Estimation of Vector Autoregressive Processes
Based on Chapter 3 of book by H. Lütkepohl: New Introduction to Multiple Time Series Analysis

Yordan Mahmudiev  Pavol Majher

December 13th, 2011
Introduction

Basic Assumptions:

- \( y_t = (y_{1t}, \ldots, y_{Kt})' \in \mathbb{R}^K \)
- available time series \( y_1, \ldots, y_T \), which is known to be generated by stationary, stable VAR(\( p \)) process

\[
y_t = \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t \tag{1}
\]

where

- \( \nu = (\nu_1, \ldots, \nu_K)' \) is \( K \times 1 \) vector of intercept terms
- \( A_i \) are \( K \times K \) coefficient matrices
- \( u_t \) is white noise with nonsingular covariance matrix \( \Sigma_u \)
- moreover \( p \) presample values for each variable, \( y_{-p+1}, \ldots, y_0 \) are assumed to be available
Notation

\[
Y := (y_1, \ldots, y_T) \quad (K \times T)
\]

\[
B := (\nu, A_1, \ldots, A_p) \quad (K \times (Kp + 1))
\]

\[
Z_t := (1, y_t, y_{t-1}, \ldots, y_{t-p+1})' \quad ((Kp + 1) \times 1)
\]

\[
Z := (Z_0, \ldots, Z_{T-1}) \quad ((Kp + 1) \times T)
\]

\[
U := (u_1, \ldots, u_T) \quad (K \times T)
\]

\[
y := \text{vec}(Y) \quad (KT \times 1)
\]

\[
\beta := \text{vec}(B) \quad (((K^2p + K) \times 1)
\]

\[
b := \text{vec}(B') \quad (((K^2p + K) \times 1)
\]

\[
u := \text{vec}(U) \quad (KT \times 1)
\]
Estimation (1)

- using this notation, the VAR(p) model (1) can be written as

\[ Y = BZ + U, \]

- after application of vec operator and Kronecker product we obtain

\[ \text{vec}(Y) = \text{vec}(BZ) + \text{vec}(U) = (Z' \otimes I_K)\text{vec}(B) + \text{vec}(U), \]

which is equivalent to

\[ y = (Z' \otimes I_K)\beta + u \]

- note that covariance matrix of \( u \) is

\[ \Sigma_u = I_T \otimes \Sigma_u \]
Estimation (2)

- multivariate LS estimation (or GLS estimation) of $\beta$ minimizes

\[
S(\beta) = u'(I_T \otimes \Sigma_u)^{-1}u = \\
= [y - (Z' \otimes I_K)\beta]'(I_T \otimes \Sigma_u^{-1})[y - (Z' \otimes I_K)\beta]
\]

- note that

\[
S(\beta) = y'(I_T \otimes \Sigma_u^{-1})y + \beta'(ZZ' \otimes \Sigma_u^{-1})\beta - 2\beta'(Z \otimes \Sigma_u^{-1})y
\]

- the first order conditions

\[
\frac{\partial S(\beta)}{\partial \beta} = 2(ZZ' \otimes \Sigma_u^{-1})\beta - 2(Z \otimes \Sigma_u^{-1})y = 0
\]

after simple algebraic exercise yield the LS estimator

\[
\hat{\beta} = ((ZZ')^{-1}Z \otimes I_K)y
\]
Estimation (3)

- the Hessian of $S(\beta)$

$$\frac{\partial^2 S}{\partial \beta \partial \beta'} = 2(ZZ' \otimes \Sigma_u^{-1})$$

is positive definite $\Rightarrow \hat{\beta}$ is minimizing vector

- the LS estimator can be written in different ways

$$\hat{\beta} = \beta + ((ZZ')^{-1} Z \otimes I_K)u = vec(YZ'(ZZ')^{-1})$$

- another possible representation is

$$\hat{b} = (I_K \otimes (ZZ')^{-1} Z)vec(Y'),$$

where we can see that multivariate LS estimation is equivalent to OLS estimation of each of the $K$ equations of (1)
Asymptotic Properties (1)

Definition

A white noise process \( u_t = (u_1 t, \ldots, u_K t)' \) is called standard white noise if the \( u_t \) are continuous random vectors satisfying \( E(u_t) = 0 \), \( \Sigma_u = E(u_t u_t) \) is nonsingular, \( u_t \) and \( u_s \) are independent for \( s \neq t \) and

\[
E|u_{it} u_{jt} u_{kt} u_{mt}| \leq c \quad \text{for } i, j, k, m = 1, \ldots, K, \text{ and all } t
\]

for some finite constant \( c \).

we need this property as a sufficient condition for the following results:

\[
\Gamma := \frac{plim}{T} \frac{ZZ'}{T} \text{ exists and is nonsingular}
\]

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \text{vec}(u_t Z_{t-1}') = \frac{1}{\sqrt{T}} (Z \otimes I_K) u \xrightarrow{d} \mathcal{N}(0, \Gamma \otimes \Sigma_u)
\]
Asymptotic Properties (2)

- the above conditions provide for consistency and asymptotic normality of the LS estimator

**Proposition**

**Asymptotic Properties of the LS Estimator**

Let $y_t$ be a stable, $K$-dimensional VAR($p$) process with standard white noise residuals, $\hat{B}$ is the LS estimator of the VAR coefficients $B$. Then

$$\hat{B} \xrightarrow{p} B \quad \text{as} \quad T \to \infty$$

and

$$\sqrt{T} \left( \hat{\beta} - \beta \right) = \sqrt{T} \ \text{vec}(\hat{B} - B) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma_u) \quad \text{as} \quad T \to \infty$$
Proposition

Asymptotic Properties of the White Noise Covariance Matrix Estimators

Let $y_t$ be a stable, $K$-dimensional VAR($p$) process with standard white noise residuals and let $\tilde{B}$ be an estimator of the VAR coefficients $B$ so that $\sqrt{T} \text{vec}(\tilde{B} - B)$ converges in distribution. Furthermore suppose that

$$\tilde{\Sigma}_u = \frac{(Y - \tilde{B}Z)(Y - \tilde{B}Z)'}{T - c},$$

where $c$ is a fixed constant. Then

$$\sqrt{T} \left( \tilde{\Sigma}_u - UU' / T \right) \xrightarrow{p} 0 \quad \text{as} \quad T \to \infty.$$
Example (1)

- three-dimensional system, data for Western Germany (1960-1978)
  - fixed investment $y_1$
  - disposable income $y_2$
  - consumption expenditures $y_3$
Example (2)

- assumption: data generated by VAR(2) process
- LS estimates are the following

\[
\hat{B} = (\hat{\nu}, \hat{A}_1, \hat{A}_2) = YZ'(ZZ')^{-1} = \\
\begin{bmatrix}
-0.017 & -0.320 & 0.146 & 0.961 & -0.161 & 0.115 & 0.934 \\
0.016 & 0.044 & -0.153 & 0.289 & 0.050 & 0.019 & -0.010 \\
0.013 & -0.002 & 0.225 & -0.264 & 0.034 & 0.355 & -0.022
\end{bmatrix}
\]

- stability of estimated process is satisfied, since all roots of the polynomial \(\det(I_3 - \hat{A}_1z - \hat{A}_2z^2)\) have modulus greater than 1
Example (3)

- we can calculate the matrix of t-ratios

\[
\begin{bmatrix}
-0.97 & -2.55 & 0.27 & 1.45 & -1.29 & 0.21 & 1.41 \\
3.60 & 1.38 & -1.10 & 1.71 & 1.58 & 0.14 & -0.06 \\
3.67 & -0.09 & 2.01 & -1.94 & 1.33 & 3.24 & -0.16
\end{bmatrix}
\]

- these quantities can be compared with critical values from a t-distribution
  - d.f. = $KT - K^2p - K = 198$ or d.f. = $T - Kp - 1 = 66$

- for a two-tailed test with significance level 5% we get critical values of approximately ±2 in both cases

- apparently several coefficients are not significant $\Rightarrow$ model contains unnecessarily many free parameters
Small Sample Properties

- difficult to analytically derive small sample properties of LS estimation
- numerical experiments are used, such as Monte Carlo method
- example process

\[
y_t = \begin{pmatrix} 0.02 \\ 0.03 \end{pmatrix} + \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix} y_{t-1} + \begin{pmatrix} 0 & 0 \\ 0.25 & 0 \end{pmatrix} y_{t-2} + u_t
\]

\[
\Sigma_u = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix} \times 10^{-4}
\]

- 1000 time series generated of length \( T = 30 \) (plus 2 presample values)
- \( u_t \sim \mathcal{N}(0, \Sigma_u) \)
## Empirical Results

<table>
<thead>
<tr>
<th>parameter</th>
<th>empirical</th>
<th>MSE</th>
<th>empirical percentiles of t-ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_1 = .02$</td>
<td>.041 .0011</td>
<td>.0015</td>
<td>-1.91 -1.04 -0.64 0.62 1.92 2.29 3.12</td>
</tr>
<tr>
<td>$\nu_2 = .03$</td>
<td>.038 .0005</td>
<td>.0006</td>
<td>-2.30 -1.40 -1.02 0.25 1.65 2.11 2.83</td>
</tr>
<tr>
<td>$\alpha_{11,1} = .5$</td>
<td>.41 .041</td>
<td>.049</td>
<td>-2.78 -2.18 -1.74 -0.43 0.92 1.28 2.01</td>
</tr>
<tr>
<td>$\alpha_{21,1} = .4$</td>
<td>.40 .018</td>
<td>.018</td>
<td>-2.61 -1.74 -1.28 0.04 1.28 1.71 2.65</td>
</tr>
<tr>
<td>$\alpha_{12,1} = .1$</td>
<td>.10 .078</td>
<td>.078</td>
<td>-2.27 -1.67 -1.35 -0.03 1.29 1.67 2.35</td>
</tr>
<tr>
<td>$\alpha_{22,1} = .5$</td>
<td>.44 .030</td>
<td>.034</td>
<td>-2.69 -1.97 -1.59 -0.35 0.89 1.30 2.06</td>
</tr>
<tr>
<td>$\alpha_{11,2} = 0$</td>
<td>-.05 .056</td>
<td>.058</td>
<td>-2.75 -1.93 -1.50 -0.24 1.02 1.38 2.09</td>
</tr>
<tr>
<td>$\alpha_{21,2} = .25$</td>
<td>.29 .023</td>
<td>.024</td>
<td>-1.99 -1.32 -0.99 0.20 1.45 1.81 2.48</td>
</tr>
<tr>
<td>$\alpha_{12,2} = 0$</td>
<td>-.07 .053</td>
<td>.058</td>
<td>-2.48 -1.91 -1.61 -0.28 0.97 1.39 2.03</td>
</tr>
<tr>
<td>$\alpha_{22,2} = 0$</td>
<td>-.01 .023</td>
<td>.024</td>
<td>-2.71 -1.72 -1.36 -0.03 1.18 1.53 2.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>degrees of freedom (d.f.)</th>
<th>percentiles of $t$-distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T - Kp - 1 = 25$</td>
<td>$-2.49 -1.71 -1.32 0 1.32 1.71 2.49$</td>
</tr>
<tr>
<td>$K(T - Kp - 1) = 50$</td>
<td>$-2.41 -1.68 -1.30 0 1.30 1.68 2.41$</td>
</tr>
<tr>
<td>$\infty$ (normal distribution)</td>
<td>$-2.33 -1.65 -1.28 0 1.28 1.65 2.33$</td>
</tr>
</tbody>
</table>
Process with Known Mean (1)

- The process mean $\mu$ is known
- The mean-adjusted VAR($p$) is given by
  
  $$(y_t - \mu) = A_1 (y_{t-1} - \mu) + \ldots + A_p (y_{t-p} - \mu) + u_t$$

- One can use LS estimation by defining:

  $$Y^0 \equiv (y_t - \mu, \ldots, y_T - \mu)$$
  $$Y^0_t \equiv \begin{bmatrix} y_t - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix}$$
  $$X \equiv (Y^0_0, \ldots, Y^0_{T-1})$$
  $$A \equiv (A_1, \ldots, A_p)$$
Process with Known Mean (2)

\[ y^0 \equiv \text{vec}(Y^0) \quad \alpha \equiv \text{vec}(A) \]

- Then the mean-adjusted VAR(p) can be rewritten as:
  \[ Y^0 = AX + U \quad \text{or} \quad y^0 = (X' \otimes I_K) \alpha + u \] where \( u \) is defined as before.

- The LS estimator is:
  \[ \hat{\alpha} = \left( (XX')^{-1} X \otimes I_K \right) y^0 \quad \text{or} \quad \hat{A} = Y^0 X' (XX')^{-1} \]

- If \( y_t \) is stable and \( u_t \) is white noise, it follows that
  \[ \sqrt{T} (\hat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{\alpha}}) \] where \( \hat{\alpha} = \Gamma_Y (0)^{-1} \otimes \Sigma_u \) with
  \[ \Gamma_Y (0) \equiv E \left( Y_t^0 (Y_t^0)' \right). \]
Process with Unknown Mean (1)

- Usually the process mean is not known and we have to estimated it:

\[ \bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t \]

- Plugging in for each \( y_t \) expressed from

\( (y_t - \mu) = A_1 (y_{t-1} - \mu) + \ldots + A_p (y_{t-p} - \mu) + u_t \) and rewriting gives:
Process with Unknown Mean (2)

\[
\bar{y} = \mu + A_1 \left[ \bar{y} + \frac{1}{T} (y_0 - y_T) - \mu \right] + \ldots + \\
A_p \left[ \bar{y} + \frac{1}{T} \left( y_{-p+1} + \ldots + y_0 - y_{T-p+1} - \ldots - y_T \right) - \mu \right] + \frac{1}{T} \sum_{t=1}^{T} u_t 
\]

- The exact meaning of elements such as \( y_{-p+1} \) for \( p > 1 \) is unclear (presample observations).

- Equivalently:

\[
(I_K - A_1 - \ldots - A_p) (\bar{y} - \mu) = \frac{1}{T} z_T + \frac{1}{T} \sum_{t=1}^{T} u_t 
\]

where \( z_T = \sum_{i=1}^{p} A_i \left[ \sum_{j=1}^{i-1} (y_{0-j} - y_{T-j}) \right] \)
Process with Unknown Mean (3)

- Obviously \( E \left( \frac{z_T}{\sqrt{T}} \right) = \frac{1}{\sqrt{T}} E (z_T) = 0 \)

- Moreover, as \( y_t \) is stable \( \text{var} \left( \frac{z_T}{\sqrt{T}} \right) = \frac{1}{T} \text{Var} (z_T) \xrightarrow{T \to \infty} 0 \)

- \( \frac{z_T}{\sqrt{T}} \) converges to zero in mean square and
  \[ (I_K - A_1 - ... - A_p)(\bar{y} - \mu) \] has the same asymptotic distribution as \( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \)

- By the CLT \( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \xrightarrow{d} \mathcal{N} (0, \Sigma_u) \)
therefore, if $y_t$ is stable and $u_t$ is white noise:

$$\sqrt{T} (\bar{y} - \mu) \xrightarrow{d} \mathcal{N} (0, \Sigma_{\bar{y}})$$

with $\Sigma_{\bar{y}} = (I_K - A_1 - \ldots - A_p)^{-1} \Sigma_u (I_K - A_1 - \ldots - A_p)'^{-1}$

another way of estimating the mean is obtained from the LS estimator:

$$\hat{\mu} = \left( I_K - \hat{A}_1 - \ldots - \hat{A}_p \right)^{-1} \hat{\nu}$$

these two ways are asymptotically equivalent
Process with Unknown Mean (5)

- Replacing \( \mu \) with \( \bar{y} \) in the vectors and matrices from before, e.g. 
  \( \hat{Y}^0 \equiv (y_t - \bar{y}, ..., y_T - \bar{y}) \) gives the corresponding LS estimator:
  \[
  \hat{\alpha} = \left( \left( \hat{X} \hat{X}' \right)^{-1} \hat{X} \otimes I_K \right) \hat{y}^0
  \]

- This estimator is asymptotically equivalent to LS estimator for a process with known mean \( \hat{\alpha} \)
  \[
  \sqrt{T} \left( \hat{\alpha} - \alpha \right) \xrightarrow{d} \mathcal{N} \left( 0, \Gamma_Y(0)^{-1} \otimes \Sigma_u \right)
  \]
  with \( \Gamma_Y(0) \equiv E \left( Y_t^0 (Y_t^0)' \right) \)
The Yule-Walker Estimator (1)

- Recall from the lecture slides that for VAR(1) it holds:
  \[ A_1 = \Gamma_y(0)\Gamma_y(1)^{-1} \] and in general \( \Gamma_y(h) = A_1\Gamma_y(h - 1) = A_1^h\Gamma_y(0) \)

- Extending to VAR(p):
  \[
  \begin{bmatrix}
  \Gamma_y(h - 1) \\
  \vdots \\
  \Gamma_y(h - p)
  \end{bmatrix} = [A_1, \ldots, A_p]
  \begin{bmatrix}
  \Gamma_y(0) & \ldots & \Gamma_y(p - 1) \\
  \vdots & \ddots & \vdots \\
  \Gamma_y(-p + 1) & \ldots & \Gamma_y(0)
  \end{bmatrix}
  \]

\[ [\Gamma_y(1), \ldots, \Gamma_y(p)] = [A_1, \ldots, A_p] \]
The Yule-Walker Estimator (2)

\[
[\Gamma_y(1), \ldots, \Gamma_y(p)] = \Lambda \Gamma_Y(0)
\]

- Hence, \( A = [\Gamma_y(1), \ldots, \Gamma_y(p)] \Gamma_Y(0)^{-1} \)

- If \( p \) presample observations are available, the mean \( \mu \) can be estimated by:

\[
\bar{y}^* = \frac{1}{T + p} \sum_{t=-p+1}^{T} y_t
\]

- Then \( \hat{\Gamma}_y(h) = \frac{1}{T + p - h} \sum_{t=-p+h+1}^{T} (y_t - \bar{y}^*) (y_{t-h} - \bar{y}^*)' \)
The Yule-Walker Estimator has the same asymptotic properties as the LS estimator for stable VAR processes. However, it could be less attractive for small samples. The following example shows that asymptotically equivalent estimators can give different results for small samples (here \(T=73\))

\[
\tilde{y} = \begin{bmatrix} 0.018 \\ 0.020 \\ 0.020 \end{bmatrix} \quad \hat{\mu} = \left( I_3 - \hat{A}_1 - \hat{A}_2 \right)^{-1} \hat{\nu} = \begin{bmatrix} 0.017 \\ 0.020 \\ 0.020 \end{bmatrix}
\]
The Yule-Walker Estimator (4)

\[
\hat{A} = \left( \hat{A}_1, \hat{A}_2 \right) = \begin{bmatrix}
-0.319 & 0.143 & 0.960 & -0.160 & 0.112 & 0.933 \\
0.044 & -0.153 & 0.288 & 0.050 & 0.019 & -0.010 \\
-0.002 & 0.224 & -0.264 & 0.034 & 0.354 & -0.023
\end{bmatrix}
\]

\[
\hat{A}_{YW} = \begin{bmatrix}
-0.319 & 0.147 & 0.959 & -0.160 & 0.115 & 0.932 \\
0.044 & -0.152 & 0.286 & 0.050 & 0.020 & -0.012 \\
-0.002 & 0.225 & -0.264 & 0.034 & 0.355 & -0.022
\end{bmatrix}
\]
Maximum Likelihood Estimation

- Assume that the VAR(p) is Gaussian, i.e.

\[ u = \text{vec}(U) = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix} \sim \mathcal{N}(0, I_T \otimes \Sigma_u) \]

- The probability density of \( u \) is

\[
f_u(u) = \frac{1}{(2\pi)^{KT/2} |I_T \otimes \Sigma_u|^{-1/2}} \exp \left[ -\frac{1}{2} u' \left( I_T \otimes \Sigma_u^{-1} \right) u \right]
\]
The Log-Likelihood Function

- From the probability density of \( \mathbf{u} \), a probability density for \( \mathbf{y} \equiv \text{vec} (\mathbf{Y}) \), \( f_\mathbf{y} (\mathbf{y}) \), can be derived.
- After some modification the log-likelihood function is given by:

\[
\ln l (\mu, \alpha, \sum_u) = -\frac{K T}{2} \ln 2\pi - \frac{T}{2} \ln |\sum_u| - \frac{1}{2} \text{tr} \left[ (\mathbf{Y}^0 - \mathbf{A} \mathbf{X})' \sum_u^{-1} (\mathbf{Y}^0 - \mathbf{A} \mathbf{X}) \right]
\]

- From \( \frac{\partial \ln(l)}{\partial \mu} \), \( \frac{\partial \ln(l)}{\partial \alpha} \), and \( \frac{\partial \ln(l)}{\partial \sum_u} \) we get the system of normal equations, which can be solved for the estimators.
The three ML Estimators:

\[
\begin{align*}
\tilde{\mu} &= \frac{1}{T} \left( I_K - \sum_{i=1}^{p} \tilde{A}_i \right)^{-1} \sum_{t=1}^{T} \left( y_t - \sum_{i=1}^{p} \tilde{A}_i y_{t-i} \right) \\
\tilde{\alpha} &= \left( (\tilde{X} \tilde{X}')^{-1} \tilde{X} \otimes I_K \right) (y - \tilde{\mu}^*) \\
\tilde{\Sigma}_u &= \frac{1}{T} \left( \tilde{Y}^0 - \tilde{A} \tilde{X} \right) \left( \tilde{Y}^0 - \tilde{A} \tilde{X} \right)'
\end{align*}
\]
The estimators are asymptotically consistent and asymptotically normal distributed:

\[
\sqrt{T} \begin{bmatrix}
\tilde{\mu} - \mu \\
\tilde{\alpha} - \alpha \\
\tilde{\sigma} - \sigma
\end{bmatrix} \xrightarrow{d} \mathcal{N} \left( 0, \begin{bmatrix}
\Sigma_{\tilde{\mu}} & 0 & 0 \\
0 & \Sigma_{\tilde{\alpha}} & 0 \\
0 & 0 & \Sigma_{\tilde{\sigma}}
\end{bmatrix} \right)
\]

where \( \tilde{\sigma} = \text{vech} \left( \tilde{\Sigma}_u \right) \) and \( \Sigma_{\tilde{\mu}} = \left( I_K - \sum_{i=1}^p \tilde{A}_i \right)^{-1} \sum_u \left( I_K - \sum_{i=1}^p \tilde{A}_i' \right)^{-1} \)
Properties of the ML Estimator (2)

- $\Sigma \tilde{\alpha} = \Gamma_Y (0)^{-1} \otimes \Sigma_u$
- $\Sigma \tilde{\sigma} = 2D_K^+ (\Sigma_u \otimes \Sigma_u) \left( D_K^+ \right)'$

where $D_K$ is given by $\text{vec} (\Sigma_u) = D_K \text{vech} (\Sigma_u)$ and $D_K^+$ is the Moore-Penrose generalized inverse.

$\sigma = \text{vech} (\Sigma_u) = \text{vech} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{31} \\ \sigma_{22} \\ \sigma_{32} \\ \sigma_{33} \end{bmatrix}$
Thank you for your attention!