Robust Estimation in Vector Autoregressive Models Based on a Robust Scale

By Marta García Ben, Universidad de Buenos Aires
Ana J. Villar, Universidad de Buenos Aires
and Víctor J. Yohai, Universidad de Buenos Aires and CONICET

Abstract. A new class of robust estimates for vector autoregressive processes is proposed. The autoregressive coefficients and the covariance matrix of the innovations are estimated simultaneously by minimizing the determinant of the covariance matrix estimate, subject to a constraint on a robust scale of the Mahalanobis norms of the innovation residuals. By choosing as robust scale a $\tau-$ estimate, the resulting estimates combine good robustness properties and asymptotic efficiency under Gaussian innovations. These estimates are asymptotically normal and in the case that the innovations have an elliptical distribution, their asymptotic covariance matrix differs only by a scalar factor from the one corresponding to the maximum likelihood estimate.

Keywords. Vector autoregressive models; multivariate time series; robust estimation, $\tau$-estimates

1 Introduction

Let $x_t = (x_{1t}, \ldots, x_{mt})'$ $(1 \leq t \leq T)$ be observations corresponding to a stationary vector autoregressive process (henceforth VAR($p$)). This means that there exist $m \times m$ matrices

$$\phi_r = (\phi_{r,ij})_{1 \leq i,j \leq m} \quad (r = 1, \ldots, p)$$

with $\phi_p \neq 0$ and $\nu \in \mathbb{R}^m$ such that

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \nu + a_t, \quad (1.1)$$
where \( a_t \) is a multivariate white noise process, i.e., the \( a_t \) are independent and identically distributed random vectors with

\[
E(a_t) = 0, \quad E(a_t a_t' ) = \Sigma,
\]

where \( \Sigma \) is a nonsingular \( m \times m \) matrix. Put

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

where \( I_m \) is the \( m \times m \) identity matrix. The stationarity of \( x_t \) requires that the roots of

\[
\text{det}\{\Phi(z)\} = 0,
\]

where \( \text{det}(A) \) denotes the determinant of \( A \), lie outside the unit circle \( |z| \leq 1 \).

The unknown parameters of model (1.1) are \( \phi_1, \ldots, \phi_p, \nu \) and \( \Sigma \).

Wilson (1973) proposed to estimate the parameters of this model by maximizing a normal likelihood conditionally on \((x_1, x_2, \ldots, x_p)\). We will call this estimate conditional maximum likelihood estimate (CMLE) and it is known that it has the same asymptotic normal distribution as the exact maximum likelihood estimate. Therefore the CMLE is asymptotically efficient under normal innovations. However, the CMLE is very sensitive to violations of the normality assumption and to the presence of a few atypical observations (outliers).

Let \( A \) be an \( m \times m \) non-singular matrix and \( b \in \mathbb{R}^m \); then the process \( x_t^* = Ax_t + b \) is also a VAR(\( p \)) process with parameters \( A\phi_i A^{-1} \) \((1 \leq i \leq p)\), \( A\nu + (I_m - A(\sum_{r=1}^{p} \phi_r) A^{-1})b \) and covariance matrix \( A\Sigma A' \). Then it is natural to require that estimates of the VAR(\( p \)) model parameters satisfy the following affine equivariance properties:

\[
\hat{\phi}_i\{(x_t^*)_{1 \leq t \leq T}\} = A\hat{\phi}_i\{(x_t)_{1 \leq t \leq T}\} A^{-1} \quad (1 \leq i \leq p),
\]

\[
A\hat{\nu}\{(x_t)_{1 \leq t \leq T}\} = A\hat{\nu}\{(x_t)_{1 \leq t \leq T}\} + (I_m - A(\sum_{r=1}^{p} \hat{\phi}_r\{(x_t)_{1 \leq t \leq T}\}) A^{-1})b,
\]

\[
A\hat{\Sigma}\{(x_t^*)_{1 \leq t \leq T}\} = A\hat{\Sigma}\{(x_t)_{1 \leq t \leq T}\} A'.
\]

Put \( z_t = (x_{t-1}^*, \ldots, x_{t-p}^*)' \) and \( \Omega = (\phi_1, \ldots, \phi_p, \nu) \). The dimension of \( z_t \) is \( r \times 1 \), where \( r = mp + 1 \) and the dimension of \( \Omega \) is \( m \times r \). Then we have

\[
x_t = \Omega z_t + a_t.
\]

Define the innovation residuals by

\[
\tilde{a}_t(\Omega) = x_t - \Omega z_t
\]
and the Mahalanobis norms of this residuals by
\[ d_t(\Omega, \Sigma) = (\hat{a}_t'(\Omega)\Sigma^{-1}\hat{a}_t(\Omega))^{1/2}. \]

It is well known that the CMLE of \( \Omega \) and \( \Sigma \) is given by
\[ (\hat{\Omega}, \hat{\Sigma}) = \arg \min_{\Omega, \Sigma} \det(\Sigma) \]
subject to
\[ \frac{1}{T-p} \sum_{t=p+1}^{T} d_t^2(\Omega, \Sigma) = m. \]

Given a sample \( u_1, \ldots, u_n \), let \( s(u_1, \ldots, u_n) \) be the square root of the mean squared error (SRMSE) scale estimate given by
\[ s(u_1, \ldots, u_n) = \left( \frac{1}{n} \sum_{i=1}^{n} u_i^2 \right)^{1/2}. \]

Then the constraint (1.3) can be also expressed as
\[ s^2(d_{p+1}(\Omega, \Sigma), \ldots, d_T(\Omega, \Sigma)) = m. \]

Let \( \omega_i \) be the \( i \)-th row of \( \Omega \) and \( \hat{\omega}_i \) its CMLE. Then \( \hat{\omega}_i \) is the least squares estimate of the regression model
\[ x_{ti} = \omega'_i z_t + a_{ti}, \quad p + 1 \leq i \leq T. \]

The covariance matrix estimate \( \hat{\Sigma} \) is the sample covariance of the \( \hat{a}_t(\hat{\Omega}) \)'s.

The CMLE is asymptotically efficient under Gaussian innovations. However, it is known that it is very sensitive to outliers and can be very inefficient even for small departures from normality. The only two classes of robust estimates proposed for VAR models are extensions of the RA—estimates (Bustos and Yohai (1986)). The first class is due to Li and Hui (1989), but these estimates are not affine equivariant. The estimates of the second class, introduced by García Ben, Martínez and Yohai (1999), are affine equivariant.

In this paper we define robust estimates of \( (\Omega, \Sigma) \) that are still of the form (1.2) and (1.5), but the SRMSE scale \( s \) given in (1.4) is replaced by a robust scale. More precisely, we propose to use a scale in the class of \( \tau \)-estimates introduced by Yohai and Zamar (1988). This approach is similar to the one used by Lopuhaä (1991) to estimate location and scatter matrix from multivariate data.
The proposed estimates, which will be called $\tau$-estimates for VAR models, are qualitatively robust for stationary VAR models, i.e., a small fraction of outliers does not change the asymptotic value of these estimates. See Hampel (1971) and Boente, Fraiman and Yohai (1987) for rigorous formalization of the concepts of qualitative robustness for stochastic processes. A Monte Carlo study also shows that they have good robustness behavior for small samples.

In Section 2 we define $\tau$–estimates for VAR processes and derive the asymptotic normal distribution of $\tau$–estimates and compute their relative efficiency with respect to the CMLE under normal innovations. In Section 3 we present a computing algorithm based on iterative weighted CMLE. In Section 4 we present the results of the Monte Carlo study. In the Appendix we derive some mathematical results.

2 Estimates based on a robust scale

A first class of robust scale estimates are the M-estimates which were introduced by Huber (1981). For a sample $u = (u_1, \ldots, u_n)$, an M-estimate of scale $s(u_1, \ldots, u_n)$ is defined by the value $s$ satisfying

$$
\frac{1}{n} \sum_{i=1}^{n} \rho \left( \frac{|u_i|}{s} \right) = \kappa,
$$

where $\kappa$ is chosen so that $\kappa = E_{H_0}(\rho(|u|))$, where $H_0$ is the nominal model for the $u_i$s. This choice guarantees that $s(u_1, \ldots, u_n)$ converges to 1 when $u_1, \ldots, u_n, \ldots$ is a stationary and ergodic stochastic process with marginal distribution $H_0$. We will require that the function $\rho$ satisfies the following properties:

A1 $\rho(0) = 0$.
A2 $0 \leq u \leq u^*$ implies $\rho(u) \leq \rho(u^*)$.
A3 $0 < A = \sup u \rho(u) < \infty$.
A4 $\rho$ is twice differentiable.

One measure of the degree of robustness of an estimate is the breakdown point introduced by Hampel (1971). Roughly speaking, the breakdown point of an estimate is the smallest fraction of outliers that is required to take the estimate to an extreme value. In the case of scale estimates, the two extreme
values are zero or infinity ("implosion" and "explosion", respectively). Huber (1981) proves that the breakdown point to infinity of an scale M-estimate is $c^*_\infty = \kappa/A$ and the breakdown point to zero is $c^*_0 = 1 - \kappa/A$. Then, the breakdown point of this scale estimate is given by

$$
\epsilon^* = \min \left( \frac{\kappa}{A}, 1 - \frac{\kappa}{A} \right). \tag{2.2}
$$

Therefore, choosing $\rho$ such that $\kappa/A = 0.5$ we obtain $\epsilon^* = 0.5$, which is the highest breakdown point for a scale-estimate (see Huber (1981)). However, (see Hossjer (1992)), the M-scale estimates can not combine a high breakdown point with high efficiency under normality. To combine both properties, Yohai and Zamar (1988) proposed $\tau-$estimates of scale.

Consider two functions $\rho_1$ and $\rho_2$ satisfying A1-A4 and put

$$
\kappa_i = E_{H_0}(\rho_i(|u|)), \ i = 1, 2. \tag{2.3}
$$

Let $s(u)$ be the M-scale estimate defined in (2.1) with $\rho = \rho_1$ and $\kappa = \kappa_1$. The $\tau$-estimate of scale of $u = (u_1, ..., u_n)$ is defined by

$$
\tau^2(u) = \frac{\tau_0^2}{\kappa_2} s^2(u) \frac{1}{n} \sum_{i=1}^{n} \rho_2 \left( \frac{|u_i|}{s(u)} \right). \tag{2.4}
$$

This estimate $\tau(u)$ converges to $\tau_0$ when $u_1, ..., u_n, ...$ is a stationary and ergodic process with marginal distribution $H_0$. Then, it is convenient to choose $\tau_0^2 = E_{H_0}(u^2)$, so that $\tau(u)$ converges to the same value as the SRMSE scale estimate.

Put $\psi_i(u) = \rho_i'(u)$, $i = 1, 2$. To guarantee consistency, we will also require that $\rho_2$ satisfies the following condition

$$
B \ 2\rho_2(u) - \psi_2(u)u > 0 \text{ for } u > 0.
$$

Yohai and Zamar (1988) proved that by properly choosing $\rho_1$, $\rho_2$, the $\tau$-estimates can combine breakdown point 0.5 with high Gaussian asymptotic efficiency. This is a consequence of the following facts.

- The scale $\tau$ differs from $s$ by a bounded factor. Based on this fact, Yohai and Zamar (1988) proved that a $\tau$-scale has the same breakdown point as the corresponding $s-$ scale.

- For any $\rho_1$, the choice of $\rho_2(u) = u^2$, and $\tau_0^2 = E_{H_0}(u^2)$ yields $\tau(u) = (1/n \sum_{i=1}^{n} u_i^2)^{1/2}$, and hence taking $\rho_2(u)$ close to $u^2$ but bounded makes $\tau$ robust and efficient for a normal distribution.
Possible choices of $\rho_1$ and $\rho_2$ are given later in this Section.

Let now $x_t = (x_{1t}, \ldots, x_{mt}), (1 \leq t \leq T)$ be observations corresponding to a VAR($p$) model given by (1.1). Then the $\tau-$estimates of $\Omega$ and $\Sigma$ are defined by

$$\left(\hat{\Omega}, \hat{\Sigma}\right) = \arg \min_{\Omega, \Sigma} \det(\Sigma)$$

subject to

$$\tau^2(d_{p+1}(\Omega, \Sigma), \ldots, d_T(\Omega, \Sigma)) = m,$$

where the scale $\tau$ is defined in (2.4) with $\tau^2 = m$.

Comparing (2.5) and (2.6) with (1.2) and (1.3), we observe that the CMLE is a particular case of $\tau-$estimate corresponding to $\rho_2(u) = u^2$.

It is easy to check that $\tau$-estimates are affine equivariant.

The following Theorem, proved in the Appendix, gives the equations for $\tau$-estimates. Define

$$d^*_t(\Omega, \Sigma) = \frac{d_t(\Omega, \Sigma)}{s(d_{p+1}(\Omega, \Sigma), \ldots, d_T(\Omega, \Sigma))}$$

and let $A_T = A_T(\Omega, \Sigma), B_T = B_T(\Omega, \Sigma)$ and $\psi^*_T = \psi_{T,\Omega\Sigma}$ be defined by

$$A_T(\Omega, \Sigma) = \frac{1}{T-p} \sum_{t=p+1}^{T} (2\rho_2(d^*_t(\Omega, \Sigma)) - \psi_2(d^*_t(\Omega, \Sigma))d^*_t(\Omega, \Sigma)),

B_T(\Omega, \Sigma) = \frac{1}{T-p} \sum_{t=p+1}^{T} (\psi_1(d^*_t(\Omega, \Sigma))d^*_t(\Omega, \Sigma)),

\psi^*_T(u) = A_T\psi_1(u) + B_T\psi_2(u)

and

$$w^*_T(u) = \frac{\psi^*_T(u)}{u}.$$

**Theorem 1.** Suppose that $\rho_1$ and $\rho_2$ are differentiable and put $\psi_1 = \rho'_1$. Then the $\tau-$estimates satisfy the following equations

$$\sum_{t=p+1}^{T} w^*_T(d^*_t(\hat{\Omega}, \hat{\Sigma}))\tilde{a}_t(\hat{\Omega})z'_t = 0$$

and

$$\hat{\Sigma} = \frac{\sum_{t=p+1}^{T} mw^*_T(d^*_t(\hat{\Omega}, \hat{\Sigma}))\tilde{a}_t(\hat{\Omega})\tilde{a}'_t(\hat{\Omega})}{s^2 \sum_{t=p+1}^{T} \psi^*_T(d^*_t(\Omega, \Sigma))d^*_t(\Omega, \Sigma)}.$$
where \( \tilde{s} = s(d_{p+1} (\tilde{\Omega}, \tilde{\Sigma}), ..., d_T (\tilde{\Omega}, \tilde{\Sigma})) \)

To determine the asymptotic distribution of \( \tau - \) estimates for VAR models, we will require that the innovations have an elliptical distribution with unimodal density. Then we consider the following assumption.

C. The distribution \( F \) of \( a_t \) has a density of the form

\[
f(a) = \frac{f^*(a')(\Sigma^{-1}a)}{\det(\Sigma)^{1/2}},
\]

where \( f^* \) is a strictly decreasing function.

An important family of unimodal elliptical distributions is the multivariate normal. In this case

\[
f^*(u) = \varphi^*(u) = \frac{\exp(-u/2)}{(2\pi)^{m/2}}.
\]

We will denote by \( (\Omega_0, \Sigma_0) \) the true values of \( (\Omega, \Sigma) \). We will calibrate the estimates so that \( \tilde{\Sigma} \) is consistent when \( f^* = f_0^* \), where \( f_0^* \) is a particular given function. For this purpose we take as \( H_0 \) in (2.3), the distribution of \( (a_0'\Sigma_0^{-1}a_0)^{1/2} \) under \( f_0^* \). Observe that this distribution does not depend on \( \Sigma \) in the elliptical case. We will take \( f_0^* = \varphi^* \), which implies that \( H_0 \) is the distribution of \( \sqrt{u} \), where \( u \) is a chi-square variable with \( m \) degrees of freedom. If \( f^* \neq f_0^* \), then \( \tilde{\Omega} \) is still a consistent estimate of \( \Omega \), but \( \tilde{\Sigma} \) will converge to \( \Sigma_0^* = \lambda_0 (H) \Sigma_0 \), where

\[
\lambda_0 (H) = \frac{k_0^2 (H)}{\sigma_0^2 (H)},
\]

and \( k_0 (H) \) and \( \sigma_0 (H) \) are defined by

\[
E_H \left( \rho_1 \left( \frac{u}{k_0 (H)} \right) \right) = \kappa_1
\]

and

\[
\frac{\sigma_0^2 (H)}{\kappa_2} E_H \left( \rho_2 \left( \frac{u}{k_0 (H)} \right) \right) = 1,
\]

where \( H \) is the distribution of \( (a_0^t \Sigma_0^{-1} a_0)^{1/2} \). Observe that \( \sigma_0 = k_0 = 1 \) if the \( a_t \) are elliptical and \( f^* = f_0^* \). In the Appendix we give an heuristic argument showing that in this case \( (\tilde{\Omega}, \tilde{\Sigma}) \) converges to \( (\Omega_0, \Sigma_0^*) \). More precisely, we show that \( (\Omega_0, \Sigma_0^*) \) satisfies the asymptotic form of equations (2.7), (2.8) and (2.6).
Theorem 2 establishes the asymptotic normality of $\tilde{\Omega}$. A sketch of its proof is given in the Appendix. We need to define the limit of the sequence of functions $\psi^*_T$. For that purpose put

$$A = A(H, \psi_1, \psi_2) = E_H \left( 2\rho_2 \left( \frac{u}{k_0} \right) - \psi_2 \left( \frac{u}{k_0} \right) \right),$$

$$B = B(H, \psi_1, \psi_2) = E_H \left( \psi_1 \left( \frac{u}{k_0} \right) \left( \frac{u}{k_0} \right) \right)$$

and

$$\psi^*_H(u) = A\psi_1 \left( \frac{u}{k_0} \right) + B\psi_2 \left( \frac{u}{k_0} \right), \quad w^*_H(u) = \frac{\psi^*_H(u)}{u}.$$

**Theorem 2.** Let $x_t, 1 \leq t \leq T$ be a VAR($p$) stationary process with parameters $\Omega_0$ and $\Sigma_0$. Suppose also that $\rho_i, i = 1, 2$ satisfy A1-A4 and B, and that the $a_t$ have a distribution satisfying C. Let $(\tilde{\Omega}_T, \Sigma_T)$ be the $\tau$-estimate corresponding to the first $T$ observations. Then $\tilde{\Omega}^*_T = T^{1/2}(\tilde{\Omega}_T - \Omega_0)$ converges in law to a normal distribution. The asymptotic covariance matrix of $\tilde{\Omega}^*_T$ is the same as the one of $\tilde{\Omega}_T = T^{1/2}(\tilde{\Omega}_T - \Omega_0)$, where $\tilde{\Omega}_T$ is the CMLE, multiplied by a constant $J(\psi_1, \psi_2, H)$ given by

$$J(\psi_1, \psi_2, H) = \frac{m^2 E_H(\psi^*_H(u))}{E_H(u^2)(E_H(\psi^*_H(u)) + (m-1)E_H(w^*_H(u)))^2},$$

where $H$ is the distribution of $(a'_iT\Sigma_0^{-1}a_t)^{1/2}$. Then the asymptotic relative efficiency of $\tilde{\Omega}_T$ with respect to $\tilde{\Omega}_T$ is $1/J(\psi_1, \psi_2, H)$.

Observe that the efficiency $1/J(\psi_1, \psi_2, H)$ depends on the dimension $m$ but not on the VAR model.

This suggests the following strategy to obtain a $\tau$-estimate which is simultaneously highly robust and highly efficient under normal innovations.

1. Choose $\rho_1$ so that

$$\kappa_1 / \max \rho_1 = 0.5. \quad (2.12)$$

This guarantees that the initial M-scale estimate $s$ has breakdown point 0.5.

2. Choose as $\rho_2(u)$ a bounded function close enough to $u^2$ to get the desired efficiency..
This can be achieved, for example, by taking $\rho_1$ and $\rho_2$ in Tukey’s bisquare family defined by

$$
\rho_{B,c}(u) = \begin{cases} 
\frac{u^2}{2} \left( 1 - \frac{u^2}{c^2} + \frac{u^4}{3c^4} \right) & \text{if } |u| \leq c \\
\frac{c^2}{6} & \text{if } |u| > c,
\end{cases}
$$

where $c$ is any positive number. Observe that when $c$ increases, $\rho_{B,c}$ approaches $u^2$.

Table 1 gives the values of $c_1$ such that $\rho_1 = \rho_{B,c_1}$ satisfies (2.12) for different values of $m$. Table 2 gives the value of $c_2$ to achieve different levels of asymptotic efficiency (taking $\rho_1 = \rho_{B,c_1}$ and $\rho_2 = \rho_{B,c_2}$).

3 Computing Algorithm

Consider the $\tau$-estimating equations (2.7) and (2.8) and suppose that we knew $\tilde{\Omega}$, $\tilde{\Sigma}$ and $\tilde{s} = s(d_{p+1}(\tilde{\Omega}, \tilde{\Sigma}), ..., d_T(\tilde{\Omega}, \tilde{\Sigma}))$. Then according to result (2.7) of Theorem 1 we have that each row of $\tilde{\Omega}$ can be obtained separately by weighted least squares (WLS), where the $t$-th observation has weight $w_\tau^T(d_t(\tilde{\Omega}, \tilde{\Sigma})/\tilde{s})$. This and (2.8) suggest the following iterative algorithm:

1. Using initial values $\tilde{\Omega}_0$ and $\tilde{\Sigma}_0$ satisfying (2.6), compute $\tilde{s}_0 = s_{T-p}(d_{p+1}(\tilde{\Omega}_0, \tilde{\Sigma}_0), ..., d_T(\tilde{\Omega}_0, \tilde{\Sigma}_0))$ and the weights $w_\tau^T(d_t(\tilde{\Omega}_0, \tilde{\Sigma}_0)/\tilde{s}_0)$ for $p + 1 \leq t \leq T$. Then these weights are used to compute each row of $\tilde{\Omega}_1$ separately by WLS.

We compute now $\tilde{s}_1 = s_{T-p}(d_{p+1}(\tilde{\Omega}_1, \tilde{\Sigma}_0), ..., d_T(\tilde{\Omega}_1, \tilde{\Sigma}_0))$.

2. We compute a matrix

$$
\tilde{\Sigma}_1^* = \frac{\sum_{t=p+1}^{T} mw_\tau^T \left( \frac{d_t(\tilde{\Omega}_1, \tilde{\Sigma}_0)}{s_t} \right) a_t(\tilde{\Omega}_1) a_t'(\tilde{\Omega}_1)}{\tilde{s}_1^2 \sum_{t=p+1}^{T} \psi_\tau^T \left( \frac{d_t(\tilde{\Omega}_1, \tilde{\Sigma}_0)}{s_t} \right) \frac{d_t(\tilde{\Omega}_1, \tilde{\Sigma}_0)}{s_t}}. \tag{3.1}
$$

3. We compute $\tilde{\tau}_1 = \tau(d_{p+1}(\tilde{\Omega}_1, \tilde{\Sigma}_1^*), ..., d_T(\tilde{\Omega}_1, \tilde{\Sigma}_1^*))$ and $\tilde{\Sigma}_1 = (\tilde{\tau}_1^2/m)\tilde{\Sigma}_1^*$. Then $(\tilde{\Omega}_1, \tilde{\Sigma}_1)$ satisfies constraint (2.6).
4. Suppose now that we have already computed \((\hat{\Omega}_h, \hat{\Sigma}_h)\) satisfying constraint (2.6). Then \((\hat{\Omega}_{h+1}, \hat{\Sigma}_{h+1})\) are computed using steps 1-3, but starting from \((\hat{\Omega}_h, \hat{\Sigma}_h)\) instead of \((\hat{\Omega}_0, \hat{\Sigma}_0)\).

5. The procedure is stopped in step \(h\) such that the relative absolute differences of all elements of the matrices \(\hat{\Omega}_h\) and \(\hat{\Omega}_{h-1}\) are smaller than a given value \(\delta\).

The convergence of this algorithm to the \(\tau\)-estimate has not been proved. In our simulations in Section 4, \((\hat{\Omega}_h, \hat{\Sigma}_h)\) always approach values satisfying (2.7) and (2.8).

4 Monte Carlo Results

To assess the efficiency and robustness of the proposed estimates we performed a Monte Carlo study. We considered two VAR(1) models of dimension 2. For the first model \(\phi_{11} = 0.9, \phi_{12} = 0.9, \phi_{21} = 0.9\), and \(\phi_{22} = 0.9\). For the second model \(\phi_{11} = 0.9, \phi_{12} = 0.9, \phi_{21} = -0.4\), and \(\phi_{22} = 0.5\). The innovations \(a_t\) were generated from a multivariate N(0, \(I_2\)) distribution.

For each model, three estimates were computed: the CMLE, a RA-estimate (see García Ben et al., 1999) and a \(\tau\)-estimate. Both RA- and \(\tau\)-estimates were based on the bisquare function, and the tuning constant were chosen so that the relative asymptotic efficiencies with respect to the CMLE be 0.90.

We considered the cases of no outliers and three types of contamination with 5% of replacement outliers. In the contaminated cases, instead of observing the VAR process \(x_t\), we observe the contaminated process \(\tilde{x}_t\), where \(\tilde{x}_t = x_t\) for 95% of the observations and \(\tilde{x}_t = d\) for the remaining 5%. The contaminated observations are fixed and equally spaced. Three values of \(d\) were considered: (1) \(d = (5, x_{t,2})\), (2) \(d = (5, 5)\) and (3) \(d = (5, -5)\).

The simulation was performed with 500 samples of size \(T = 100\). The \(\tau\)-estimates were computed using the algorithm described in Section 3, and the RA-estimates using the one described in Section 4 of García Ben et al. (1999). For both, RA- and \(\tau\)-estimates, the initial values were separate MM-estimate of regression (see Yohai (1987)) for each component with 90% efficiency. In Tables 3 and 4 we report the mean squared errors (MSE) of the three estimates and the relative efficiencies (RE) of the RA- and \(\tau\)-estimates with respect to the CMLE. The mean squared errors are displayed multiplied by 100. These results show that, when there are no outliers, the relative efficiencies of the RA- and \(\tau\)-estimates are always similar to the asymptotic
value 0.90. In the case of outlier contamination, the efficiency of the RA- and \( \tau \)-estimates is in general large. However, there are some isolated cases where the efficiency is less than one. The explanation for this unexpected improvement of the efficiency of the CMLE is that in these cases, the contaminated observations act as “good” leverage points. Consider for example Model 1 with contamination of type (1). In this case, the only contaminated variable is \( x_{t,1} \). Since in the equation \( x_{t,2} = \phi_{2,1}x_{t-1,1} + \phi_{2,2}x_{t-1,2} + \nu_2 + a_{t,2} \) the coefficient \( \phi_{2,1} = 0 \), the contaminated observations are good leverage points and this explains the improvement of the performance of the CMLE of \( \nu_2, \phi_{2,1} \), and \( \phi_{2,2} \).

**TABLE 3 ABOUT HERE**

**TABLE 4 ABOUT HERE**

For model 1, the \( \tau \)-estimates behaves better than the RA-estimates under the three types of contaminations. For model 2, the \( \tau \)-estimates behave better than the RA-estimates under contaminations of types 1 and 3. Instead, for contamination of type 2, the RA-estimates behaves better.

## 5 Conclusions

We have presented a new class of equivariant estimates for VAR models based on a robust scale: the \( \tau \)-estimates. These estimates depend on two functions \( \rho_1 \) and \( \rho_2 \), and by a proper choice of these functions, they combine high efficiency under normal errors and good robustness properties under outlier contamination. A Monte Carlo study results confirm these properties under normal errors and under isolated additive outliers. Further study is required to determine their robust behavior under other type of outliers as innovation outliers and patchy outliers.

### Appendix

**Proof of Theorem 1.** Put

\[
T(\Omega, \Sigma) = \text{det}(\Sigma)\tau^{2m}(d_{p+1}(\Omega, \Sigma), ..., d_T(\Omega, \Sigma)) ,
\]

then, it is easy to show that for any real \( \lambda \)

\[
T(\Omega, \lambda \Sigma) = T(\Omega, \Sigma).
\]
This shows that the $\tau$–estimate $(\hat{\Omega}, \hat{\Sigma})$ also minimizes $T(\Omega, \Sigma)$ without restrictions (observe that $(\hat{\Omega}, \lambda \hat{\Sigma})$ minimizes $T(\Omega, \Sigma)$, too), and therefore it should satisfy the following equations

$$\frac{\partial T(\Omega, \Sigma)}{\partial \Omega} = 0, \quad (A.1)$$

and

$$\frac{\partial T(\Omega, \Sigma)}{\partial \Sigma} = 0. \quad (A.2)$$

We will show that $A.1$ is equivalent to (2.7) and $A.2$ to (2.8).

Denote by $s^*(\Omega, \Sigma) = s(d_{p+1}(\Omega, \Sigma), \ldots, d_T(\Omega, \Sigma))$ and $R(\Omega, \Sigma) = (1/(T - p)) \sum_{t=p+1}^{T} \rho_2(d_t(\Omega, \Sigma)/s^*(\Omega, \Sigma))$. In the rest of the proof we will write $d_t$, $d_t^*$, $R$ and $s^*$ instead of $d_t(\Omega, \Sigma)$, $d_t^*(\Omega, \Sigma)$, $R(\Omega, \Sigma)$ and $s^*(\Omega, \Sigma)$ respectively. Finally put $\Omega = (\omega_{ij})$, then (A.1) is equivalent to

$$2s^* \frac{\partial s^*}{\omega_{ij}} R + s^2(\Omega, \Sigma) \frac{\partial R}{\omega_{ij}} = 0, \quad 1 \leq i \leq m, \quad 1 \leq j \leq r. \quad (A.3)$$

It is easy to show that

$$\frac{\partial s(d_{p+1}, \ldots, d_T)}{\partial d_t} = \psi_1(d_t/s(d_{p+1}, \ldots, d_T)) \frac{\sum_{h=p+1}^{T} \psi_1(d_h/s(d_{p+1}, \ldots, d_T))(d_h/s(d_{p+1}, \ldots, d_T))}{\sum_{h=p+1}^{T} \psi_1(d_h/s(d_{p+1}, \ldots, d_T))(d_h/s(d_{p+1}, \ldots, d_T))}$$

and that

$$\frac{\partial d_t}{\partial \omega_{ij}} = -\frac{\sum_{v=1}^{m} \sigma^{iv} a_{tv}(\Omega)}{d_t} z_{lj},$$

where $\Sigma^{-1} = (\sigma^{ij})$. Therefore

$$\frac{\partial s^*}{\partial \omega_{ij}} = -\frac{\sum_{t=p+1}^{T} (\psi_1(d_t^*)/d_t^*) \sum_{v=1}^{m} \sigma^{iv} z_{lj} a_{tv}(\Omega)}{s^* \sum_{t=p+1}^{T} \psi_1(d_t^*)d_t^*}. \quad (A.4)$$

We also have

$$\frac{\partial R}{\partial \omega_{ij}} = \frac{1}{T - p} \sum_{t=p+1}^{T} \psi_2(d_t^*) \left( \frac{\partial d_t}{\partial \omega_{ij}} s^* - d_t \frac{\partial s^*}{\partial \omega_{ij}} / s^{*2} \right). \quad (A.5)$$

Substituting (A.4) in (A.5) we get

$$\frac{\partial R}{\partial \omega_{ij}} = -\frac{\sum_{t=p+1}^{T} \psi_1(d_t^*)d_t^*}{(T - p) s^{*2} \sum_{t=p+1}^{T} \psi_1(d_t^*)d_t^*} \left( \sum_{t=p+1}^{T} (\psi_2(d_t^*)/d_t^*) \sum_{v=1}^{m} \sigma^{iv} z_{lj} a_{tv}(\Omega) \right).$$
\[
+ \left( \sum_{t=p+1}^T \psi_2(d_t^*) d_t^* \right) \left( \sum_{t=p+1}^T \psi_1(d_t^*) / d_t^* \right) \frac{1}{\sigma^*} \frac{\partial \sigma^*}{\partial \Sigma} \sum_{i=1}^m \sigma_{ij} w_{tj}^* \left( \frac{\partial^* t}{\partial^* t} \right).
\]

(A.6)

Substituting (A.4) and (A.6) in (A.3), we get

\[- \frac{m}{\sum_{i=1}^m \sigma_{ij} \sum_{t=p+1}^T \psi_1(d_t^*) d_t^*} \left( \frac{\partial^* t}{\partial^* t} \right) = 0, \quad 1 \leq i \leq m, \quad 1 \leq j \leq r\]

which is equivalent to

\[- \Sigma^{-1} \sum_{t=p+1}^T \psi_1(d_t^*) d_t^* \left( \frac{\partial^* t}{\partial^* t} \right) = 0, \quad \Sigma^{-1} \sum_{t=p+1}^T \psi_1(d_t^*) d_t^* = 0,\]

and this is equivalent to (2.7).

(A.2) is equivalent to

\[\frac{\partial L(\Omega, \Sigma)}{\partial \Sigma} = 0, \quad \text{(A.7)}\]

where

\[L(\Omega, \Sigma) = \log(T(\Omega, \Sigma)) = \log(\det(\Sigma)) + 2m \log(s^*) + m \log \left( \frac{1}{T - p} \sum_{t=p+1}^T \rho_2(d_t^*) \right). \quad \text{(A.8)}\]

It is easy to show that

\[\frac{\partial s^*}{\partial \Sigma} = \frac{1}{\sum_{t=p+1}^T \psi_1(d_t^*) \frac{\partial^* t}{\partial^* t}} \sum_{t=p+1}^T \psi_1(d_t^*) d_t^* \]

and since

\[\frac{\partial d_t}{\partial \Sigma} = - \frac{1}{2d_t} \Sigma^{-1} a_t d_t \Sigma^{-1}, \quad \text{(A.9)}\]

then

\[\frac{\partial s^*}{\partial \Sigma} = \frac{1}{2} \Sigma^{-1} \left( \sum_{t=p+1}^T \psi_1(d_t^*) d_t \right) \Sigma^{-1} = 0, \quad \text{(A.10)}\]

It is well known that, for a symmetric matrix \( \Sigma \)

\[\frac{\partial \log(\det(\Sigma))}{\partial \Sigma} = \Sigma^{-1}.\]

Differentiating (A.8) we get
\[
\frac{\partial L(\Omega, \Sigma)}{\partial \Sigma} = \Sigma^{-1} + \frac{2m}{s^*} \frac{\partial s^*}{\partial \Sigma} \frac{m(1/(T-p)) \sum_{t=p+1}^{T} \psi_2(d_t^*) \left( \frac{\partial \psi_2}{\partial \Sigma} - \frac{d_t^*}{s^*} \frac{\partial s^*}{\partial \Sigma} \right)}{(1/(T-p)) \sum_{t=p+1}^{T} \rho_2(d_t^*)}.
\]

(A.11)

Substituting (A.9) and (A.10) in (A.11) we get

\[
\frac{\partial L(\Omega, \Sigma)}{\partial \Sigma} = \Sigma^{-1} - \frac{m}{s^*} \Sigma^{-1} \left( \sum_{t=p+1}^{T} w_1(d_t^*) \frac{\partial \Sigma}{\partial \Sigma} \right) \Sigma^{-1}
\]

\[
- \frac{(m/(T-p)) \sum_{t=p+1}^{T} w_2(d_t^*) \Sigma^{-1} a_t a_t^* \Sigma^{-1}}{2s^2 \left( \frac{1}{(T-p)} \sum_{t=p+1}^{T} \rho_2(d_t^*) \right)}
\]

\[
+ \left[ \frac{(m/(T-p)) \sum_{t=p+1}^{T} \psi_2(d_t^*) d_t^*}{2s^2 \left( \frac{1}{(T-p)} \sum_{t=p+1}^{T} \rho_2(d_t^*) \right) \left( \sum_{t=p+1}^{T} \psi_1(d_t^*) d_t^* \right)} \right] \left[ \Sigma^{-1} \left( \sum_{t=p+1}^{T} w_1(d_t^*) a_t a_t^* \right) \Sigma^{-1} \right].
\]

(A.12)

Then, by (A.7)

\[
2 \Sigma \left( \frac{1}{T-p} \sum_{t=p+1}^{T} \rho_2(d_t^*) \right) \left( \sum_{t=p+1}^{T} w_1(d_t^*) \right)
\]

\[
= \frac{2m}{s^2} \left( \frac{1}{T-p} \sum_{t=p+1}^{T} \rho_2(d_t^*) \right) \left( \sum_{t=p+1}^{T} w_1(d_t^*) a_t a_t^* \right)
\]

\[
+ \frac{m}{s^2} \left( \sum_{t=p+1}^{T} \psi_1(d_t^*) d_t^* \right) \left( \frac{1}{T-p} \sum_{t=p+1}^{T} w_2(d_t^*) a_t a_t^* \right)
\]

\[
+ \frac{m}{s^2} \left( \frac{1}{T-p} \sum_{t=p+1}^{T} w_2(d_t^*) \right) \left( \sum_{t=p+1}^{T} w_1(d_t^*) a_t a_t^* \right).
\]

(A.13)

Solving \(\Sigma\) from (A.13) we get
\[
\Sigma = \frac{m}{s^{2}} \frac{\sum_{t=p+1}^{T} w_{T}^{*}(d_{t}^{*})a_{t}a_{t}'}{2 \left( \frac{1}{T-p} \sum_{t=p+1}^{T} \rho_{2}(d_{t}^{*}) \left( \sum_{t=p+1}^{T} \psi_{1}(d_{t}^{*})d_{t}^{*} \right) \right)}. \tag{A.14}
\]

We also have
\[
2 \left( \frac{1}{T - p} \sum_{t=p+1}^{T} \rho_{2}(d_{t}^{*}) \right) \left( \sum_{t=p+1}^{T} \psi_{1}(d_{t}^{*})d_{t}^{*} \right) = 2BT \sum_{t=p+1}^{T} \rho_{2}(d_{t}^{*}). \tag{A.15}
\]

From the definition of \( A_{T} \) we get
\[
2 \sum_{t=p+1}^{T} \rho_{2}(d_{t}^{*}) = (T - p)A_{T} + \sum_{t=p+1}^{T} \psi_{2}(d_{t}^{*})d_{t}^{*}, \tag{A.16}
\]
and then
\[
2BT \sum_{t=p+1}^{T} \rho_{2}(d_{t}^{*}) = \left( \sum_{t=p+1}^{T} \psi_{1}(d_{t}^{*})d_{t}^{*} \right) A_{T} + \left( \sum_{t=p+1}^{T} \psi_{2}(d_{t}^{*})d_{t}^{*} \right) B_{T} = \sum_{t=p+1}^{T} \psi_{2}^{*}(d_{t}^{*})d_{t}^{*}.
\]

Then by (A.14) and (A.15), we obtain
\[
\Sigma = \frac{m}{s^{2}} \frac{\sum_{t=p+1}^{T} w_{T}^{*}(d_{t}^{*})a_{t}a_{t}'}{\sum_{t=p+1}^{T} \psi_{T}(d_{t}^{*})d_{t}^{*}},
\]
and this proves that (A.2) is equivalent to (2.8). \( \blacksquare \)

Heuristic Proof of Consistency. Let \( s_{0}(F, \Omega, \Sigma) \) and \( \tau_{0}(F, \Omega, \Sigma) \) be the limit values under \( F \) of \( s(d_{p+1}(\Omega, \Sigma), ..., d_{T}(\Omega, \Sigma)) \) and \( \tau(d_{p+1}(\Omega, \Sigma), ..., d_{T}(\Omega, \Sigma)) \) respectively. Therefore
\[
E_{F} \left( \rho_{1} \left( \frac{d_{t}(\Omega, \Sigma)}{s_{0}(F, \Omega, \Sigma)} \right) \right) = \kappa_{1} \tag{A.17}
\]
and
\[
\tau_{0}^{2}(F, \Omega, \Sigma) = \frac{m}{\kappa_{2}} s_{0}^{2}(F, \Omega, \Sigma) E_{F} \left( \rho_{2} \left( \frac{d_{t}(\Omega, \Sigma)}{s_{0}(F, \Omega, \Sigma)} \right) \right). \tag{A.18}
\]
Taking limits in (2.7) and (2.8), we derive that the limit value of \((\tilde{\Omega}, \tilde{\Sigma})\) should satisfy
\[
E_F w_t^* (d_t^*(\Omega, \Sigma)) \hat{a}_t(\Omega) z_t' = 0
\]  
(A.19)

and
\[
\Sigma = \frac{m E_F (w_t^* (d_t(\Omega, \Sigma)/s_0(F, \Omega, \Sigma)) \hat{a}_t(\Omega) \hat{a}_t'(\Omega))}{s_0(F, \Omega, \Sigma) E_F \psi_H (d_t(\Omega, \Sigma)/s_0(F, \Omega, \Sigma)) d_t(\Omega, \Sigma)}.
\]  
(A.20)

Since \(\tau_2^2(d_{p+1}(\tilde{\Omega}, \tilde{\Sigma}), ..., d_T(\tilde{\Omega}, \tilde{\Sigma})) = m\), the limit value of \((\tilde{\Omega}, \tilde{\Sigma})\) should also satisfy
\[
\tau_0^2(F, \Omega, \Sigma) = m.
\]  
(A.21)

Let \((\Omega_0, \Sigma_0)\) be the true values of \((\Omega, \Sigma)\) and \(\Sigma_0^* = \lambda_0(H)\Sigma_0\), where \(\lambda_0(H)\) is defined in (2.9). We will show that \((\Omega_0, \Sigma_0^*)\) satisfies (A.19), (A.20) and (A.21). We start proving that \(s_0(F, \Omega_0, \Sigma_0^*) = \sigma_0\) and that \(\tau_0^2(F, \Omega_0, \Sigma_0^*) = m\). By (2.10)
\[
E_F \left( \rho_1 \left( \frac{d_t(\Omega_0, \Sigma_0^*)}{\sigma_0} \right) \right) = E_F \left( \rho_1 \left( \frac{(a_t' \Sigma_0^{-1} a_t)^{1/2}}{\sigma_0} \right) \right) = E_F \left( \rho_1 \left( \frac{(a_t' \Sigma_0^{-1} a_t)^{1/2}}{k_0} \right) \right) = E_H \left( \rho_1 \left( \frac{u}{k_0} \right) \right) = \kappa_1,
\]
and therefore \(s_0(F, \Omega_0, \Sigma_0^*) = \sigma_0\). Then, using (2.11), we obtain
\[
\frac{m}{\kappa_2} \sigma_0^2 E_F \left( \rho_2 \left( \frac{(a_t' \Sigma_0^{-1} a_t)^{1/2}}{\sigma_0} \right) \right) = \frac{m}{\kappa_2} \sigma_0^2 E_F \left( \rho_2 \left( \frac{(a_t' \Sigma_0^{-1} a_t)^{1/2}}{k_0} \right) \right) = \frac{m}{\kappa_2} \sigma_0^2 E_H \left( \rho_2 \left( \frac{u}{k_0} \right) \right) = m
\]

and then
\[
\tau_0^2(F, \Omega_0, \Sigma_0^*) = m.
\]  
(A.22)

Since \(a_t\) and \(z_t\) are independent, we have
where (A.20) and (A.21).

we have that for any function

$$\lim_{T \to \infty} T$$

equation (2.7) is equivalent to

Put $a^*_t = \Sigma_0^{-1/2} a_t$, and let $F^*$ be its distribution. Since $F^*$ is spherical, we have that for any function $q$

$$E_F \left( \frac{q(||a^*_t||)}{||a^*_t||} a^*_t a^*_t \right) = \frac{1}{m} E_{F^*} \left( q(||a^*_t||)||a^*_t|| \right),$$

where $|| \cdot ||$ denotes Euclidean norm. Then, taking $q(u) = \psi^*(u/\sigma_0(H))$, we get

$$\frac{mE_F (w^*(d_t(\Omega_0, \Sigma_0^*))/s_0(F, \Omega_0, \Sigma_0^*))}{s_0(F, \Omega_0, \Sigma_0^*)} E_F (\psi^*(d_t(\Omega_0, \Sigma_0^*)/s_0(F, \Omega_0, \Sigma_0^*))a_t a_t^*/(a_t \Sigma_0^{-1} a_t)^{1/2})$$

$$\frac{E_F (\psi^*((a_t \Sigma_0^{-1} a_t)^{1/2}/s_0(F, \Omega_0, \Sigma_0^*))a_t a_t^*/(a_t \Sigma_0^{-1} a_t)^{1/2})}{mE_F (\psi^*(||a^*_t||)/s_0(F, \Omega_0, \Sigma_0^*))a^*_t a^*_t/||a^*_t|| \Sigma^{-1/2} \Sigma_0^{-1/2}}$$

$$= \Sigma_0^{-1/2} \Sigma_0^{-1/2} = \Sigma_0^*.$$

It follows from (A.23), (A.24) and (A.22) that $(\Omega_0, \Sigma_0^*)$ satisfies (A.19), (A.20) and (A.21).

**Proof of Theorem 2.** We will start assuming that $\Sigma_0 = I_n$. Since

$$\lim_{T \to \infty} \tilde{\Omega} = \Omega_0$$

and $\lim_{T \to \infty} \tilde{\Sigma} = (k_0^2/\sigma_0^2) \Sigma_0$, it can be proved that asymptotically equation (2.7) is equivalent to

$$Q(\tilde{\Omega}) = 0,$$

where

$$Q(\Omega) = \sum_{t=p+1}^{T} w^*(||\tilde{a}_t(\Omega)||)\tilde{a}_t(\Omega) z^*_t.$$

Put $Q(\Omega) = (q_{ij}(\Omega))$, then, by the mean value theorem we have

17
\[ q_{ij}(\tilde{\Omega}) \simeq q_{ij}(\Omega_0) + \sum_{u=1}^{m} \sum_{v=1}^{r} \frac{\partial q_{ij}(\Omega_0)}{\partial \omega_{uv}} \bigg|_{\Omega=\Omega_0} (\tilde{\omega}_{uv} - \omega_{0uv}), \quad (A.25) \]

where \( \simeq \) means asymptotically equivalent.

On the other hand

\[ \frac{\partial q_{ij}(\Omega)}{\partial \omega_{uv}} \bigg|_{\Omega=\Omega_0} = \sum_{t=p+1}^{T} w^*(||a_t||) \frac{\partial |\hat{a}_t(\Omega)||}{\partial \omega_{uv}} \bigg|_{\Omega=\Omega_0} a_t z_t + w^*(||a_t||) \frac{\partial \hat{a}_t(\Omega)}{\partial \omega_{uv}} \bigg|_{\Omega=\Omega_0} z_t, \quad (A.26) \]

\[ \frac{\partial ||a_t||}{\partial \omega_{uv}} \bigg|_{\Omega=\Omega_0} = \frac{-a_t z_t}{||a_t||} \quad (A.27) \]

and

\[ \frac{\partial \hat{a}_{ti}(\Omega)}{\partial \omega_{uv}} \bigg|_{\Omega=\Omega_0} = -z_t \delta_{iu}. \quad (A.28) \]

Using (A.27), (A.28) and (A.26) we obtain

\[ \frac{\partial q_{ij}(\Omega)}{\partial \omega_{uv}} \bigg|_{\Omega=\Omega_0} = - \sum_{t=p+1}^{T} \left[ w^*_H(||a_t||)(a_t a_{tu}/||a_t||) z_t z_t + \delta_{iu} w^*(||a_t||) z_t z_t \right]. \quad (A.29) \]

Put \( \tilde{\omega}_{uv} = (T - p)^{1/2}(\tilde{\omega}_{uv} - \omega_{0uv}) \),

\[ f_{ij}^{uv} = \frac{1}{T - p} \sum_{t=p+1}^{T} \left[ w^*_H(||a_t||)(a_t a_{tu}/||a_t||) z_t z_t + \delta_{iu} w^*(||a_t||) z_t z_t \right] \]

and

\[ r_{T,ij} = \frac{1}{(T - p)^{1/2}} \sum_{t=p+1}^{T} w^*(||a_t||) a_t z_t. \]

Replacing (A.29) in (A.25) we get the following approximating equation

\[ r_{T,ij} \simeq \sum_{u=1}^{m} \sum_{v=1}^{r} f_{uv}^{ij}\tilde{\omega}_{uv}^*. \quad (A.30) \]

By the ergodic theorem we have

\[ \lim_{T \to \infty} f_{uv}^{ij} = \begin{cases} G_{m_{jv}} & \text{if } u = i \\ 0 & \text{if } u \neq i \text{ a.s.} \end{cases} \]
where
\[ G = E(w_H^*(||a_t||)||a_t||)/m + E(w^*(||a_t||)) \]
and
\[ m_{jv} = E(z_{lv}z_{tv}). \]

Then by (A.30) a new asymptotic equation is
\[ \rho_{T,ij}' = G \rho_{X,v} = 1, \]
and putting \( \bar{\omega}_i = (\bar{\omega}_{i1}, ..., \bar{\omega}_{ir})', \rho_{T,i} = (\rho_{T,i1}, ..., \rho_{T,ir})' \), and \( M = (m_{ij}) \) we get
\[ \bar{\omega}_i \sim \frac{1}{G} M^{-1} \rho_{T,i}. \]  

The central limit theorem implies that the matrix \( \rho_{T,ij} \) (\( i = 1, ..., r, j = 1, ..., r \)) converges to a Gaussian process \( r_{ij} \) with mean 0 and covariance matrix
\[ \text{cov}(r_{ij}r_{i'j'}) = \begin{cases} 
  dm_{jj'} & \text{if } i = i' \\
  0 & \text{if } i \neq i' 
\end{cases}, \]
where
\[ d = E(w_H^2(||a_t||)||a_t||)/m. \]

Then from (A.31) \( \bar{\omega}_i \) is asymptotically \( N(0, J(\psi_1, \psi_2, H)M^{-1}) \) where
\[ J(\psi_1, \psi_2, H) = \frac{mE(w_H^2(||a_t||)||a_t||)^2}{(E(w_H^2(||a_t||)||a_t||) + mE(w_H^2(||a_t||)))^2} = \frac{mE(\psi^2(||a_t||))}{(E(\psi^2(||a_t||)) + (m - 1)E(w_H^2(||a_t||)))^2} = \frac{mE_H(\psi^2(u))}{(E_H(\psi^2(u)) + (m - 1)E(w_H^2(u)))^2}. \]

The CMLE corresponds to \( \rho_2(u) = \rho_0(u) = u^2 \) and \( \rho_1 \) arbitrary, then
\[ 2\rho_2(u) - \psi_2(u) = 0 \] and therefore \( A = 0 \). Then, \( \psi_H(u) = 2Bu/k_0, E_H(\psi^2(u)) = 4B^2E_H(u^2)/k_0^2, E_H(\psi^2(u)) = E(w_H^2(u)) = 2B/k_0. \) Then \( J(\psi_1, \psi_0, H) = E_H(u^2)/m \), and then Theorem 2 follows for the case that \( \Sigma_0 = I_m \).

Consider now the case of an arbitrary \( \Sigma_0 \). Let \( x_t^* = \Sigma_0^{-1/2}x_t \) be the transformed process, which is also a VAR(p) process satisfying
\[ x_t^* = \phi_{t1}x_{t-1} + \cdots + \phi_{tp}x_{t-p} + \nu_t^* + a_t^*, \]  

(A.32)
where $\phi_0^* = \Sigma_0^{-1/2} \phi_0 \Sigma_0^{1/2}$, $\nu_0^* = \Sigma_0^{-1/2} \nu_0$ and now the density of $a_t^*$ is $f^*(||a_t||)$. Then Theorem 2 also holds because of the equivariance of the $\tau-$estimates.

REFERENCES


Table 1. Values of $c_1$ and $\kappa_1$ for the bisquare function

<table>
<thead>
<tr>
<th>Dimension</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$</td>
<td>1.55</td>
<td>2.66</td>
<td>3.45</td>
<td>4.10</td>
<td>4.65</td>
<td>6.77</td>
</tr>
<tr>
<td>$\kappa_1$</td>
<td>0.20</td>
<td>0.59</td>
<td>0.99</td>
<td>1.40</td>
<td>1.80</td>
<td>3.82</td>
</tr>
</tbody>
</table>

Table 2. Values of the constant $c_2$ and $k_2$ for the bisquare function and different values of the relative asymptotic efficiency

<table>
<thead>
<tr>
<th>RAE</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$c_2$</td>
</tr>
<tr>
<td>0.80</td>
<td>3.98</td>
</tr>
<tr>
<td>0.90</td>
<td>4.97</td>
</tr>
<tr>
<td>0.95</td>
<td>6.04</td>
</tr>
<tr>
<td>Parameter</td>
<td>Estimate</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
</tr>
<tr>
<td>ν₁</td>
<td>CMLE</td>
</tr>
<tr>
<td></td>
<td>RA</td>
</tr>
<tr>
<td></td>
<td>τ</td>
</tr>
<tr>
<td>ν₂</td>
<td>CMLE</td>
</tr>
<tr>
<td></td>
<td>RA</td>
</tr>
<tr>
<td></td>
<td>τ</td>
</tr>
<tr>
<td>φ₁₁</td>
<td>CMLE</td>
</tr>
<tr>
<td></td>
<td>RA</td>
</tr>
<tr>
<td></td>
<td>τ</td>
</tr>
<tr>
<td>φ₁₂</td>
<td>CMLE</td>
</tr>
<tr>
<td></td>
<td>RA</td>
</tr>
<tr>
<td></td>
<td>τ</td>
</tr>
<tr>
<td>φ₂₁</td>
<td>CMLE</td>
</tr>
<tr>
<td></td>
<td>RA</td>
</tr>
<tr>
<td></td>
<td>τ</td>
</tr>
<tr>
<td>φ₂₂</td>
<td>CMLE</td>
</tr>
<tr>
<td></td>
<td>RA</td>
</tr>
<tr>
<td></td>
<td>τ</td>
</tr>
</tbody>
</table>
Table 4. Mean square error (MSE) and relative efficiency (RE) of the estimates - Model 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Contamination type</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>none</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSE</td>
<td>RE</td>
<td>MSE</td>
<td>RE</td>
<td>MSE</td>
</tr>
<tr>
<td>$\nu_1$</td>
<td>CMLE</td>
<td>2.30</td>
<td>1.00</td>
<td>8.98</td>
<td>1.00</td>
<td>7.38</td>
</tr>
<tr>
<td></td>
<td>RA</td>
<td>2.46</td>
<td>0.94</td>
<td>3.94</td>
<td>2.28</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>2.50</td>
<td>0.92</td>
<td>3.91</td>
<td>2.30</td>
<td>4.00</td>
</tr>
<tr>
<td>$\nu_2$</td>
<td>CMLE</td>
<td>1.55</td>
<td>1.00</td>
<td>2.19</td>
<td>1.00</td>
<td>8.49</td>
</tr>
<tr>
<td></td>
<td>RA</td>
<td>1.69</td>
<td>0.91</td>
<td>2.06</td>
<td>1.07</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>1.80</td>
<td>0.86</td>
<td>1.95</td>
<td>1.13</td>
<td>2.38</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>CMLE</td>
<td>0.67</td>
<td>1.00</td>
<td>18.9</td>
<td>1.00</td>
<td>14.8</td>
</tr>
<tr>
<td></td>
<td>RA</td>
<td>0.74</td>
<td>0.90</td>
<td>2.76</td>
<td>6.86</td>
<td>2.65</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>0.75</td>
<td>0.88</td>
<td>2.26</td>
<td>8.38</td>
<td>4.74</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>CMLE</td>
<td>0.57</td>
<td>1.00</td>
<td>7.31</td>
<td>1.00</td>
<td>9.28</td>
</tr>
<tr>
<td></td>
<td>RA</td>
<td>0.64</td>
<td>0.89</td>
<td>1.27</td>
<td>5.76</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>0.65</td>
<td>0.87</td>
<td>1.10</td>
<td>6.64</td>
<td>3.40</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>CMLE</td>
<td>0.51</td>
<td>1.00</td>
<td>2.54</td>
<td>1.00</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>RA</td>
<td>0.56</td>
<td>0.91</td>
<td>0.68</td>
<td>3.73</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>0.58</td>
<td>0.88</td>
<td>0.67</td>
<td>3.81</td>
<td>0.91</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>CMLE</td>
<td>0.60</td>
<td>1.00</td>
<td>1.54</td>
<td>1.00</td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td>RA</td>
<td>0.66</td>
<td>0.90</td>
<td>0.82</td>
<td>1.88</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>0.65</td>
<td>0.92</td>
<td>0.66</td>
<td>2.31</td>
<td>1.19</td>
</tr>
</tbody>
</table>