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

Contents

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

$$y_t = \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t \quad (1)$$

  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 Σ_u
- moreover p presample values for each variable, $y_{-p+1}, \ldots, y_0$ are assumed to be available

Notation

$$
\begin{aligned}
Y &:= (y_1, \dots, y_T) && (K \times T) \\
B &:= (\nu, A_1, \dots, A_p) && (K \times (Kp+1)) \\
Z_t &:= (1, y_t, y_{t-1}, \dots, y_{t-p+1})' && ((Kp+1) \times 1) \\
Z &:= (Z_0, \dots, Z_{T-1}) && ((Kp+1) \times T) \\
U &:= (u_1, \dots, u_T) && (K \times T) \\
\mathbf{y} &:= vec(Y) && (KT \times 1) \\
\beta &:= vec(B) && ((K^2p+K) \times 1) \\
\mathbf{b} &:= vec(B') && ((K^2p+K) \times 1) \\
\mathbf{u} &:= vec(U) && (KT \times 1)
\end{aligned}
$$

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

$$vec(Y) = vec(BZ) + vec(U) = (Z' \otimes I_K)vec(B) + vec(U),$$

which is equivalent to

$$\mathbf{y} = (Z' \otimes I_K)\beta + \mathbf{u}$$

- note that covariance matrix of **u** is

$$\Sigma_{\mathbf{u}} = I_T \otimes \Sigma_u$$

Estimation (2)

- multivariate LS estimation (or GLS estimation) of β minimizes

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

- note that

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

- the first order conditions

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

after simple algebraic exercise yield the LS estimator

$$\hat{\beta} = ((ZZ')^{-1}Z \otimes I_K)\mathbf{y}$$

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 differen ways

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

- another possible representation is

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

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*

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

for some finite constant c.

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

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

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} vec(u_t Z'_{t-1}) = \frac{1}{\sqrt{T}} (Z \otimes I_K)\mathbf{u} \xrightarrow[T \to \infty]{d} \mathcal{N}(0, \Gamma \otimes \Sigma_u)$$

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

$$\hat{B} \xrightarrow[T\to\infty]{p} B$$

and

$$\sqrt{T}\,(\hat{\beta} - \beta) = \sqrt{T}\,vec(\hat{B} - B) \xrightarrow[T\to\infty]{d} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma_u)$$

.

Asymptotic Properties (3)

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 $\bar{B}$ be an estimator of the VAR coefficients B so that $\sqrt{T}\, vec(\bar{B} - B)$ converges in distribution. Furthermore suppose that

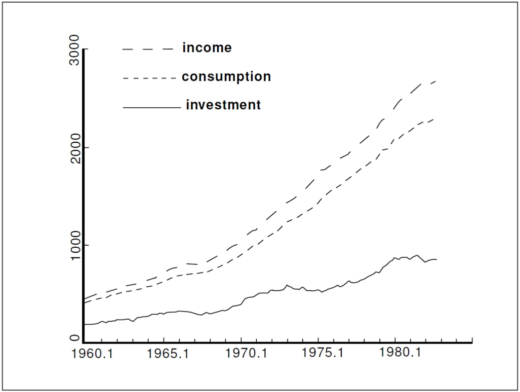
$$\bar{\Sigma}_u = \frac{(Y - \bar{B}Z)(Y - \bar{B}Z)'}{T - c},$$

where c is a fixed constant. Then

$$\sqrt{T}\,(\bar{\Sigma}_u - UU'/T) \xrightarrow[T \to \infty]{p} 0.$$

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

$$
\begin{aligned}
\widehat{B} &= (\widehat{\nu}, \widehat{A}_1, \widehat{A}_2) = YZ'(ZZ')^{-1} \\
&= \begin{bmatrix} -.017 & -.320 & .146 & .961 & -.161 & .115 & .934 \\ .016 & .044 & -.153 & .289 & .050 & .019 & -.010 \\ .013 & -.002 & .225 & -.264 & .034 & .355 & -.022 \end{bmatrix}
\end{aligned}
$$

- stability of estimated process is satisfied, since all roots of the polynomial $\det(I_3 - \hat{A}_1 z - \hat{A}_2 z^2)$ 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 do 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

	empirical			empirical percentiles of t-ratios						
parameter	mean	variance	MSE	1.	5.	10.	50.	90.	95.	99.
$\nu_1 = .02$	.041	.0011	.0015	−1.91	−1.04	−0.64	0.62	1.92	2.29	3.12
$\nu_2 = .03$	.038	.0005	.0006	−2.30	−1.40	−1.02	0.25	1.65	2.11	2.83
$\alpha_{11,1} = .5$	.41	.041	.049	−2.78	−2.18	−1.74	−0.43	0.92	1.28	2.01
$\alpha_{21,1} = .4$	.40	.018	.018	−2.61	−1.74	−1.28	0.04	1.28	1.71	2.65
$\alpha_{12,1} = .1$	.10	.078	.078	−2.27	−1.67	−1.35	−0.03	1.29	1.67	2.38
$\alpha_{22,1} = .5$	.44	.030	.034	−2.69	−1.97	−1.59	−0.35	0.89	1.30	2.06
$\alpha_{11,2} = 0$	−.05	.056	.058	−2.75	−1.93	−1.50	−0.24	1.02	1.38	2.09
$\alpha_{21,2} = .25$	.29	.023	.024	−1.99	−1.32	−0.99	0.20	1.45	1.81	2.48
$\alpha_{12,2} = 0$	−.07	.053	.058	−2.48	−1.91	−1.61	−0.28	0.97	1.39	2.03
$\alpha_{22,2} = 0$	−.01	.023	.024	−2.71	−1.72	−1.36	−0.03	1.18	1.53	2.18
	degrees of freedom(d.f.)			percentiles of t-distributions						
				1.	5.	10.	50.	90.	95.	99.
	$T - Kp - 1 = 25$			−2.49	−1.71	−1.32	0	1.32	1.71	2.49
	$K(T - Kp - 1) = 50$			−2.41	−1.68	−1.30	0	1.30	1.68	2.41
	∞ (normal distribution)			−2.33	−1.65	−1.28	0	1.28	1.65	2.33

Process with Known Mean (1)

- The process mean μ is known
- The mean-adjusted VAR(p) is given by

$$(y_t - \mu) = A_1 (y_{t-1} - \mu) + ... + A_p (y_{t-p} - \mu) + u_t$$

- One can use LS estimation by defining:

$$Y^0 \equiv (y_t - \mu, ..., y_T - \mu) \qquad A \equiv (A_1, ..., A_p)$$

$$Y_t^0 \equiv \begin{bmatrix} y_t - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix} \qquad X \equiv \left(Y_0^0, ..., Y_{T-1}^0 \right)$$

Process with Known Mean (2)

$\mathbf{y^0} \equiv vec(Y^0)$ $\alpha \equiv vec(A)$

- Then the mean-adjusted VAR(p) can be rewritten as:

$Y^0 = AX + U$ or $\mathbf{y}^0 = (X' \otimes I_K)\,\alpha + \mathbf{u}$ where $\mathbf{u}$ is defined as before.

- The LS estimator is:

$\hat{\alpha} = \left((XX')^{-1} X \otimes I_K\right) \mathbf{y^0}$ or $\hat{A} = Y^0 X' (XX')^{-1}$

- 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 $(y_t - \mu) = A_1 (y_{t-1} - \mu) + ... + A_p (y_{t-p} - \mu) + u_t$ and rewriting gives:

Process with Unknown Mean (2)

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

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

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

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 $E\left(z_T/\sqrt{T}\right) = \frac{1}{\sqrt{T}} E\left(z_T\right) = 0$
- Moreover, as y_t is stable $var\left(z_T/\sqrt{T}\right) = \frac{1}{T} Var\left(z_T\right) \underset{T\to\infty}{\to} 0$
- $z_T/\sqrt{T}$ converges to zero in mean square and $\left(I_K - A_1 - ... - A_p\right)\left(\bar{y} - \mu\right)$ has the same asymptotic distribution as $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t$
- By the CLT $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t \xrightarrow{d} \mathcal{N}\left(0, \Sigma_u\right)$

Process with Unknown Mean (4)

- 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 $\Sigma_{\bar{y}} = (I_K - A_1 - ... - A_p)^{-1}\,\Sigma_u\,(I_K - A_1 - ... - A_p)^{\prime -1}$

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

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

- these two ways are asymptotically equivalent

Process with Unknown Mean (5)

- Replacing μ with $\bar{y}$ in the vectors and matrices from before, e.g. $\widehat{Y}^0 \equiv (y_t - \bar{y}, ..., y_T - \bar{y})$ gives the corresponding LS estimator:

$$\hat{\hat{\alpha}} = \left(\left(\widehat{X}\widehat{X}' \right)^{-1} \widehat{X} \otimes I_K \right) \widehat{\mathbf{y}}^{\mathbf{0}}$$

- This estimator is asymptotically equivalent to LS estimator for a process with known mean $\hat{\alpha}$

$$\sqrt{T} \left(\hat{\hat{\alpha}} - \alpha \right) \xrightarrow{d} \mathcal{N} \left(0, \Gamma_Y (0)^{-1} \otimes \Sigma_u \right)$$

with $\Gamma_Y (0) \equiv E \left(Y_t^0 \left(Y_t^0 \right)' \right)$

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): $\Gamma_y(h) = [A_1, ..., A_p] \begin{bmatrix} \Gamma_y(h-1) \\ \vdots \\ \Gamma_y(h-p) \end{bmatrix}$ or:

$$[\Gamma_y(1), ..., \Gamma_y(p)] = [A_1, ..., A_p] \begin{bmatrix} \Gamma_y(0) & \dots & \Gamma_y(p-1) \\ \vdots & \ddots & \vdots \\ \Gamma_y(-p+1) & \dots & \Gamma_y(0) \end{bmatrix}$$

The Yule-Walker Estimator (2)

$$[\Gamma_y(1), ..., \Gamma_y(p)] = A\Gamma_Y(0)$$

- Hence, $A = [\Gamma_y(1), ..., \Gamma_y(p)]\,\Gamma_Y(0)^{-1}$
- If p presample observations are available, the mean μ 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 (3)

- 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)

$$\bar{y} = \begin{bmatrix} 0.018 \\ 0.020 \\ 0.020 \end{bmatrix} \qquad \hat{\mu} = \left(I_3 - \hat{A}_1 - \hat{A}_2\right)^{-1} \hat{\nu} = \begin{bmatrix} 0.017 \\ 0.020 \\ 0.020 \end{bmatrix}$$

The Yule-Walker Estimator (4)

$$\widehat{\widehat{A}} = \left(\widehat{\widehat{A}}_1, \widehat{\widehat{A}}_2\right) = \begin{bmatrix} -.319 & .143 & .960 & -.160 & .112 & .933 \\ .044 & -.153 & .288 & .050 & .019 & -.010 \\ -.002 & .224 & -.264 & .034 & .354 & -.023 \end{bmatrix}$$

$$\hat{A}_{YW} = \begin{bmatrix} -.319 & .147 & .959 & -.160 & .115 & .932 \\ .044 & -.152 & .286 & .050 & .020 & -.012 \\ -.002 & .225 & -.264 & .034 & .355 & -.022 \end{bmatrix}$$

Maximum Likelihood Estimation

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

$$\mathbf{u} = vec\left(U\right) = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix} \sim \mathcal{N}\left(0, I_T \otimes \Sigma_u\right)$$

- The probability density of $\mathbf{u}$ is

$$f_{\mathbf{u}}(\mathbf{u}) = \frac{1}{(2\pi)^{KT/2}} \left|I_T \otimes \sum_u\right|^{-1/2} exp\left[-\frac{1}{2}\mathbf{u}' \left(I_T \otimes \sum_u^{-1}\right) \mathbf{u}\right]$$

The Log-Likelihood Function

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

$$ln \quad l\,(\mu, \alpha, \sum_u) =$$
$$-\frac{KT}{2} ln 2\pi - \frac{T}{2} ln \left|\sum_u\right| - \frac{1}{2} tr \left[\left(Y^0 - AX\right)' \sum_u^{-1} \left(Y^0 - AX\right)\right]$$

- 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 ML Estimators

- The three ML Estimators:

$$\tilde{\mu} = \frac{1}{T}\left(I_K - \sum_{i=1}^{p}\tilde{A}_i\right)^{-1}\sum_{t=1}^{T}\left(y_t - \sum_{i=1}^{p}\tilde{A}_i y_{t-i}\right)$$

$$\tilde{\alpha} = \left(\left(\widetilde{X}\widetilde{X}'\right)^{-1}\widetilde{X}\otimes I_K\right)(y - \widetilde{\mu}^*)$$

$$\tilde{\Sigma}_u = \frac{1}{T}\left(\widetilde{Y}^0 - \widetilde{A}\widetilde{X}\right)\left(\widetilde{Y}^0 - \widetilde{A}\widetilde{X}\right)'$$

Properties of the ML Estimator (1)

- 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 $\tilde{\sigma} = vech\left(\tilde{\Sigma}_u\right)$ 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 $vec(\Sigma_u) = D_K vech(\Sigma_u)$ and D_K^+ is the Moore-Penrose generalized inverse.

$$\sigma = vech(\Sigma_u) = vech \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{31} \\ \sigma_{22} \\ \sigma_{32} \\ \sigma_{33} \end{bmatrix}$$

Thank you for your attention!