Multivariate Least Weighted Squares (MLWS)

Stochastic Modelling in Economics and Finance 2
Supervisor: Prof. RNDr. Jan Ámos Víšek, CSc.

Petr Jonáš
12th April 2010
Contents

1 Multivariate regression

2 Weighted regression

3 Multivariate Least Weighted Squares Algorithm

4 Robust VAR estimator
Multivariate regression

Consider the multivariate regression model

$$Y_t = B' X_t + U_t,$$

(1)

- $Y_t \in \mathbb{R}^q$, $X_t \in \mathbb{R}^p$, $t = 1, \ldots, n$,
- $U_t \sim (0, \Sigma)$, $t = 1, \ldots, n$ are i.i.d random variables,
- $\Sigma$ is symmetric and positive definite matrix $q \times q$.
- $B \in \mathbb{R}^{p \times q}$ contains the regression coefficients.
Using notation

- $X = (X_1, \ldots, X_n)'$, $Y = (Y_1, \ldots, Y_n)'$, $U = (U_1, \ldots, U_n)'$,
- $y = \text{vec}(Y)$, $\beta = \text{vec}(B)$, $u = \text{vec}(U)$,

we can rewrite (1) as follows

$$Y = XB + U,$$  \hspace{1cm} (2)

or

$$y = (I \otimes X)\beta + u.$$  \hspace{1cm} (3)

$u \sim (0, \Sigma \otimes I)$. 
Applying Aitken’s (or LS) estimator on (3) we obtain
\[ \hat{\beta} = (I \otimes (X'X)^{-1}X')y \]
or equivalently
\[ \hat{B}_{LS} = (X'X)^{-1}X'Y \] (4)
and \( \Sigma \) can be estimated as
\[ \hat{\Sigma}_{LS} = \frac{1}{n}(Y - X\hat{B}_{LS})(Y - X\hat{B}_{LS})' = \frac{1}{n} \sum_{i=1}^{n}(Y_i - \hat{B}_{LS}'X_i)(Y_i - \hat{B}_{LS}'X_i)' \]. (5)

Replacing \( n \) in the denominator by \( n - p \), we obtain an unbiased estimate of \( \Sigma \). For details see Judge et al. (1988) or Lütkepohl (2005).
The goal is to generalize least weighted squares (LWS) estimator to multivariate case.

\[ \hat{\beta}_{LWS}^{n,w} := \arg \min_{\beta \in \mathbb{R}^n} \sum_{i=1}^{n} w \left( \frac{i - 1}{n} \right) r_{(i)}^2(\beta), \]  

(6)

The Mahalanobis distance is defined as

\[ d_t(B, \Sigma) = \left( (Y_t - B'X_t)' \Sigma^{-1}(Y_t - B'X_t) \right)^{1/2}. \]

We can define the multivariate least squares (MLS) estimator by

\[ \hat{B}_{MLS} = \left\{ B^* \mid (B^*, \Sigma^*) \in \arg \min_{(B, \Sigma); \|\Sigma\| = 1} \sum_{i=1}^{n} d_i^2(B, \Sigma) \right\}. \]

Denote

\[ D = \left\{ (B, \Sigma) \mid B \in \mathbb{R}^{p \times q}, \Sigma \in \mathbb{R}^{q \times q} \text{ symmetric, positive definite matrix, } \|\Sigma\| = 1 \right\}. \]
Denote a data set $Z_n = \{(X'_i, Y'_i)' \mid i = 1, \ldots, n\} \subset \mathbb{R}^{p+q}$ and assume that not all $n$ points are lying in the same subspace of $\mathbb{R}^{p+q}$. Formally it means that for all $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$

$$\#\{(X'_i, Y'_i)' \mid \alpha' X_i + \beta' Y_i = 0\} < n. \quad (7)$$

**Lemma 1**

$$\hat{B}_{LS} = \hat{B}_{MLS},$$

for all datasets satisfying condition (7).

**Remark**  If $(X'_i, Y'_i)'$ comes as a sample from continuous distribution, condition (7) is satisfied with probability 1.
To prove Lemma 1 the following lemma will be useful. It could be found in Agulló et al. (2008).

**Lemma 2**

Let $z = (x, y)$ be a $(p+q)$-dimensional random variable having distribution $K$. Suppose that $E_K[xx']$ is a strictly positive definite matrix. Define $B_{LS}(K) = E_K[xx']^{-1}E_K[xy']$ and $\Sigma_{LS}(K) = E_K[uu']$ where $u := y - B_{LS}(K)'x$. Then among all pairs $(b, \Delta)$ with $b \in \mathbb{R}^{p \times q}$ and $\Delta$ positive definite and symmetric matrix of size $q$ such that

$$E_K[(y - b'x)'\Delta^{-1}(y - b'x)] = q, \quad (8)$$

the unique pair which minimizes det $\Delta$ is given by $\left(B_{LS}(K), \Sigma_{LS}(K)\right)$.

**Remark**  If not all point of a data set are lying on the same subspace of $\mathbb{R}^{p+q}$ (condition (7)), then Lemma 2 can be applied by taking for $K$ the empirical distribution function associated to the data.
**Proof:**

Let us denote $\tilde{\Sigma}_{LS} = (\det \hat{\Sigma}_{LS})^{-1/q} \hat{\Sigma}_{LS}$, so that $|\tilde{\Sigma}_{LS}| = 1$. First of all we give three equations which will be useful in the rest of the proof. Using the properties of traces

$$\frac{1}{n} \sum_{i=1}^{n} d_{i}^2(\hat{B}_{LS}, \hat{\Sigma}_{LS}) = \frac{1}{n} \sum_{i=1}^{n} (Y_{i} - \hat{B}_{LS}'X_{i})' \hat{\Sigma}_{LS}^{-1} (Y_{i} - \hat{B}_{LS}'X_{i})$$

$$= \text{tr}(\frac{1}{n} \sum_{i=1}^{n} (Y_{i} - \hat{B}_{LS}'X_{i})' \hat{\Sigma}_{LS}^{-1} (Y_{i} - \hat{B}_{LS}'X_{i}))$$

$$= \text{tr}(\hat{\Sigma}_{LS}^{-1} \frac{1}{n} \sum_{i=1}^{n} (Y_{i} - \hat{B}_{LS}'X_{i})' (Y_{i} - \hat{B}_{LS}'X_{i}))$$

$$= \text{tr}(I_{q}) = q,$$

(9)

thus

$$\sum_{i=1}^{n} d_{i}^2(\hat{B}_{LS}, \hat{\Sigma}_{LS}) = (\det \hat{\Sigma}_{LS})^{-1/q} \sum_{i=1}^{n} d_{i}^2(\hat{B}_{LS}, \hat{\Sigma}_{LS}).$$

(10)
Combining (9) and (10) yields

\[ \sum_{i=1}^{n} d_i^2(\hat{B}_{LS}, \tilde{\Sigma}_{LS}) = nq(\det \hat{\Sigma}_{LS})^{1/q}. \]  

(11)

Suppose there exist some \((B_1, \Sigma_1) \in D\) which

\[ \sum_{i=1}^{n} d_i^2(B_1, \Sigma_1) < \sum_{i=1}^{n} d_i^2(\hat{B}_{LS}, \tilde{\Sigma}_{LS}). \]

Using equation (11) yields

\[ \sum_{i=1}^{n} d_i^2(B_1, \Sigma_1) < \sum_{i=1}^{n} d_i^2(\hat{B}_{LS}, \tilde{\Sigma}_{LS}) \]

\[ \sum_{i=1}^{n} d_i^2(B_1, \Sigma_1) < nq(\det \hat{\Sigma}_{LS})^{1/q} \]

\[ \frac{1}{n} \sum_{i=1}^{n} d_i^2(B_1, (\det \hat{\Sigma}_{LS})^{1/q} \Sigma_1) < q, \]
So there exists a constant $0 < c < 1$ such that

$$
\frac{1}{c} \left( \frac{1}{n} \sum_{i=1}^{n} d_i^2(B_1, (\det \hat{\Sigma}_{LS})^{1/q} \Sigma_1) \right) = q
$$

(12)

$$
\frac{1}{n} \sum_{i=1}^{n} d_i^2(B_1, c(\det \hat{\Sigma}_{LS})^{1/q} \Sigma_1) = q.
$$

The equation (12) satisfies condition (8) and by employing Lemma 2 we obtain

$$
\det(c(\det \hat{\Sigma}_{LS})^{1/q} \Sigma_1) > \det \hat{\Sigma}_{LS},
$$

$$
c^q \det(\Sigma_1) > 1,
$$

which is a contradiction with our assumption $\det(\Sigma_1) = 1$ and thus for any $(B_1, \Sigma_1) \in D$

$$
\sum_{i=1}^{n} d_i^2(B_1, \Sigma_1) \geq \sum_{i=1}^{n} d_i^2(\hat{B}_{LS}, \hat{\Sigma}_{LS}).
$$
Now suppose there exist some \((B_2, \Sigma_2) \in D, B_2 \neq \hat{B}_{LS}\) satisfying

\[
\sum_{i=1}^{n} d_i^2(B_2, \Sigma_2) = \sum_{i=1}^{n} d_i^2(\hat{B}_{LS}, \hat{\Sigma}_{LS}).
\]

Using equation (11) yields

\[
\sum_{i=1}^{n} d_i^2(B_2, \Sigma_2) = nq(det \hat{\Sigma}_{LS})^{1/q}.
\]

\[
\frac{1}{n} \sum_{i=1}^{n} d_i^2(B_2, (det \hat{\Sigma}_{LS})^{1/q} \Sigma_2) = q.
\]

The equation (13) satisfies condition (8) and therefore employing Lemma 2 (uniqueness, \(\hat{B}_{LS} \neq B_2\)) we obtain

\[
\det((det \hat{\Sigma}_{LS})^{1/q} \Sigma_2) > \det \hat{\Sigma}_{LS},
\]

\[
\det \hat{\Sigma}_{LS} \det(\Sigma_2) > \det \hat{\Sigma}_{LS},
\]
It is again a contradiction with our assumption $\det(\Sigma_2) = 1$. It means that $(\hat{B}_{LS}, \tilde{\Sigma}_{LS})$ is the unique pair which minimizes

$$\sum_{i=1}^{n} d_i^2(B, \Sigma), \ (B, \Sigma) \in D$$

and it concludes the proof.
Weighted regression

To generalize the previous approach for weighted regression we have to define some proper weight function (or weights).

**Definition 3**
Let the function $w : [0, 1] \rightarrow [0, 1]$ be nonincreasing and continuous on $[0, 1]$. Then the function $w$ is called weight function.

Similarly as in the one dimensional case, multiplying for each $t = 1, \ldots, n$ the equation (1) by positive weight $\sqrt{w_t}$ leads to Multivariate Weighted Least Squares (MWLS) estimator.

$$\sqrt{w_t} Y_t = \sqrt{w_t} B' X_t + \sqrt{w_t} U_t, \quad t = 1, \ldots, n.$$  \hspace{1cm} (14)
Denote the vector of weights by \( w := (w_1, \ldots, w_n) \) and the matrix by \( W := \text{diag}\{w_1, \ldots, w_n\} \). Now we can rewrite (2) in the form

\[
W^{1/2}Y = W^{1/2}XB + W^{1/2}U.
\]  

(15)

and the MWLS estimator is therefore given by

\[
\hat{B}_{MWLS}^{n,w} = (X'WX)^{-1}X'WY.
\]  

(16)

Using (5) we obtain

\[
\hat{\Sigma}_{MWLS}^{n,w} = \frac{1}{n}(Y - X\hat{B}_{MWLS}^{n,w})'(W(Y - X\hat{B}_{MWLS}^{n,w})).
\]  

(17)
Denote $Y^* = W^{1/2} Y$, $X^* = W^{1/2} X$, $U^* = W^{1/2} U$ and $d_i^* = (Y_i^* - B' X_i^*)' \Sigma^{-1} (Y_i^* - B' X_i^*)$. The MWLS estimator could be equivalently obtained as

$$\hat{B}_{MWLS}^{n,w} = \arg \min_{(B, \Sigma) \in D} \sum_{s=1}^{n} d_s^*(B, \Sigma)$$

$$= \arg \min_{(B, \Sigma) \in D} \sum_{s=1}^{n} (Y_s^* - B' X_s^*)' \Sigma^{-1} (Y_s^* - B' X_s^*)$$

$$= \arg \min_{(B, \Sigma) \in D} \sum_{s=1}^{n} w_s (Y_s - B' X_s)' \Sigma^{-1} (Y_s - B' X_s)$$

$$= \arg \min_{(B, \Sigma) \in D} \sum_{s=1}^{n} w_s d_s^2(B, \Sigma)$$

$$\tilde{\Sigma}_{MWLS}^{n,w} = (\det(\hat{\Sigma}_{MWLS}^{n,w}))^{-1/q} \hat{\Sigma}_{MWLS}^{n,w}.$$  \hspace{1cm} (18)
Multivariate Least Weighted Squares

Definition 4
Let us define the multivariate least weighted squares (MLWS) estimator of the matrix of regression coefficients $B$ as

$$\hat{B}_{MLWS}^{n,w} = \arg \min_{(B, \Sigma) \in D} \sum_{s=1}^{n} w \left( \frac{s - 1}{n} \right) d_{(s)}^2(B, \Sigma),$$

where $d_{(1)}(B, \Sigma) \leq d_{(2)}(B, \Sigma) \leq \cdots \leq d_{(n)}(B, \Sigma)$ is the ordered sequence of the residual Mahalanobis distances.
For any \( m \in \mathbb{N} \) denote by \( \Pi_m \) the set of all permutations of \( \{1, \ldots, m\} \) and for any \( \pi \in \Pi_n \), \( \pi_j \) stands for the \( j \)-th coordinate of the vector \( \pi \). For any \((B, \Sigma) \in D\) put

\[
\pi(B, \Sigma, i) = j \in \{1, \ldots, n\} \iff d_i^2(B, \Sigma) = d_{\pi_j}^2(B, \Sigma) \quad (20)
\]

and \( \pi(B, \Sigma) = (\pi(B, \Sigma, 1), \ldots, \pi(B, \Sigma, n)) \). Now we can express the equation (19) in the equivalent form

\[
\hat{B}_{MLWS}^n = \arg \min_{(B, \Sigma) \in D} \sum_{i=1}^{n} w \left( \frac{\pi(B, \Sigma, i) - 1}{n} \right) d_i^2(B, \Sigma), \quad (21)
\]
For an arbitrary $\pi \in \Pi_n$ denote the function we want to minimize by

$$MF(B, \Sigma, \pi) = \sum_{i=1}^{n} w \left( \frac{\pi_i - 1}{n} \right) d_i^2(B, \Sigma).$$  \hspace{1cm} (22)

To prove the existence of the solution of our minimization problem (19) we can follow the steps of the proof for LWS presented in Víšek (2008) and Víšek (2009).
1. \( \pi(B, \Sigma) \in \Pi_n \) therefore

\[
\min_{(B, \Sigma) \in D} \min_{\pi \in \Pi_n} \sum_{i=1}^{n} w \left( \frac{\pi_i - 1}{n} \right) d_i^2(B, \Sigma) \leq \leq \min_{(B, \Sigma) \in D} \sum_{i=1}^{n} w \left( \frac{\pi(B, \Sigma, i) - 1}{n} \right) d_i^2(B, \Sigma)
\]

\[
\Updownarrow
\]

\[
\min_{(B, \Sigma) \in D} \min_{\pi \in \Pi_n} MF(B, \Sigma, \pi) \leq \min_{(B, \Sigma) \in D} MF(B, \Sigma, \pi(B, \Sigma)). \quad (23)
\]
2. For an arbitrary fixed \((\tilde{B}, \tilde{\Sigma}) \in D\) the equations (20) and (22) yield

\[
MF(\tilde{B}, \tilde{\Sigma}, \pi(\tilde{B}, \tilde{\Sigma})) = \sum_{i=1}^{n} w \left( \frac{\pi(\tilde{B}, \tilde{\Sigma}, i) - 1}{n} \right) d_{i}^{2}(\tilde{B}, \tilde{\Sigma}) = \sum_{i=1}^{n} w \left( \frac{i - 1}{n} \right) d_{(i)}^{2}(\tilde{B}, \tilde{\Sigma}).
\]

(24)

The last equation in (24) means, that the largest distances are multiplied by the smallest weights and any other combination of weights and distances can’t produce smaller value of the sum. Hence for any \((B, \Sigma) \in D\) and \(\pi \in \Pi_{n}\) we have

\[
MF(B, \Sigma, \pi(B, \Sigma)) \leq MF(B, \Sigma, \pi).
\]

(25)
3. (23) and (25) yields

\[
\min_{(B, \Sigma) \in D} \min_{\pi \in \Pi_n} MF(B, \Sigma, \pi) = \min_{(B, \Sigma) \in D} MF(B, \Sigma, \pi(B, \Sigma)). \tag{26}
\]
4. Fix arbitrary $\omega_0 \in \Omega$, and $\pi \in \Pi_n$, we can evaluate the Multivariate Weighted Least Squares estimator with the weights given by the weight matrix

$$W(\pi) = \text{diag}\{w \left( \frac{\pi_i - 1}{n} \right), \ldots, w \left( \frac{\pi_n - 1}{n} \right)\}$$

$$(\hat{B}_{MWLS}^{n, \pi}, \tilde{\Sigma}_{MWLS}^{n, \pi}) = \arg\min_{(B, \Sigma) \in D} \sum_{s=1}^{n} w \left( \frac{\pi_s - 1}{n} \right) d^2_s(B, \Sigma)$$

For any $(B, \Sigma) \in D$

$$MF(\hat{B}_{MWLS}^{n, \pi}, \tilde{\Sigma}_{MWLS}^{n, \pi}, \pi) \leq MF(B, \Sigma, \pi). \quad (27)$$
5. We can repeat the previous step for all $\pi \in \Pi_n$ and for our fixed $\omega_0$ we can define

$$\pi(\omega_0) = \arg\min_{\pi \in \Pi_n} MF(\hat{B}^{n,\pi}_{MWLS}, \tilde{\Sigma}^{n,\pi}_{MWLS}, \pi)$$

$$= \arg\min_{\pi \in \Pi_n} MF(\hat{B}^{n,\pi}_{MWLS}(\omega_0), \tilde{\Sigma}^{n,\pi}_{MWLS}(\omega_0), \pi).$$

(28)

The estimates $(\hat{B}^{n,\pi}_{MWLS}(\omega_0), \tilde{\Sigma}^{n,\pi}_{MWLS}(\omega_0))$ of course depends on $\omega_0$ but we will use the shorter form $(\hat{B}^{n,\pi}_{MWLS}, \tilde{\Sigma}^{n,\pi}_{MWLS})$ instead. For any $\pi \in \Pi_n$ (27) and (28) yields

$$MF(\hat{B}^{n,\pi(\omega_0)}_{MWLS}, \tilde{\Sigma}^{n,\pi(\omega_0)}_{MWLS}, \pi(\omega_0)) \leq MF(\hat{B}^{n,\pi}_{MWLS}, \tilde{\Sigma}^{n,\pi}_{MWLS}, \pi)$$

(29)
6. Using (27) and (29) for any \( \tilde{\pi} \in \Pi_n \) and any \( (\tilde{B}, \tilde{\Sigma}) \in D \)

\[
MF(\hat{B}_{MWLS}^n, \tilde{\Sigma}_{MWLS}^n, \pi(\omega_0)) \leq MF(\hat{B}_{MWLS}^n, \tilde{\Sigma}_{MWLS}^n, \tilde{\pi}) \leq MF(\tilde{B}, \tilde{\Sigma}, \tilde{\pi})
\]

\[
\updownarrow
\]

\[
MF(\hat{B}_{MWLS}^n, \tilde{\Sigma}_{MWLS}^n, \pi(\omega_0)) = \min_{(B, \Sigma) \in D} \min_{\pi \in \Pi_n} MF(B, \Sigma, \pi)
\]  

(30)
7. Applying (24) and (26) to the equality (30) we obtain

$$MF(\hat{B}_{MWLS}^n, \pi(\omega_0), \tilde{\Sigma}_{MWLS}^n, \pi(\omega_0)) = \min_{(B, \Sigma) \in D} MF(B, \Sigma, \pi(B, \Sigma))$$

$$= \min_{(B, \Sigma) \in D} \sum_{i=1}^{n} w \left( \frac{i-1}{n} \right) d^2_{(i)}(B, \Sigma)$$

and due to the definition of $\hat{B}_{MLWS}^{n,w}(\omega_0)$ we finally obtain

$$\hat{B}_{MLWS}^{n,w}(\omega_0) = \hat{B}_{MWLS}^{n,\pi(\omega_0)}(\omega_0). \quad (31)$$
Repeating steps 5-7 for all $\omega \in \Omega$ completes the proof. Hence, we have just proven, that there allways exists the solution of the minimization problem (21) (and also (19)).

**Remark** As a consequence of (31), for any $\omega_0 \in \Omega$ the MLWS estimator is equal to the MWLS estimator with weights $\pi(\omega_0)$ defined in (28).
Normal equations

\[
\hat{B}_{MWLS}^{n,w} = \arg \min_{B \in \mathbb{R}^{p \times q}} \sum_{i=1}^{n} w_i (Y_i - B' X_i)'
\]

\[
\left[ (\det(\frac{1}{n} \sum_{i=1}^{n} w_i (Y_i - B' X_i)' (Y_i - B' X_i)))^{-1/q} \times \frac{1}{n} \sum_{i=1}^{n} w_i (Y_i - B' X_i)' (Y_i - B' X_i) \right]^{-1} (Y_i - B' X_i)
\]

\[
= \arg \min_{B \in \mathbb{R}^{p \times q}} \sum_{i=1}^{n} w_i (Y_i - B' X_i)'
\]

\[
\left[ (\det(\frac{1}{n} (Y - XB)' W (Y - XB)))^{-1/q} \times \frac{1}{n} (Y - XB)' W (Y - XB) \right]^{-1} (Y_i - B' X_i).
\]
The last equation gives us the way how to compute MLWS. We can minimize

$$MF(\hat{B}_{MWLS}^{n,\pi}, \tilde{\Sigma}_{MWLS}^{n,\pi}, \pi)$$

over the set of all permutations $\Pi_n$. Asymptotic complexity of this algorithm is $n!$. Impossible to compute for $n$ larger than 10. $10! = 3628800$.

It is better to modify iterative algorithm for LWS presented in Mašíček (2004).
Algorithm

1. \( \hat{B} = 0, \hat{\Sigma} = I, \hat{MF} = +\infty \)
2. choose initial values \( B_1, \Sigma_1, j := 1 \)
3. compute Mahalanobis distances \( d_i(B_j, \Sigma_j) \)
4. assign weights to observations (smaller MD = higher weight)
5. \( B_{j+1}, \Sigma_{j+1} \) MWLS with weights computed in step 4
6. \( B_j \neq B_{j+1}; j:=j+1 \) and step 3
7. if \( MF(B_j, \Sigma_j, w_j) < \hat{MF} \) then
   \( \hat{B} := B_j, \hat{\Sigma} = \Sigma_j, \hat{MF} = MF(B_j, \Sigma_j, w_j) \)
8. repeat step 2 until terminal condition
9. return values \( \hat{B}, \hat{\Sigma}, \hat{MF} \)
Previous algorithm is still too slow. Possible improvements:

- choice of initial values

- heuristic on inperspective solutions
Choice of initial values

Randomly choose $m = 1000$ subsets

- $H_i$ of size $h \approx n/2, n/4$ and compute $\hat{B} = B_{LS}(H_i)$ and $\hat{\Sigma} = \Sigma_{LS}(H_i)$.

- $H_i$ of size $h = p + q$ and compute the coefficients of the hyperplane through $H_i$. If $H_i$ does not define a unique hyperplane extend $H_i$ by adding random observations until it does. $\hat{B} =$coefficients of this hyperplane $\hat{\Sigma} = i$. 
Heuristic on inperspective solutions

- perform 2 iteration steps with all $m$ solutions
- choose $a = 10$ with lowest MFs
- perform 30 iteration steps with $a$ solutions
- choose one with lowest MF
- iterate until convergence (or terminal condition)
Vector autoregressive (VAR) model

\textbf{Definition 5}

\(K\)-dimensional random process \(\{Y_t, t \in \mathbb{Z}\}\) is called \textit{K-dimensional vector autoregressive process of order p}, when

\[
Y_t = \nu + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + U_t, \quad t \in \{-p+1, \ldots, T\},
\]

where

\begin{itemize}
  \item \(A_1, \ldots, A_p (A_p \neq 0)\) are \((K \times K)\) coefficient matrices,
  \item \(\nu\) is \((K \times 1)\) vector of intercept term,
  \item \(\{U_t, t \in \mathbb{Z}\}\) is a white noise \((\mathbb{E}U_t = 0, \mathbb{E}U_t U_s = \delta_{ts}\Sigma_U)\).
\end{itemize}
For $t = 1, \ldots, T$ denote

$$X_t = (1, Y_{t-1}', \ldots, Y_{t-p}')' \in \mathbb{R}^{pk+1},$$

$$B = (\nu, A_1, \ldots, A_p)' ,$$

$$X = (X_1, \ldots, X_T)',$$

$$Y = (Y_1, \ldots, Y_T)',$$

We can rewrite the equation (32) to

$$Y = BX + U.$$  \hfill (33)

$$\hat{B}_{LS} = (X'X)^{-1}X'Y.$$  \hfill (34)
$Y_t = \begin{bmatrix} .01 \\ .02 \end{bmatrix} + \begin{bmatrix} .40 & .03 \\ .04 & .20 \end{bmatrix} Y_{t-1} + \begin{bmatrix} .100 & .005 \\ .010 & .080 \end{bmatrix} Y_{t-2} + \begin{bmatrix} U_{1t} \\ U_{2t} \end{bmatrix}$

where $U_t \sim N(0, \Sigma)$

$$\Sigma = \begin{bmatrix} 1 & .2 \\ .2 & 1 \end{bmatrix}$$

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

$$\text{MSE} = \sum_{i=1}^{q} \sum_{j=1}^{p} \frac{1}{\text{nsim}} \sum_{s=1}^{\text{nsim}} (\hat{B}_{ij}^s - B_{ij})^2$$

Source: Croux and Joossens (2008), $\text{nsim} = 1000$. 
Figure 1: Simulated Bias and Mean Squared Error for the LS, and the robust MLTS and MLWS estimators of a bivariate VAR(2) model, in presence of m additive outliers in a series of length 500.
Bibliography I


Bibliography II


Thank You For Your Attention!