9}$$

where s is the M-scale estimator defined as in (5) with $\rho = \rho_1$ and $\kappa = 0.5$.

Step 2. The MM-estimator of the regular VAR(p) model is defined by

$$\widehat{\beta}_T^{\mathrm{MM}} = \arg\min_{\beta} \sum_{t=p+1}^{T} \rho_2\left(M(\overline{\mathbf{u}}_t(\beta), \widetilde{\Sigma}_T^S)\right). \tag{10}$$

Since we are using $\kappa = 0.5$, Eqs. (5), (8) and (9) guarantee the Fisher consistency of $\widetilde{\Sigma}_T^S$ when the innovations are multivariate normal. Note that $\mathbf{u}_t$ depends not only on $\mathbf{y}_t$ but also on $\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-p}$. Therefore when $\mathbf{y}_t$ is an outlier, its effect is propagated to $p+1$ innovation residuals: $\overline{\mathbf{u}}_t(\beta), \overline{\mathbf{u}}_{t+1}(\beta), \ldots, \overline{\mathbf{u}}_{t+p}(\beta)$. Since the S, τ-and MM-estimators depends on the $\overline{\mathbf{u}}_t(\beta)'s$, its degree of robustness is rather low, specially when p is large.

To overcome this problem we introduce a class of MM-estimators based on residuals computed using the BIP–VAR model. For a given $\beta = (\mu, \mathbf{\Phi})$, a positive definite matrix Σ and $k > 0$, according to (4), the innovation residuals for the BIP–VAR model are defined by

$$\begin{aligned}\widehat{\mathbf{u}}_t(\beta, \Sigma, k) = \mathbf{y}_t - \mu + \sum_{r=1}^{p} \phi_r \mu - \sum_{r=1}^{p} \phi_r \left(\mathbf{y}_{t-r} - \widehat{\mathbf{u}}_{t-r}(\beta, \Sigma, k)\right) \\ - \sum_{r=1}^{p} \phi_r \left(w\left(\frac{M(\widehat{\mathbf{u}}_{t-r}(\beta, \Sigma, k), \Sigma)}{k}\right) \widehat{\mathbf{u}}_{t-r}(\beta, \Sigma, k)\right).\end{aligned} \tag{11}$$

Then, given ρ_1, ρ_2 and κ, we define the bounded MM-(BMM-) estimator by the following three steps:

Step 1. We compute an S-estimator $\left(\widehat{\beta}_T^S, \widehat{\Sigma}_T^S\right)$ of (β, Σ) based on the residuals given in (11) by

$$\left(\widehat{\beta}_T^S, \widehat{\Sigma}_T^S, \widehat{k}_T^S\right) = \arg\min_{\beta,\Sigma,k} \det(\Sigma) \tag{12}$$

subject to

$$s(M(\widehat{\mathbf{u}}_{p+1}(\beta, \Sigma, k), \Sigma), \ldots, M(\widehat{\mathbf{u}}_T(\beta, \Sigma, k), \Sigma)) = 1, \tag{13}$$

where s is the M-scale estimator defined as in (5) with $\rho = \rho_1$ and $\kappa = 0.5$.

Step 2. Define

$$\widehat{\beta}_{1,T} = \arg\min_{\beta} \sum_{t=p+1}^{T} \rho_2\left(M(\overline{\mathbf{u}}_t(\beta), \widehat{\Sigma}_T^S)\right) \tag{14}$$

and

$$\widehat{\beta}_{2,T} = \arg\min_{\beta} \sum_{t=p+1}^{T} \rho_2\left(M(\widehat{\mathbf{u}}_t(\beta, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S)\right), \tag{15}$$

where $\overline{\mathbf{u}}_t$ are the residuals corresponding to a regular VAR model which are calculated as in (7) and $\widehat{\mathbf{u}}_t$ are the filtered residuals corresponding to a BIP–VAR model which are calculated as in (11).

Step 3. Put

$$a_{1,T} = \sum_{t=p+1}^{T} \rho_2\left(M(\overline{\mathbf{u}}_{p+1}(\widehat{\beta}_{1,T}), \widehat{\Sigma}_T^S)\right), \qquad a_{2,T} = \sum_{t=p+1}^{T} \rho_2\left(M(\widehat{\mathbf{u}}_t(\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S)\right).$$

Then, the BMM-estimator $\widehat{\beta}_T^{\mathrm{BMM}}$ is defined by

$$\widehat{\beta}_T^{\mathrm{BMM}} = \begin{cases} \widehat{\beta}_{1,T} & \text{if } a_{1,T} \le a_{2,T} \\ \widehat{\beta}_{2,T} & \text{if } a_{1,T} > a_{2,T}. \end{cases} \tag{16}$$

Remark 1. Note that in (15) we compute the residuals using a BIP–VAR model with $k = 1$ instead of the estimated value $\widehat{k}_T^S$. The reason for this choice is that using $k = 1$, it is possible to show that $\widehat{\beta}_T^{\mathrm{BMM}}$ has the same asymptotic distribution than

$\widehat{\beta}_T^{\mathrm{MM}}$. Besides, we do not know the asymptotic distribution of $\widehat{\beta}_T^{\mathrm{BMM}}$ when we replace k by $\widehat{k}_T^S$. We choose w so that when $k = 1$ the corresponding estimator has a good behavior. For example, in our Monte Carlo study, we use the function given in (27). For this function the probability of bounding the innovation under a correct VAR model when $k = 1$ is 0.025.

Remark 2. Note that when the sample follows a VAR model without outliers, the definition of $\widehat{\beta}_{1,T}$ is based on the correct model. Then for large T we should have that $a_{1,T} < a_{2,T}$ and then $\widehat{\beta}_T^{\mathrm{BMM}} = \widehat{\beta}_{1,T}$. Instead, if the sample contains large additive outliers it is likely that $\widehat{\beta}_T^{\mathrm{BMM}} = \widehat{\beta}_{2,T}$ due to the fact that, as explained in Section 2, the propagation of the effect of outliers in the BIP–VAR model is less important than in the VAR model.

In the next proposition we give an equivalent way to compute $\left(\widehat{\beta}_T^S, \widehat{\Sigma}_T^S, \widehat{k}_T^S\right)$ solving an unconstrained optimization problem.

Proposition 1. *Let $(\overline{\Sigma}_T^S, \overline{\beta}_T^S)$ be defined by*

$$(\overline{\Sigma}_T^S, \overline{\beta}_T^S) = \arg\min_{\Sigma,\beta} D(\Sigma, \beta, 1), \tag{17}$$

where

$$D(\Sigma, \beta, k) = s^{2m}(M(\widehat{\mathbf{u}}_{p+1}(\beta, \Sigma, k), \Sigma), \ldots, M(\widehat{\mathbf{u}}_T(\beta, \Sigma, k), \Sigma)) \det(\Sigma).$$

Then we have $\widehat{\beta}_T^S = \overline{\beta}_T^S$,

$$\widehat{\Sigma}_T^S = s^2 \left(M(\widehat{\mathbf{u}}_{p+1}(\overline{\beta}_T^S, \overline{\Sigma}_T^S, 1), \overline{\Sigma}_T^S), \ldots, M(\widehat{\mathbf{u}}_T(\overline{\beta}_T^S, \overline{\Sigma}_T^S, 1), \overline{\Sigma}_T^S)\right) \overline{\Sigma}_T^S \tag{18}$$

and

$$\widehat{k}_T^S = \frac{1}{s(M(\widehat{\mathbf{u}}_{p+1}(\overline{\beta}_T^S, \overline{\Sigma}_T^S, 1), \overline{\Sigma}_T^S), \ldots, M(\widehat{\mathbf{u}}_T(\overline{\beta}_T^S, \overline{\Sigma}_T^S, 1), \overline{\Sigma}_T^S))}. \tag{19}$$

This proposition is proved in the Appendix.

Note that in the case that $\widehat{\beta}_T^{\mathrm{BMM}} = \widehat{\beta}_{2,T}$, the one step forecast for y_t according to (3) can be defined as

$$\begin{aligned}\widehat{\mathbf{y}}_t &= \widehat{\mu} - \sum_{r=1}^{p} \phi_r \widehat{\mu} + \sum_{r=1}^{p} \widehat{\phi}_r \left(\mathbf{y}_{t-i} - \widehat{\mathbf{u}}_{t-r}(\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1)\right) \\ &\quad + \sum_{r=1}^{p} \widehat{\phi}_r w\left(M(\widehat{\mathbf{u}}_{t-r}(\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S)\right) \widehat{\mathbf{u}}_{t-r}\left((\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S\right).\end{aligned} \tag{20}$$

This equation can also be written as

$$\widehat{\mathbf{y}}_t = \widehat{\mu} + \sum_{r=1}^{p} \widehat{\phi}_r(\mathbf{y}_t^c - \widehat{\mu}),$$

where $\mathbf{y}_t^c$ is the *cleaned* series defined by

$$\begin{aligned}\mathbf{y}_t^c &= \widehat{\mathbf{y}}_t + w\left(M(\widehat{\mathbf{u}}_t(\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S)\right) \widehat{\mathbf{u}}_t(\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1) \\ &= \mathbf{y}_t - \widehat{\mathbf{u}}_t(\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1) + w\left(M(\widehat{\mathbf{u}}_t(\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S)\right) \widehat{\mathbf{u}}_t(\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1) \\ &= w\left(M(\widehat{\mathbf{u}}_t(\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S)\right) \mathbf{y}_t + \left(1 - w\left(M(\widehat{\mathbf{u}}_t(\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S)\right)\right) \widehat{\mathbf{y}}_t.\end{aligned} \tag{21}$$

This shows that, $\mathbf{y}_t^c$ is a weighted average of the observed value and the predicted value. Note the similarity with the predicting equation for a multivariate exponential smoothing model (see for example Eq. (1) in Croux et al. (2010)). Observe that if $M(\widehat{\mathbf{u}}_t(\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S) \le k_0$, then according to (2), $w\left(M(\widehat{\mathbf{u}}_t(\widehat{\beta}_{2,T}, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S)\right) = 1$ and so $y_t^c = y_t$.

The functions ρ_1 and ρ_2 can be chosen for example in Tukey's bisquare family. This family of functions is defined as

$$\rho_c(x) = \begin{cases} \dfrac{3x^2}{c^2}\left(1 - \dfrac{x^2}{c^2} + \dfrac{x^4}{3c^4}\right) & \text{if } |x| \le c \\ 1 & \text{if } |x| > c \end{cases} \tag{22}$$

where $c > 0$ is any positive number. The choices of the constant c for determining ρ_1 and ρ_2 are discussed in Section 6.

Remark 3. The Mahalanobis distances $M(\widehat{\mathbf{u}}_t(\widehat{\beta}_T^{\mathrm{BMM}}, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S)$ can be used to detect outlying observations. Let χ_m^2 be the chi-squared distribution function with m degrees of freedom. Since when the $\mathbf{u}_t$'s are normal, the distribution of $M^2(\widehat{\mathbf{u}}_t(\widehat{\beta}_T^{\mathrm{BMM}}, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S)$ is approximated by χ_m^2, outliers can be identified as those observations such that $M^2(\widehat{\mathbf{u}}_t(\widehat{\beta}_T^{\mathrm{BMM}}, \widehat{\Sigma}_T^S, 1), \widehat{\Sigma}_T^S) \ge (\chi_m^2)^{-1}(1 - \alpha)$ where α is small, for example $\alpha = 0.025$.

4. Asymptotics

Let us consider observations $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T$ following a stationary VAR(p) process with parameter vector β and covariance matrix Σ. In this section we give the asymptotic distribution of the MM-and BMM-estimators of β. We use heuristic arguments to support our results rather than rigorous proofs.

Assume that ρ_1 and ρ_2 are twice continuously differentiable. Let $\widehat{\beta}_T^{\mathrm{MM}}$ be the MM-estimator defined in (10). Differentiating the right hand side of (10) it can be shown that $\widehat{\beta}_T^{\mathrm{MM}}$ should satisfy the following equation

$$\sum_{t=p+1}^{T} \frac{\psi_2\left(M\left(\overline{\mathbf{u}}_t'(\beta), \widetilde{\Sigma}_S^{-1}\right)\right)}{M\left(\overline{\mathbf{u}}_t'(\beta), \widetilde{\Sigma}_S^{-1}\right)} \mathbf{d}_t \overline{\mathbf{u}}_t'(\widehat{\beta}_T^{\mathrm{MM}}) = \mathbf{0},$$

where $\psi_2 = \rho_2'$ and $\mathbf{d}_t = (\mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}', \mathbf{1}_{1\times m})'$. Let s_0 be defined by

$$E\left(\rho_1\left(\frac{M(\mathbf{u}_t, \Sigma)}{s_0}\right)\right) = \kappa.$$

According to (8), $s_0 = 1$ when the $\mathbf{u}_t$'s are multivariate normal. Then, proceeding as in García Ben et al. (2001), it is possible to show that $\widetilde{\Sigma}_T^S$ converges to $s_0^2\Sigma$ in probability and that $T^{1/2}(\widehat{\beta}_T^{\mathrm{MM}} - \beta) \to^D N(0, V)$, where $\to^D$ denotes convergence in distribution and

$$V = \frac{s_0^2 m^2 E(\psi_2^2(v/s_0))}{E(v^2)\left(\mathbf{E}\left(\psi_2'(v/s_0) + (m-1)\, W(v/s_0)\right)\right)^2} C, \tag{23}$$

where $\psi_2 = \rho_2'$, $W(v) = \psi_2(v)/v$, $v = M(\mathbf{u}_t, \Sigma)$ and C is the asymptotic covariance matrix of the maximum likelihood (ML) estimator. This matrix C is given by

$$C = \frac{E(v^2)}{m}\begin{bmatrix} G^{-1}\otimes\Sigma & 0_{mp\times m} \\ 0_{m\times mp} & \left(I_m - \sum_{r=1}^{p}\phi_r\right)^2 \end{bmatrix},$$

where G is the $mp \times mp$ matrix defined as $G = E(\mathbf{z}_t'\mathbf{z}_t)$ where $\mathbf{z}_t = (\mathbf{y}_{t-1}' - \mu', \ldots, \mathbf{y}_{t-p}' - \mu')$. Then, the relative efficiency of the MM-estimator with respect to the maximum likelihood estimator is given by

$$EFF = \frac{E(v^2)\left(\mathbf{E}\left(\psi_2'(v/s_0) + (m-1)\, W(v/s_0)\right)\right)^2}{s_0^2 m^2 E(\psi_2^2(v/s_0))}. \tag{24}$$

According to Remark 2, we can expect that there exists T_0 such that $\widehat{\beta}_T^{\mathrm{BMM}} = \widehat{\beta}_{1,T}$ for $T \geq T_0$. Therefore the asymptotic distribution of $\widehat{\beta}_T^{\mathrm{BMM}}$ should be the same as that of $\widehat{\beta}_{1,T}$ A rigorous proof of this fact for the univariate case can be found in Muler et al. (2009). Moreover, the only difference between $\widehat{\beta}_{1,T}$ and $\widehat{\beta}_T^{\mathrm{MM}}$ is that the first estimator uses as covariance matrix $\widehat{\Sigma}_T^S$ and the second $\widetilde{\Sigma}_T^S$. Since both $\widehat{\Sigma}_T^S$ and $\widetilde{\Sigma}_T^S$ converge to $s_0^2\Sigma$ in probability, it can be proved that $\widehat{\beta}_{1,T}$ and $\widehat{\beta}_T^{\mathrm{MM}}$ have the same asymptotic distribution. Then $\widehat{\beta}_T^{\mathrm{BMM}}$ has the same asymptotic distribution that $\widehat{\beta}_T^{\mathrm{MM}}$, that is, $T^{1/2}(\widehat{\beta}_T^{\mathrm{BMM}} - \beta) \to^D N(0, V)$, where V is as in (23).

5. Computational aspects

We now define an algorithm to obtain an initial value to compute Step 1 of the MM- and BMM-estimators. Let $\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})$, $t = 1, \ldots, T$, be observations following a VAR(p) series with parameters $\mu = (\mu_1, \ldots, \mu_m)$ and $\phi = (\phi_1, \ldots, \phi_p)$. We can write the VAR($p$) process as a multivariate linear model as follows

$$\mathbf{y}_t = \sum_{r=1}^{p} \phi_r \mathbf{x}_{rt} + \mu + u_t, \quad t = p+1, \ldots, T, \tag{25}$$

where $\mathbf{x}_{rt} = \mathbf{y}_{t-r} - \mu$. We estimate μ by the coordinate-wise median $\widehat{\mu}$, then we define a first estimator of $\Phi = \left(\phi_1, \ldots, \phi_p\right)$ and Σ using the following subsampling procedure. Consider N random subsamples of size $h = mp + m$ of the dataset $(\mathbf{y}_t^*, \mathbf{x}_{1t}^*, \ldots, \mathbf{x}_{pt}^*)$, $t = p+1, \ldots, T$ where $\mathbf{y}_t^* = \mathbf{y}_t - \widehat{\mu}$ and $\mathbf{x}_{it}^* = \mathbf{y}_{t-i} - \widehat{\mu}$, $1 \leq i \leq p$. For each of these subsamples we estimate $\Phi = (\phi_1, \ldots, \phi_p)$ by ordinary least squares and compute the sample covariance matrix of the residuals. In this way we obtain N candidates $(\widetilde{\mathbf{\Phi}}_1, \widetilde{\Sigma}_1), \ldots, (\widetilde{\mathbf{\Phi}}_N, \widetilde{\Sigma}_N)$. We now define another set of N candidates $(\mathbf{\Phi}_1^*, \Sigma_1^*), \ldots, (\mathbf{\Phi}_N^*, \Sigma_N^*)$ using a concentration step similar to the one proposed by Rousseeuw and Van Driessen (1999). Let $\widetilde{\Phi}_j = (\widetilde{\phi}_1^j, \ldots, \widetilde{\phi}_p^j)$ and compute $\overline{\mathbf{u}}_t(\widehat{\mu}, \widetilde{\mathbf{\Phi}}_j)$, $p+1 \leq t \leq T$. Consider the subsample whose elements are the $[(T-p)/2]$ observations that have smallest Mahalanobis distances $M(\overline{\mathbf{u}}_t(\widehat{\mu}, \widetilde{\mathbf{\Phi}}_j), \widetilde{\Sigma}_j)$. We define Φ_j^* as the ordinary least squares estimator of these $[(T-p)/2]$

observations and Σ_j^* as an standardizing scalar times the sample covariance matrix of the corresponding $[(T-p)/2]$ residuals $\overline{\mathbf{u}}_t(\widehat{\mu}, \mathbf{\Phi}_j^*)$. The standardizing scalar is chosen so that this estimator is consistent under a multivariate normal model. It is easy to show that the scalar is equal to $(\chi^2_m)^{-1}(0.5)$.

Based on Proposition 1, an initial estimator for the S-estimator of the step 1 of the MM-estimator is defined by $(\widehat{\mu}, \mathbf{\Phi}_{j^*}^*, \Sigma_{j^*}^*)$, where

$$j^* = \arg\min_{1\le j\le N} s^{2m}\left(M\left(\overline{\mathbf{u}}_{p+1}(\mu, \mathbf{\Phi}_j^*), \Sigma_j^*\right), \ldots, \overline{\mathbf{u}}_{\mathbf{p+1}}(\mu, \mathbf{\Phi}_{\mathbf{j}}^*)\right) \det\left(\Sigma_j^*\right).$$

Then, since the VAR model can be written as a multivariate regression model and the S-estimators form a subclass of the τ-estimators, we can use the recursive algorithm proposed in García Ben et al. (2006) for τ-estimators of multivariate regression. As it is the case with most subsampling procedures, this initial estimator would not be largely affected by outliers if some of the subsamples do not contain outliers. This occurs with high probability when $1/(m(p+1))$ is larger than the fraction of contamination and the number of subsamples is large enough.

For the BMM-estimator the initial estimator for the S-estimator of step 1 is $(\widehat{\mu}, \mathbf{\Phi}_{j^*}^*, \Sigma_{j^*}^*)$, where

$$j^* = \arg\min_{1\le j\le N} s^{2m}\left(M\left(\widehat{\mathbf{u}}_{p+1}\left(\widehat{\mu}, \mathbf{\Phi}_j^*, \Sigma_j^*, 1\right), \Sigma_j^*\right), \ldots, M\left(\widehat{\mathbf{u}}_T\left(\widehat{\mu}, \mathbf{\Phi}_j^*, \Sigma_j^*, 1\right), \Sigma_j^*\right)\right) \det(\Sigma_j^*).$$

Using these initial values, the final estimators were computed using the function *fminsearch* of the MATLAB software. As it is usual when minimizing non-convex functions, the convergence of this algorithm to the global minimum depends on how close the initial estimator is to this minimum. In our case the Monte Carlo results show that for the cases that were considered, the solution obtained by the algorithm was satisfactory.

6. Monte Carlo results

We have performed a Monte Carlo study to compare the BMM-estimator with the ML- MM-, RA- and LTS-estimators. We have simulated two Gaussian stationary bivariate VAR(1) models and one Gaussian stationary bivariate VAR(2) model. We consider two cases: no outliers, and 10% of additive outliers. For each case the outliers are all equal and of additive type. Then if $\mathbf{y}_t$, $1 \le t \le T$, is a series that follows the VAR model, the series with additive outliers is obtained by replacing a given percentage of observations $\mathbf{y}_t$ equally spaced in the time by

$$\mathbf{y}_t + \mathbf{a}, \tag{26}$$

where $\mathbf{a}$ is a fixed value in R^m. In the case of the VAR(1) model we take samples of size $T = 100$ and for the VAR(2) model $T = 200$. The corresponding matrices Φ for the VAR(1) models (models 1 and 2) are

$$\Phi = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}, \qquad \Phi = \begin{pmatrix} 0.9 & 0 \\ -0.4 & 0.5 \end{pmatrix}$$

and the VAR(2) model (model 3) corresponds to

$$\phi_1 = \begin{pmatrix} 0.40 & 0.03 \\ 0.04 & 0.20 \end{pmatrix}, \qquad \phi_2 = \begin{pmatrix} 0.100 & 0.005 \\ 0.010 & 0.080 \end{pmatrix}.$$

Without loss of generality we took $\mu = (0, 0)'$ and $\Sigma = I_2$, where I_p the $p \times p$ identity matrix. For the three models The number of simulated replications for all the cases was $N = 500$.

The MM- and BMM-estimators are based on functions ρ_1 and ρ_2 in the bisquare family given in (22) with c equal to $c_1 = 2.66$ and $c_2 = 3.94$ respectively. The value of c_1 was chosen so that (8) holds. This condition is required if we want the scale S to have breakdown point 0.5 and $\widehat{\Sigma}_T^S$ to be consistent for multivariate normal innovations. The constant c_2 was calculated using (24) so that the MM-estimator has an efficiency of 85% when the $\mathbf{u}_t$'s have a multivariate normal distribution. Larger values of c_2 increase the efficiency but decrease the robustness of the estimators, and the opposite occurs when this constant decreases. We found that the chosen value for c_2 provides a good trade-off between robustness and efficiency.

To compare different estimators, we introduce a measure of the predictive power of an estimator $\widehat{\beta}$ of the parameter $\beta = (\mu, \Phi)$ of a VAR(p) model. Let $\mathbf{y}_t$ be a process following this model with covariance matrix Σ. We have $\mathbf{y}_t = \mathbf{w}_t + \mathbf{u}_t$, where $\mathbf{w}_t = \mu + \phi_1(\mathbf{y}_{t-1} - \mu) + \cdots + \phi_p(\mathbf{y}_{t-p} - \mu)$. Since $\mathbf{u}_t$ is not predictable, we only consider the error for predicting $\mathbf{w}_t$. Suppose that we want to use $\beta^* = (\mu^*, \mathbf{\Phi}^*)$, where $\mathbf{\Phi}^* = (\phi_1^*, \ldots, \phi_p^*)$ to predict $\mathbf{w}_t$ given $\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-p}$. The predictor of $\mathbf{w}_t$ is

$$\widehat{\mathbf{w}}_t(\beta^*) = \mu^* + \phi_1^*(\mathbf{y}_{t-1} - \mu^*) + \cdots + \phi_p^*(\mathbf{y}_{t-p} - \mu^*)$$

and we can measure the goodness of this predictor by

$$g(\beta^*) = \det E\left[(\widehat{\mathbf{w}}_t(\beta^*) - \mathbf{w}_t)(\widehat{\mathbf{w}}_t(\beta^*) - \mathbf{w}_t)'\right].$$

Table 1
Relative MFE efficiencies.

Model	Estimator			
	RA	LTS	MM	BMM
1	0.85	0.84	0.83	0.84
2	0.89	0.83	0.80	0.80
3	0.86	0.82	0.79	0.78

Table 2
Computing time.

p	T		
	100	200	500
1	0.3	0.5	1.1
2	0.7	1.5	3.5
3	1.5	2.5	6.9
4	2.5	5.2	13.2
5	3.0	5.8	14.9

It is easy to show that

$$g(\beta^*) = \sum_{l=1}^{p}\sum_{r=1}^{p}\left(\phi_l^* - \phi_l\right)\Sigma^{(r-l)}\left(\phi_r^* - \phi_r\right)' + \left(\mu^* - \mu - \sum_{r=1}^{p}\phi_r^*\left(\mu^* - \mu\right)\right)\left(\mu - \overline{\mu} - \sum_{r=1}^{p}\phi_r^*\left(\mu^* - \mu\right)\right)',$$

where $\Sigma^{(k)} = E((y_t - \mu)(y_{t-k} - \mu)')$. Therefore $g(\beta^*)$ is minimized when $\beta^* = \beta$ and in that case $g(\beta) = \mathbf{0}$. Then we define, as measure of forecasting efficiency, *the mean forecasting error* (MFE) of an estimator $\widehat{\beta}$ as $\mathrm{MFE}(\widehat{\beta}) = E(g(\widehat{\beta}))$.

Suppose that the series contains additive outliers of the form (26). Since in all the models considered in the Monte Carlo simulation $\Sigma = I_2$, for all the studied estimators $\widehat{\beta}$, $\mathrm{MFE}(\widehat{\beta})$ depend on $\mathbf{a}$ only through $\|\mathbf{a}\|$. Then, without loss of generality we take $\mathbf{a} = (b, b)$. We take a grid of values of $b : b = 0.5i$, $0 \le i \le 20$.

For each of the estimator $\widehat{\beta}$ studied in the Monte Carlo study, we estimate $\mathrm{MFE}(\widehat{\beta})$ by

$$\widehat{\mathrm{MFE}}(\widehat{\beta}) = \frac{1}{N}\sum_{i=1}^{N} g\left(\widehat{\beta}^{(i)}\right),$$

where $\widehat{\beta}^{(i)}$ is the value of $\widehat{\beta}$ corresponding to the i-th sample. We took as w the function defined by

$$w(x) = \begin{cases} 1 & \text{if } x \le k_0 \\ 1 - \dfrac{x - k_0}{l_0 - k_0} & \text{if } k_0 < x \le l_0 \\ 0 & \text{if } x > l_0. \end{cases} \tag{27}$$

We choose as k_0^2 the quantile 0.975 of the chi-square distribution with two degrees of freedom and as l_0^2 the quantile 0.995 of this distribution.

In Table 1 we show the relative efficiencies based on $\widehat{\mathrm{MFE}}$ of the different robust estimators with respect to the ML-estimator for the three models when there are not outliers.

We note that for model 1 all the estimators have a similar efficiency, close to 0.85. For models 2 and 3 the RA estimator is a little more efficient than the remaining robust estimators. In Fig. 1 we plot $\widehat{\mathrm{MFE}}$ as a function of the outlier size b, when the sample contains 10% of additive outliers. To avoid scale distortions, in the case of the ML estimator, we only plot the part of the curve corresponding to small values of b. We note that for the three models the robust methods behave much better than the ML-estimator and that the BMM-estimator performs better than the other robust estimators.

In Table 2 we show the approximate computing time (in minutes) of the BMM-estimator when fitting a bivariate VAR(p) model with 100, 200 and 500 observations and $1 \le p \le 5$. This times were obtained using a MATLAB code in a PC computer with an Intel Core Duo Processor E7500, 2.93 GHz.

7. An example with real data

In this example we consider two monthly series: the one year (GS1) and the three year (GS3) Treasury constant maturity rates in the period July 1991 to January 2009. Both series were obtained from the Economic Database of the Federal Reserve Bank of St Louis. The same series were used by Croux and Joossens (2008) at a different period. The series were transformed into stationary ones using first differences of logarithms. We denote by $\mathbf{y}_t = (y_{1t}, y_{2t})'$ $t = 1, \ldots, 211$ the vector of the two stationary series. We fit bivariate VAR(p) models using ML-, RA-, LTS-, MM- and BMM-estimators. In each case we choose

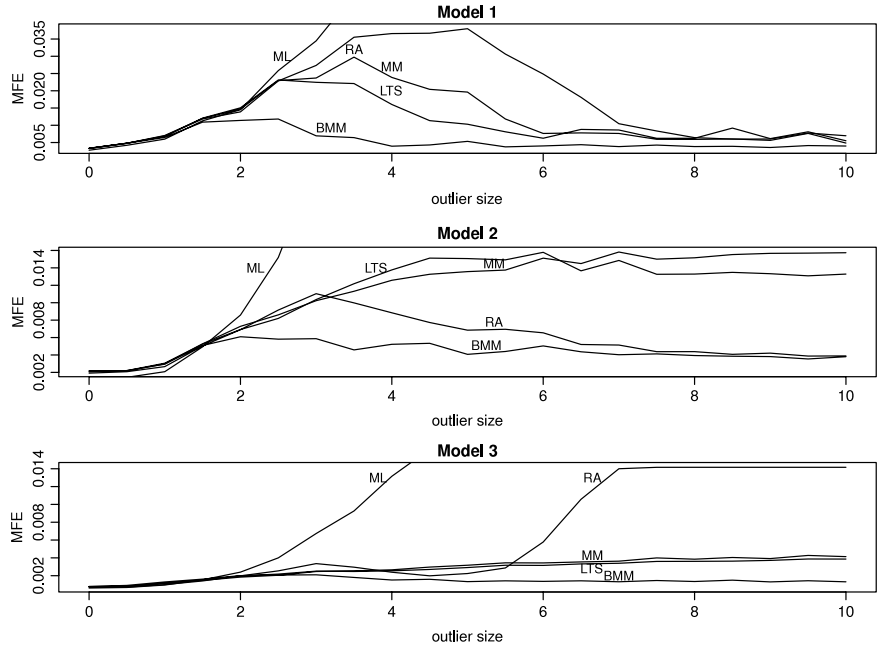

Fig. 1. Mean forecasting errors under outlier contamination.

Table 3
Parameter estimates for the example.

Parameter	Estimator				
	ML	RA	LTS	MM	BMM
ϕ_{11}	0.057	0.230	−0.058	−0.039	0.493
ϕ_{12}	0.584	0.235	0.532	0.474	0.0002
ϕ_{21}	−0.284	−0.127	−0.365	−0.372	−0.028
ϕ_{22}	0.685	0.532	0.774	0.758	0.429

the order of the model using the modified Akaike information criterion proposed by Croux and Joossens (2008). For all the estimators this criterion selected a VAR(1) model. In Table 3 we show the value of the five estimators.

In Fig. 2 we show the observed and cleaned values of y_{1t} and y_{2t}. The cleaned series, which were computed using (21), coincides with the observed ones in almost all periods. In the periods where both series differ, the cleaned values are marked with a small circle.

In Fig. 3 we compare the quantiles (up to the 0.8-quantile) of the absolute values of the residuals of the BMM-estimator of y_{1t} (plotted as a solid line) with those corresponding to the other four estimators. We observe that the quantiles corresponding to the BMM-estimator are smaller. As was expected, the ML-estimator is the one with the largest quantiles. The reason why we show up to the 0.8-quantile is to avoid that the large residuals to cause a scale distortion. A similar figure for y_{2t} does not show much differences among the estimators, and for this reason it is not included. The cause that explains why the BMM-estimator gives a better fit only for y_{1t} may be, as Fig. 2 show, that the series y_{1t} contains larger outliers than those in y_{2t}.

We have also simulated the performance of all the estimators studied in this section for the fitted model corresponding to the BMM-estimator. The estimated covariance matrix with this estimator is

$$\Sigma = \begin{pmatrix} 0.0031 & 0.0034 \\ 0.0034 & 0.0040 \end{pmatrix}.$$

We have performed 500 simulations using samples of size 100 with 10% of additive outliers of different sizes as it is done in Section 6. In Fig. 4 we plot $\widehat{\text{MFE}}$ as a function of the outlier size b. We note that for this model the BMM-estimator has a better overall behavior than the other estimators.

8. Concluding remarks

We have proposed a class of robust estimators for VAR models, the BMM-estimators, which include a mechanism that avoids the propagation of the effect of the outliers on the residuals. This is achieved by embedding the VAR models in the class of BIP–VAR models and computing an MM-estimator for this class. When the observations follow a VAR

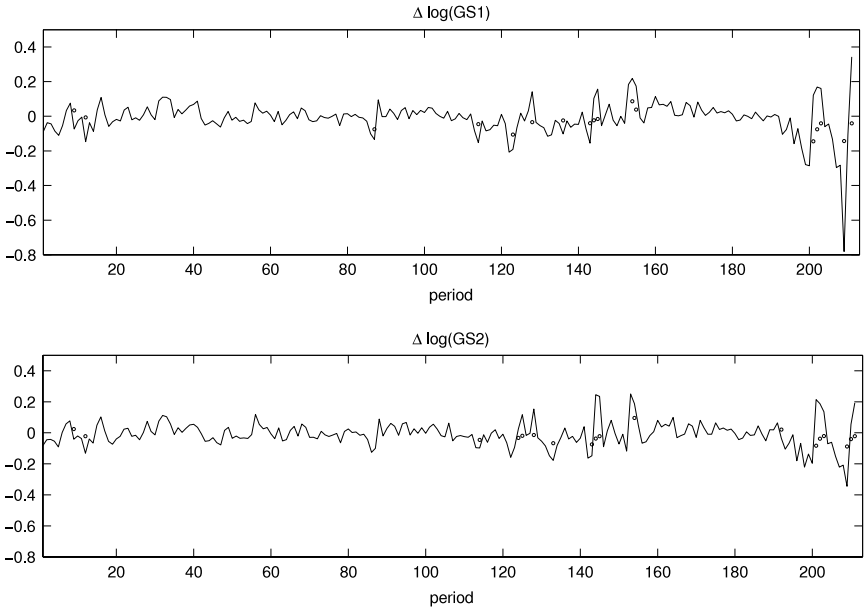

Fig. 2. Observed and cleaned values of Δ log GS1 and Δ log GS3.

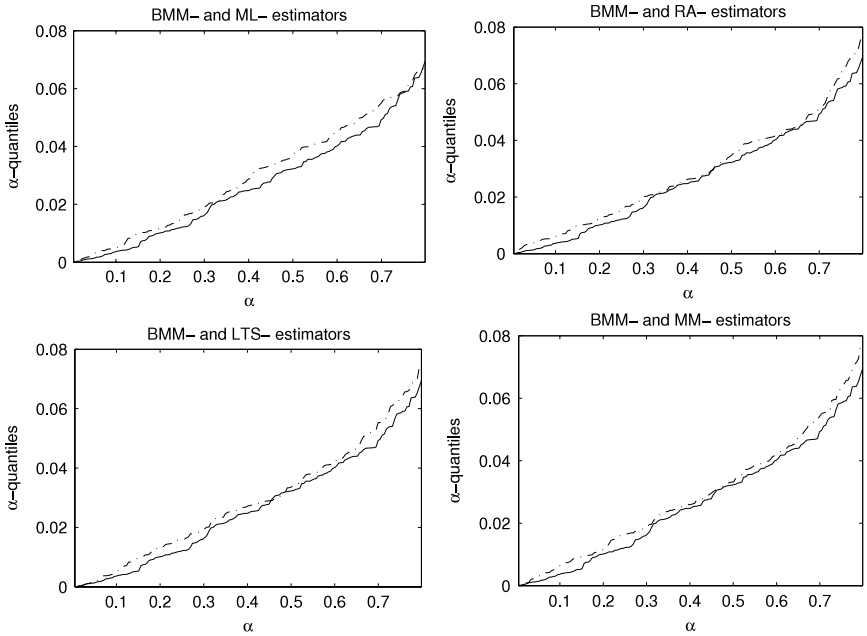

Fig. 3. Quantiles of the absolute values of the residuals of GS1.

model without outliers, the BMM-estimators have the same asymptotic distribution as regular MM-estimators. A Monte Carlo study shows that the BMM-estimators for VAR model compare favorably in efficiency and robustness with respect to other robust estimators. Extension of the BMM-estimators to VARMA models is also possible and a matter of future research.

Acknowledgments

We thank an anonymous referee and the Associate Editor for valuable insights and suggestions that largely contributed to the improvement of the paper. This research was supported in part by Grants 20020100100276 from Universidad of Buenos Aires, PID 112-200801-00216 from CONICET and PICT 00899 from ANPCYT, Argentina.

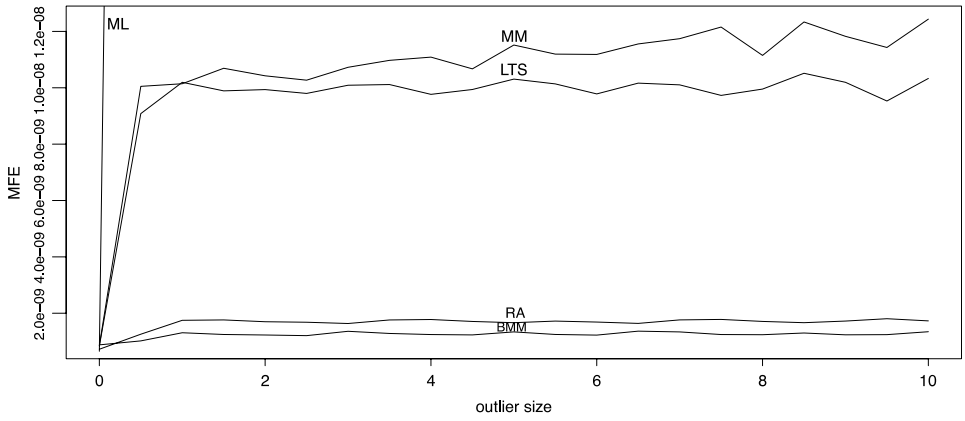

Fig. 4. Mean forecasting errors under outlier contamination for the fitted model.

Appendix

Proof of Proposition 1. It is immediate to show that for all $\lambda > 0$

$$D\left(\beta, \lambda^2 \Sigma, \frac{k}{\lambda}\right) = D(\beta, \Sigma, k), \tag{28}$$

and then

$$D(\widehat{\beta}_T^S, \widehat{\Sigma}_T^S, \widehat{k}_T^S) = D(\overline{\beta}_T^S, \overline{\Sigma}_T^S, 1). \tag{29}$$

We also have

$$s\left(M(\widehat{\mathbf{u}}_{p+1}(\widehat{\beta}_T^S, \widehat{\Sigma}_T^S, \widehat{k}_T^S), \widehat{\Sigma}_T^S), \ldots, M(\widehat{\mathbf{u}}_T(\widehat{\beta}_T^S, \widehat{\Sigma}_T^S, \widehat{k}_T^S), \widehat{\Sigma}_T^S)\right) = 1. \tag{30}$$

Take any (β, Σ, k) such that

$$s^2\left(M(\widehat{\mathbf{u}}_{p+1}(\beta, \Sigma, k), \Sigma), \ldots, M(\widehat{\mathbf{u}}_T(\beta, \Sigma, k), \Sigma)\right) = 1. \tag{31}$$

Then using (12) and (28)–(31) we have

$$\begin{aligned}\det(\Sigma) &= D(\beta, \Sigma, k) = D(\beta, k^2\Sigma, 1)\\ &\geq D(\overline{\beta}_T^S, \overline{\Sigma}_T^S, 1) = D(\widehat{\beta}_T^S, \widehat{\Sigma}_T^S, \widehat{k}_T^S) = \det(\widehat{\Sigma}_T^S),\end{aligned}$$

and this proves Proposition 1. □

References

Bianco, A.M., Garcia Ben, M., Martinez, E., Yohai, V.J., 1996. Robust procedures for regression models with ARIMA errors. In: Prat, Albert (Ed.), COMPSTAT 96, Proceeding of Computational Statistics. Physica-Verlag, Heidelberg, pp. 27–38.
Bilodeau, M., Duchesne, P., 2000. Robust estimation of the SUR model. Canadian Journal of Statistics 28, 277–288.
Boudt, K., Croux, C., 2010. Robust M-estimation of multivariate GARCH models. Computational Statistics and Data Analysis 54, 2459–2469.
Bustos, O.H., Yohai, V.J., 1986. Robust estimates for ARMA models. Journal of the American Statistical Association 81, 155–168.
Croux, C., Gelper, S., Mahieu, K., 2010. Robust exponential smoothing of multivariate time series. Computational Statistics & Data Analysis 54, 2999–3006.
Croux, C., Joossens, K., 2008. Robust estimation of the vector autoregressive model by a least trimmed squares procedure. In: COMPSTAT 2008: Proceedings in Computational Statistics, pp. 489–501.
Galeano, P., Peña, D., Tsay, R.S., 2006. Outlier detection in multivariate time series by projection pursuit. Journal of the American Statistical Association 101, 654–669.
García Ben, M., Martinez, E., Yohai, V.J., 1999. Robust estimation in vector ARMA models. Journal of Time Series 20, 381–399.
García Ben, M., Martínez, E., Yohai, V.J., 2006. Robust estimation for the multivariate linear model based on a τ-scale. Journal of Multivariate Analysis 97, 1600–1622.
García Ben, M., Villar, A.J., Yohai, V.J., 2001. Robust estimation in vector autoregressive models based on a robust scale. Estadística 53, 397–434.
Hössjer, O., 1992. On the optimality of S-estimators. Statistics and Probability Letters 14, 413–419.
Huber, P., 1981. Robust Statistics. Wiley, New York.
Kudraszow, N.L., Maronna, R.A., 2010. Estimates of MM type for the multivariate linear model. arXiv:1004.4883v2 [math.ST].
Maronna, R.A., Martin, R.D., Yohai, V.J., 2006. Robust Statistics: Theory and Methods. Wiley, West Sussex.
Martin, R.D., Samarov, A., Vandaele, W., 1983. Robust methods for ARIMA models. In: Zellner, E. (Ed.), Applied Time Series Analysis of Economic Data. Washington Bureau of the Census, pp. 153–177.
Muler, N., Peña, D., Yohai, V., 2009. Robust estimation for ARMA models. Annals of Statistics 37, 816–840.
Muler, N., Yohai, V.J., 2008. Robust estimates for GARCH processes. Journal of Statistical Planning and Inference 138, 2918–2940.

Rousseeuw, P.J., Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. Technometrics 41, 212–223.
Rousseeuw, P.J., Yohai, V.J., 1984. Robust regression by means of *S* estimators. In: Franke, J., Härdle, W., Martin, R.D. (Eds.), Robust and Nonlinear Time Series. In: Lecture Notes in Statistics, vol. 25. Springer Verlag, New York, pp. 256–272.
Van Aelst, S., Willems, G., 2005. Multivariate regression *S*-estimators for robust estimation and inference. Statistica Sinica 15, 981–1001.
Yohai, V.J., 1987. High breakdown point and high efficiency estimates for regression. Annals of Statistics 15, 642–656.
Yohai, V.J., Zamar, R.H., 1988. High breakdown-point estimates of regression by means of the minimization of an efficient scale. Journal of the American Statistical Association 83, 406–413.