Multivariate identification

Andreas Jakobsson

Estimation

Estimation of the unknown parameters of a multivariate process works similar to the univariate case, but one has to take a bit more care with the dimensions.

The mean is estimated as

$$\hat{\mathbf{m}}_{\mathbf{y}} = \frac{1}{N}\sum_{t=1}^{N} \mathbf{y}_t$$

and the (biased) autocovariance function as

$$\hat{\mathbf{R}}_{\mathbf{y}}(k) = \frac{1}{N}\sum_{t=k+1}^{N} \left(\mathbf{y}_t - \hat{\mathbf{m}}_{\mathbf{y}}\right)\left(\mathbf{y}_{t-k} - \hat{\mathbf{m}}_{\mathbf{y}}\right)^*$$

yielding the ACF estimate

$$\hat{\boldsymbol{\rho}}_{\mathbf{y}}(k) = \hat{\mathbf{P}}_{\mathbf{y}}^{-1/2}\hat{\mathbf{R}}_{\mathbf{y}}(k)\hat{\mathbf{P}}_{\mathbf{y}}^{-1/2}$$

where

$$\hat{\mathbf{P}}_{\mathbf{y}} = \begin{bmatrix} \hat{r}_{11}(0) & 0 & \ldots & 0 \\ 0 & \hat{r}_{22}(0) & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \hat{r}_{mm}(0) \end{bmatrix}$$

Estimation

Example:
Consider the $m \times 1$ multivariate ARX(p) process

$$\mathbf{y}_t + \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} = \mathbf{B}_0\mathbf{x}_t + \ldots + \mathbf{B}_r\mathbf{x}_{t-r} + \mathbf{e}_t$$

with $\mathbf{x}_t$ being an s-dimensional, possibly stochastic, vector. Thus,

$$\mathbf{y}_t^* = -\sum_{k=1}^{p} \mathbf{y}_{t-k}^*\mathbf{A}_k^* + \sum_{k=0}^{r} \mathbf{x}_{t-k}^*\mathbf{B}_k^* + \mathbf{e}_t^* = \mathbf{X}_t^*\boldsymbol{\theta} + \mathbf{e}_t^*$$

where

$$\mathbf{X}_t^* = \begin{bmatrix} -\mathbf{y}_{t-1}^* & \ldots & -\mathbf{y}_{t-p}^* & \mathbf{x}_{t-1}^* & \ldots & \mathbf{x}_{t-r}^* \end{bmatrix}$$
$$\boldsymbol{\theta} = \begin{bmatrix} \mathbf{A}_1 & \ldots & \mathbf{A}_p & \mathbf{B}_0 & \ldots & \mathbf{B}_r \end{bmatrix}$$

or, equivalently, assuming that $p \geq r$,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_{p+1}^* \\ \vdots \\ \mathbf{y}_N^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{p+1}^* \\ \vdots \\ \mathbf{X}_N^* \end{bmatrix}\boldsymbol{\theta} + \begin{bmatrix} \mathbf{e}_{p+1}^* \\ \vdots \\ \mathbf{e}_N^* \end{bmatrix} = \mathbf{X}\boldsymbol{\theta} + \mathbf{E}$$

which suggests the LS parameter estimate

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^*\mathbf{X}\right)^{-1}\mathbf{X}^*\mathbf{Y}$$

Let $\boldsymbol{\phi} = \mathrm{vec}\{\boldsymbol{\theta}\}$. Then, $V\left\{\hat{\boldsymbol{\phi}}\right\} = \boldsymbol{\Sigma} \otimes [\mathbf{X}^*\mathbf{X}]^{-1}$.

Estimation

The maximum likelihood (ML) estimate is formed as the parameters maximising the likelihood function given the observed data.

Restricting ourselves to the *real-valued* multivariate process

$$\mathbf{y}_t = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}_t$$

for $t = 1, \ldots, N$, where $\mathbf{e}_t$ is a zero-mean multivariate Gaussian process with covariance matrix $\boldsymbol{\Sigma}$. Let

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \ldots & \mathbf{y}_N \end{bmatrix}$$

Then,

$$\begin{aligned} f(\mathbf{Y}) &= \prod_{t=1}^{N}\left[(2\pi)^m \det(\boldsymbol{\Sigma})\right]^{-1/2} \exp\left\{-\frac{1}{2}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]^T \boldsymbol{\Sigma}^{-1}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]\right\} \\ &= \left[(2\pi)^m \det(\boldsymbol{\Sigma})\right]^{-N/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{N}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]^T \boldsymbol{\Sigma}^{-1}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]\right\} \end{aligned}$$

The log likelihood is thus

$$\ln f(\mathbf{Y}) = -\frac{N}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\sum_{t=1}^{N}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]^T \boldsymbol{\Sigma}^{-1}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right] + c$$

where c denotes a constant that does not depend on $\boldsymbol{\theta}$.

Estimation

Assuming first that $\mathbf{\Sigma}$ is known, the maximisation of

$$\ln f(\mathbf{Y}) = -\frac{N}{2}\ln\det\left(\mathbf{\Sigma}\right) - \frac{1}{2}\sum_{t=1}^{N}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]^T \mathbf{\Sigma}^{-1}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right] + c$$

simplifies to

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{ML} &= \arg\max_{\boldsymbol{\theta}} \ln f(\mathbf{Y})\\ &= \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\sum_{t=1}^{N}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]^T \mathbf{\Sigma}^{-1}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]\\ &= \arg\min_{\boldsymbol{\theta}} \operatorname{tr}\left(\mathbf{\Sigma}^{-1}\hat{\mathbf{D}}_{\boldsymbol{\theta}}\right)\end{aligned}$$

where

$$\hat{\mathbf{D}}_{\boldsymbol{\theta}} = \frac{1}{N}\sum_{t=1}^{N}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]^T$$

is an estimate of $\mathbf{\Sigma}$.

Estimation

In case $\mathbf{\Sigma}$ is unknown, the ML estimate is formed by maximising $f(\mathbf{Y})$ over both $\boldsymbol{\theta}$ and $\mathbf{\Sigma}$, i.e.,

$$\begin{aligned}\left\{\hat{\boldsymbol{\theta}}_{ML}, \hat{\mathbf{\Sigma}}\right\} &= \arg\min_{\boldsymbol{\theta},\mathbf{\Sigma}}\left\{N\ln\det\left(\mathbf{\Sigma}\right) + \sum_{t=1}^{N}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]^T \mathbf{\Sigma}^{-1}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]\right\}\\ &= \arg\min_{\boldsymbol{\theta}}\left\{\ln\det\left(\hat{\mathbf{D}}_{\boldsymbol{\theta}}\right)\right\}\end{aligned}$$

where

$$\hat{\mathbf{D}}_{\boldsymbol{\theta}} = \frac{1}{N}\sum_{t=1}^{N}\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]\left[\mathbf{y}_t - \mathbf{X}\boldsymbol{\theta}\right]^T$$

The resulting cost function is often multi-modal and requires an accurate initialisation.

Estimation

Model order selection:

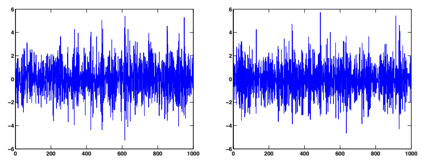
$$\begin{aligned}Q^* &= N^2\sum_{\ell=1}^{K}(N-\ell)^{-1} tr\left\{\hat{\mathbf{R}}_e^T(\ell)\hat{\mathbf{R}}_e^{-1}(0)\hat{\mathbf{R}}_e^T(\ell)\hat{\mathbf{R}}_e^{-1}(0)\right\}\\ M_p &= -\left(N - p - mp - \frac{1}{2}\right)\ln\left[\frac{|\hat{\mathbf{\Sigma}}_p|}{|\hat{\mathbf{\Sigma}}_{p-1}|}\right]\\ AIC(p) &= N\ln\left[|\hat{\mathbf{\Sigma}}_p|\right] + 2pm^2\\ BIC(p) &= N\ln\left[|\hat{\mathbf{\Sigma}}_p|\right] + pm^2\ln N\\ FPE(p) &= \left[\frac{N+mp+1}{N-mp-1}\right]^m |\hat{\mathbf{\Sigma}}_p|\end{aligned}$$

where $Q^* \in \chi^2_{1-\alpha}\left\{m^2(K-p-q)\right\}$ (use `lbpTest`) and $M_p \in \chi^2_{1-\alpha}(m^2)$

Here, $\hat{\mathbf{\Sigma}}_p$ denotes the covariance matrix of the residual when using a model of order p, i.e.,

$$\hat{\mathbf{\Sigma}}_p = \frac{1}{N-p}\left(\mathbf{Y} - \mathbf{X}_p^*\hat{\boldsymbol{\theta}}_p\right)^*\left(\mathbf{Y} - \mathbf{X}_p^*\hat{\boldsymbol{\theta}}_p\right) = \frac{\mathbf{Y}^*\mathbf{\Pi}^{\perp}_{\mathbf{X}_p}\mathbf{Y}}{N-p}$$

Example



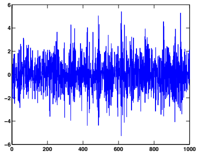
We consider a VAR(2), with

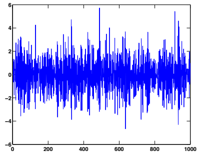
$$\mathbf{y}_t + \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} = \mathbf{e}_t$$

where

$$\mathbf{A}_1 = \begin{bmatrix} 0.5 & 0.4 \\ 0.1 & 0.8 \end{bmatrix} \qquad \mathbf{A}_2 = \begin{bmatrix} -0.2 & -0.1 \\ 0.3 & 0.6 \end{bmatrix}$$

Example





	$k=1$	$k=2$	$k=3$	$k=4$
$\boldsymbol{\rho}_{\mathbf{y}}(k)$	$\begin{bmatrix} -0.67 & -0.36 \\ 0.31 & -0.57 \end{bmatrix}$	$\begin{bmatrix} 0.41 & 0.50 \\ -0.43 & -0.08 \end{bmatrix}$	$\begin{bmatrix} -0.14 & -0.37 \\ 0.33 & 0.43 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.13 \\ -0.13 & -0.38 \end{bmatrix}$
	$\begin{bmatrix} - & - \\ + & - \end{bmatrix}$	$\begin{bmatrix} + & + \\ - & - \end{bmatrix}$	$\begin{bmatrix} - & - \\ - & + \end{bmatrix}$	$\begin{bmatrix} \cdot & + \\ - & - \end{bmatrix}$
$\phi_{k,k}$	$\begin{bmatrix} -0.67 & 0.31 \\ -0.36 & -0.57 \end{bmatrix}$	$\begin{bmatrix} 0.16 & 0.10 \\ -0.11 & -0.52 \end{bmatrix}$	$\begin{bmatrix} -0.01 & -0.04 \\ 0.05 & -0.04 \end{bmatrix}$	$\begin{bmatrix} 0.07 & -0.01 \\ -0.03 & 0.01 \end{bmatrix}$
	$\begin{bmatrix} - & + \\ - & - \end{bmatrix}$	$\begin{bmatrix} + & + \\ - & - \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} + & \cdot \\ \cdot & \cdot \end{bmatrix}$

We consider a VAR(2), with

$$\mathbf{y}_t + \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} = \mathbf{e}_t$$

where

$$\mathbf{A}_1 = \begin{bmatrix} 0.5 & 0.4 \\ 0.1 & 0.8 \end{bmatrix} \qquad \mathbf{A}_2 = \begin{bmatrix} -0.2 & -0.1 \\ 0.3 & 0.6 \end{bmatrix}$$

Our identification indicates that:

- The Box-Cox plot indicates that no transformation is required.
- No detrending seems to be required.
- We estimate ACF and PACF to determine model structure; these suggest a VAR structure of order 2 (here, $2/\sqrt{N} \approx 0.063$).
- The multivariate Jarque-Bera test indicates that the ACF and PACF are normal distributed (use `mjbtest`).

Example

p	0	1	2	3	4	5
Q^*	393	315	15	18	16	12
M_p		1404	412	3.7	6.4	4.3
AIC	1602	352	21	27	30	34
BIC	1622	391	80	106	128	152
FPE	8.614	1.626	1.021	1.029	1.035	1.037

Here, the 95% quantile for Q^* is approximately 102, and for M_p about 9.5.

Note that the AIC, BIC, and FPE all have minimum for order 2.

Using LS (use `lsVAR`), we estimate the unknown parameters

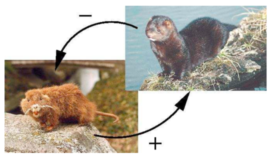
$$\mathbf{A}_1 = \begin{bmatrix} 0.52 & 0.36 \\ 0.07 & 0.28 \end{bmatrix} \qquad \mathbf{A}_2 = \begin{bmatrix} -0.17 & -0.12 \\ 0.24 & 0.58 \end{bmatrix}$$

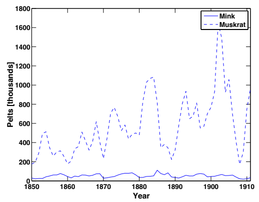
Recall that the true parameters were

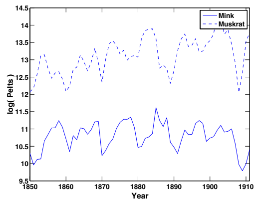
$$\mathbf{A}_1 = \begin{bmatrix} 0.5 & 0.4 \\ 0.1 & 0.8 \end{bmatrix} \qquad \mathbf{A}_2 = \begin{bmatrix} -0.2 & -0.1 \\ 0.3 & 0.6 \end{bmatrix}$$

Note that, even with 2×1000 samples, the estimates are still not all that good.

Minks and mustrats







Denote the muskrat and mink time series $y_{t,1}$ and $y_{t,2}$, respectively, and form

$$\mathbf{z}_t = \begin{bmatrix} \log y_{t,1} - \hat{m}_{\tilde{y}_1} \\ \log y_{t,2} - \hat{m}_{\tilde{y}_2} \end{bmatrix}$$

where $\hat{m}_{\tilde{y}_1} = 10,79$ and $\hat{m}_{\tilde{y}_2} = 13,12$ denote the estimated mean of the transformed muskrat and mink data sets, respectively.

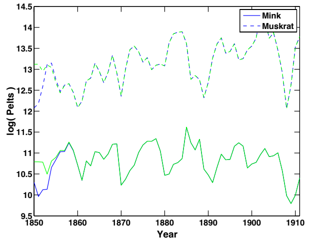
Minks and mustrats

	$k=1$	$k=2$	$k=3$	$k=4$
$\boldsymbol{\rho}_{\mathbf{y}}(k)$	$\begin{bmatrix} 0.68 & 0.47 \\ -0.13 & 0.73 \end{bmatrix}$	$\begin{bmatrix} 0.29 & 0.38 \\ -0.27 & 0.37 \end{bmatrix}$	$\begin{bmatrix} -0.02 & 0.25 \\ -0.16 & 0.20 \end{bmatrix}$	$\begin{bmatrix} -0.25 & 0.08 \\ 0.03 & 0.13 \end{bmatrix}$
	$\begin{bmatrix} + & + \\ \cdot & + \end{bmatrix}$	$\begin{bmatrix} + & + \\ - & + \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$
$\boldsymbol{\phi}_{k,k}$	$\begin{bmatrix} 0.68 & -0.13 \\ 0.47 & 0.73 \end{bmatrix}$	$\begin{bmatrix} -0.17 & -0.27 \\ 0.11 & -0.18 \end{bmatrix}$	$\begin{bmatrix} -0.24 & 0.10 \\ 0.08 & 0.15 \end{bmatrix}$	$\begin{bmatrix} -0.27 & -0.19 \\ 0.07 & -0.07 \end{bmatrix}$
	$\begin{bmatrix} + & \cdot \\ + & + \end{bmatrix}$	$\begin{bmatrix} \cdot & - \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ \cdot & \cdot \end{bmatrix}$

	$k=5$	$k=6$	$k=7$	$k=8$
$\boldsymbol{\rho}_{\mathbf{y}}(k)$	$\begin{bmatrix} -0.34 & -0.11 \\ 0.23 & 0.08 \end{bmatrix}$	$\begin{bmatrix} -0.30 & -0.33 \\ 0.35 & 0.05 \end{bmatrix}$	$\begin{bmatrix} -0.12 & -0.45 \\ 0.41 & 0.11 \end{bmatrix}$	$\begin{bmatrix} 0.10 & -0.47 \\ 0.40 & 0.25 \end{bmatrix}$
	$\begin{bmatrix} + & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} - & - \\ + & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & - \\ + & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & - \\ + & \cdot \end{bmatrix}$
$\boldsymbol{\phi}_{k,k}$	$\begin{bmatrix} -0.21 & -0.12 \\ 0.00 & -0.22 \end{bmatrix}$	$\begin{bmatrix} -0.18 & -0.23 \\ -0.20 & -0.11 \end{bmatrix}$	$\begin{bmatrix} 0.02 & -0.15 \\ -0.03 & -0.09 \end{bmatrix}$	$\begin{bmatrix} -0.02 & -0.22 \\ 0.04 & 0.11 \end{bmatrix}$
	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$

Minks and mustrats

	$k=1$	$k=2$	$k=3$	$k=4$
$\rho_{\mathbf{y}}(k)$	$\begin{bmatrix} 0.68 & 0.47 \\ -0.13 & 0.73 \end{bmatrix}$	$\begin{bmatrix} 0.29 & 0.38 \\ -0.27 & 0.37 \end{bmatrix}$	$\begin{bmatrix} -0.02 & 0.25 \\ -0.16 & 0.20 \end{bmatrix}$	$\begin{bmatrix} -0.25 & 0.08 \\ 0.03 & 0.13 \end{bmatrix}$
	$\begin{bmatrix} + & + \\ \cdot & + \end{bmatrix}$	$\begin{bmatrix} + & + \\ - & + \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$
$\phi_{k,k}$	$\begin{bmatrix} 0.68 & -0.13 \\ 0.47 & 0.73 \end{bmatrix}$	$\begin{bmatrix} -0.17 & -0.27 \\ 0.11 & -0.18 \end{bmatrix}$	$\begin{bmatrix} -0.24 & 0.10 \\ 0.08 & 0.15 \end{bmatrix}$	$\begin{bmatrix} -0.27 & -0.19 \\ 0.07 & -0.07 \end{bmatrix}$
	$\begin{bmatrix} + & \cdot \\ + & + \end{bmatrix}$	$\begin{bmatrix} \cdot & - \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ \cdot & \cdot \end{bmatrix}$
	$k=5$	$k=6$	$k=7$	$k=8$
$\rho_{\mathbf{y}}(k)$	$\begin{bmatrix} -0.34 & -0.11 \\ 0.23 & 0.08 \end{bmatrix}$	$\begin{bmatrix} -0.30 & -0.33 \\ 0.35 & 0.05 \end{bmatrix}$	$\begin{bmatrix} -0.12 & -0.45 \\ 0.41 & 0.11 \end{bmatrix}$	$\begin{bmatrix} 0.10 & -0.47 \\ 0.40 & 0.25 \end{bmatrix}$
	$\begin{bmatrix} + & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} - & - \\ + & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & - \\ + & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & - \\ + & \cdot \end{bmatrix}$
$\phi_{k,k}$	$\begin{bmatrix} -0.21 & -0.12 \\ 0.00 & -0.22 \end{bmatrix}$	$\begin{bmatrix} -0.18 & -0.23 \\ -0.20 & -0.11 \end{bmatrix}$	$\begin{bmatrix} 0.02 & -0.15 \\ -0.03 & -0.09 \end{bmatrix}$	$\begin{bmatrix} -0.02 & -0.22 \\ 0.04 & 0.11 \end{bmatrix}$
	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$



p	0	1	2	3	4	5
Q^*	41	24	29	20	18	16
M_p		122.0	12.6	8.5	8.7	5.3
AIC	-189	-317	-324	-327	-330	-330
BIC	-181	-300	-299	-293	-288	-279
FPE	0.0506	0.0065	0.0057	0.0055	0.0052	0.0053

Here, the 95% quantile for Q^* is approximately 55.8, and for M_p about 9.5.