VAR Processes with Linear Constraints

Anna Pauls Anja Breitwieser

Vienna: January 17, 2012

Contents

Contents II

K-dimensional stationary, stable VAR(p) process:

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

with

$v = (v_1, ...v_K)'$ a $(K \times 1)$ vector of intercepts

A_i $(k \times K)$ coefficient matrices and

u_t white noise (with a non-singular covariance matrix Σ_u

$\rightarrow$ same assumptions as in the unconstrained case (see: Lütkepohl, chapter 3)

Compact Form

in compact form model (1) can be written as:

$Y = BZ + U$ with

$$Y \equiv [y_1, \ldots, y_T], Z \equiv [Z_0, \ldots, Z_{T-1}] \text{ with } Z_t \equiv \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix},$$

$$B \equiv [v, A_1, \ldots, A_p], U \equiv [u_1, \ldots, u_T]$$

constraints for B: $\beta \equiv vec(B) = R\gamma + r$

β	$(K(Kp+1) \times 1)$ matrix
R	$(K(Kp+1) \times M)$ matrix
γ	$(M \times 1)$ vector
r	$(M \times 1)$ vector

by vectorizing and plugging in $R\gamma + r$ for β we get:

$$\boldsymbol{y} \equiv vec(Y) = (Z' \otimes I_K)vec(B) + vec(U)$$
$$= (Z' \otimes I_K)(R\gamma + r) + \boldsymbol{u}$$

or

$$\boldsymbol{z} = (Z' \otimes I_K)R\gamma + \boldsymbol{u}$$

with $\boldsymbol{z} \equiv \boldsymbol{y} - (Z' \otimes I_K)r$

and $\boldsymbol{u} \equiv vec(U)$

$\rightarrow$ while there are other forms of representing linear constraints, the chosen form is convenient as it allows to derive the estimators in the same manner as in the unconstrained model (see: Lütkepohl, chapter 3) minimizing:

$$S(\gamma) = \boldsymbol{u'}\left(I_T \otimes \Sigma_u^{-1}\right)\boldsymbol{u}$$
$$= [\boldsymbol{z} - (Z' \otimes I_K)R\gamma]'(I_T \otimes \Sigma_u^{-1})[\boldsymbol{z} - (Z' \otimes I_K)R\gamma]$$

yields the GLS estimator for γ:

$$\hat{\gamma} = [R'(ZZ' \otimes \Sigma_u^{-1})R]^{-1}R'(Z \otimes \Sigma_u^{-1})\,\boldsymbol{z}$$

asymptotic properties:

Under the condition, that y_t is a K-dimensional stable,stationary VAR(p) process, u_t is white noise with bounded fourth moments and $\beta = R\gamma + r$ with $rk(R) = M$

- $\hat{\gamma}$ is a consistent estimator of γ
- $\sqrt{T}(\hat{\gamma} - \gamma)$ is asymptotically normally distributed with covariance matrix: $[R'(\Gamma \otimes \Sigma_u^{-1})R]^{-1}$

In practice an Σ_u is not known and has to be estimated. One consistent estimator of Σ_u is:

$$\hat{\Sigma}_u = \frac{1}{T-Kp-1}(Y - \hat{B}Z)(Y - \hat{B}Z)'$$

where $\hat{B} = YZ'(ZZ')^{-1}$ is the unconstraint multivariate LS estimator of the coefficient matrix B

$$\rightarrow \hat{\hat{\gamma}} = [R'(ZZ' \otimes \hat{\Sigma}_u^{-1})R]^{-1}R'(Z \otimes \hat{\Sigma}_u^{-1})\boldsymbol{z}$$

asymptotic properties:

Under the previous conditions and if plim$\hat{\Sigma} = \Sigma$ the EGLS estimator is asymptotically equivalent to the GLS estimator

$\rightarrow$ from $\hat{\hat{\gamma}}$ the implied restricted EGLS estimator $\hat{\hat{\beta}}$ can be obtained:

$$\hat{\hat{\beta}} = R\hat{\hat{\gamma}} + r$$

asymptotic properties of:

- $\hat{\hat{\beta}}$ is a consistent
- $\sqrt{T}(\hat{\hat{\beta}} - \beta)$ is asymptotically normally distributed with covariance matrix: $R[R'(\Gamma \otimes \Sigma_u^{-1})R]^{-1}R'$

constrained vs. unconstrained

How does the covariance matrix of the restricted estimator compare to the covariance matrix of the unrestricted estimator?

Under the assumption that the restrictions are valid:

- the asymptotic variances of the restricted estimator is smaller or equal to the asymptotic variances of the unrestricted estimator
- asymptotic efficiency gains

optimal $h-step$ forecast of process (1):

$$y_t(h) = v + A_1 y_t(h-1) + ... + A_p y_t(h-p)$$
with $y_t(j) = y_{t+j}$ for $J \leq 0$

replacing true coefficients by estimators one gets (i.e. $B = (v, A_1, ..., A_p)$ with $B = (\hat{v}, \hat{A}_1, ..., \hat{A}_p)$:

$$\hat{y}_t(h) = \hat{v} + \hat{A}_1 \hat{y}_t(h-1) + ... + \hat{A}_p \hat{y}_t(h-p)$$
with $\hat{y}_t(j) = y_{t+j}$ for $J \leq 0$

and the forecast error matrix: $\Sigma_{\hat{y}}(h) = \Sigma_y(h) + MSE[y_t(h) - \hat{y}_t(h)]$

with $\Sigma_y(h) = \sum_{i=0}^{h-1} \phi_i \Sigma_u \phi_i'$

under the assumption, that only data up to the forecast origin is used for estimation it can be approximated by: $\Sigma_{\hat{y}}(h) = \Sigma_y(h) + \frac{1}{T}\Omega(h)$

$\rightarrow$ general results from forecasting without linear constraints remain valid,as before we get:

$$\Sigma_{\hat{y}}(h) = \Sigma_y(h) + \frac{1}{T}\Omega(h)$$

with $\Sigma_y(h)$ and $\Omega(h) \equiv E[(\partial y_t(h)/\partial\beta')\Sigma_{\hat{\beta}}(\partial y_t(h)'/\partial\beta)]$

in the case with parameter restrictions, $\Sigma_{\hat{\beta}}$ has the following form (compared to the unrestricted case, where $\Sigma_{\hat{\beta}} = \Gamma \otimes \Sigma_u$) :

$$\Sigma_{\hat{\beta}} = R[R'(\Gamma \otimes \Sigma_u^{-1})R]^{-1}R'$$

$\rightarrow$ the covariance matrix is (under the assumption that the restrictions are valid) smaller in the restricted case than in the unrestricted case, hence $\Omega(h)$ will also become smaller

Specification of Subset VAR Models

A model with zero constraints on the coefficients is a subset model of the general VAR model

Zero restrictions can be written formally as $r = \overline{r} = 0$

As the choice of restrictions may not always be undebatable, statistical procedures may be used to detect possible zero constraints or confirm intuition

One common solution: Fit all possible subsets of a VAR(p) process with p known and select the one that optimizes the chosen criterion.

- For instance,modified AIC, SC or HQ may be used

Choosing the constraints

Subsets with j elements from K^2p coefficients can be chosen, such that a total of

$\sum_{j=0}^{K^2p-1} \binom{K^2p}{j}$ VAR models, that is

$\binom{K^2p}{j}$ subsets from K^2p coefficients have to be estimated and compared.

There are several approaches to reduce the number of potential models.

$AIC = ln\tilde{\sigma}^2 + \frac{2}{T}$ (number of estimated parameters)

$BIC = ln\tilde{\sigma}^2 + \frac{lnT}{T}$ (number of estimated parameters)

$HQ = ln\tilde{\sigma}^2 + \frac{2lnlnT}{T}$ (number of estimated parameters)

with $ln\tilde{\sigma}^2$ being the sum of squared estimation residuals

AIC in the restricted case

The $k-th$ equation of the system may be written as

$$yk = \begin{bmatrix} y_{k1} \\ \vdots \\ y_{kT} \end{bmatrix} = Z'b_k + u_k = Z'\overline{R}_k c_k$$

$$\hat{c}_k = (\overline{R}_k' ZZ' \overline{R}_k)^{-1} \overline{R}_k' Z y_{(k)} + u_{(k)}$$

A corresponding estimator for the residual variance is

$$\tilde{\sigma}^2(\overline{R}_k) = (y_{(k)} - Z'\hat{b}_k)/T$$

$$AIC(\overline{R}_k) = ln\tilde{\sigma}^2(\overline{R}_k) + \frac{2}{T} rk(\overline{R}_k)$$

Several zero restrictions for example are not linear independent $\Rightarrow \downarrow \frac{2}{T} rk(\overline{R}_k)$

but if the eliminated coefficients are significantly different from zero, $\uparrow ln\tilde{\sigma}^2(\overline{R}_k)$ and $AIC(\overline{R}_k)$ will not become smaller for these restrictions.

Elimination of complete matrices

Penm & Terrel (1982)

Number of models to be compared:

$\sum_{j=0}^{p} \binom{p}{j}$ from $K^2pj = 2^p$

Useful approach for data with strong seasonality, but somewhat strong assumption while at the same time precluding further zero coefficients.

Top-Down Strategy

The unrestricted model with $\overline{R}_k = I_{(Kp+1)}$ is estimated first and the corresponding $AIC(I_{(Kp+1)})$ obtained

Restriction by eliminating the last column of $I_{(Kp+1)} \to \overline{R}_k^{(1)}$ if

$$AIC(\overline{R}_k^{(1)}) \leq AIC(I_{(Kp+1)})$$

$\Longrightarrow$ go on like this!

Bottom-Up Strategy

Considering lags of each variable separately and choosing its lag order based on the minimization of some model selection criteria, that is

$$y_{kt} = v_k + \alpha_{k1,1} y_{1,t-1} + \ldots + \alpha_{1,t-n} + u_{kt}$$

with n ranging from zero to some prespecified upper pound p

Variables are added sequentially, the variables that have been evaluated already are held fixed:

$$y_{kt} = v_k + \alpha_{k1,1} y_{1,t-1} + \ldots + \alpha_{1,t-p_1} + \alpha_{k2,1} y_{2,t-1} + \ldots + \alpha_{2,t-p_2} + u_{kt}$$

This can be combined with the top-down approach to account for the problem of omitted variable effects that may lead to overstatement of some lag lengths in the final equation

Sequential elimination of regressors

Individual zero coefficients are chosen on the basis of the t-ratios of the parameter estimators until all t-ratios are greater than some threshold value in absolute value

If the threshold value is chosen accordingly this is equivalent to sequential elimination of regressors by minimizing some model selection criteria

Residual autocovariances and autocorrelations

$$\hat{c}_h := vec(\hat{C}_h),$$

$$\hat{C}_h := (\hat{C}_1, ..., \hat{C}_h)$$

$$\hat{C}_i := \frac{1}{T} \sum_{t=i+1}^{T} \hat{u}_t \hat{u}'_{t-i} i = 0, 1, ..., h$$

$$\hat{r}_h := vec(\hat{R}_h),$$

$$\hat{R}_h := (\hat{R}_1, ..., \hat{R}_h)$$

$$\hat{R}_i := \hat{D}^{-1} \hat{C}_i \hat{D}^{-1}, i = 0, 1, ..., h$$

with $\hat{D}$ being a diagonal matrix with the square roots of the diagonal elements of $\hat{C}$ on the diagonal

The Portmanteau statistic

$$Q_h := T\Sigma_{i=1}^{h} tr(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1})$$

$$= T\hat{c}_h'(I_h \otimes \hat{C}_0^{-1} \otimes \hat{C}^{-1})\hat{c}_h$$

has a different asymptotic distribution than in the unrestricted case

$\lambda_{LM}(h)$ converges in distribution to $\chi^2(hK^2)$
which is the same asymptotic distribution as in the unrestricted case

Estimation of the logarithmized investment, income and consumption data

1961.2 - 1978.4

$T = 71$

Top-down strategy

VAR-order $p = 4$

EGLS estimates

Table 5.1. EGLS estimates of subset VAR models for the investment/income/consumption data

	model selection criterion	
	AIC	HQ-SC
$\widehat{\nu}$	$\begin{bmatrix} \underset{(.006)}{.015^*} \\ \underset{(.003)}{.015} \\ \underset{(.003)}{.013} \end{bmatrix}$	$\begin{bmatrix} \underset{(.006)}{.015} \\ \underset{(.001)}{.020} \\ \underset{(.003)}{.016} \end{bmatrix}$
$\widehat{A}_1$	$\begin{bmatrix} \underset{(.104)}{-.219} & 0 & 0 \\ 0 & 0 & \underset{(.133)}{.235} \\ 0 & \underset{(.082)}{.274} & \underset{(.116)}{-.391} \end{bmatrix}$	$\begin{bmatrix} \underset{(.104)}{-.225} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \underset{(.081)}{.261} & \underset{(.095)}{-.439} \end{bmatrix}$
$\widehat{A}_2$	$\begin{bmatrix} 0 & 0 & 0 \\ \underset{(.024)}{.010} & 0 & 0 \\ 0 & \underset{(.073)}{.335} & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \underset{(.074)}{.329} & 0 \end{bmatrix}$
$\widehat{A}_3$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \underset{(.076)}{.095} & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
$\widehat{A}_4$	$\begin{bmatrix} \underset{(.103)}{.340} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} \underset{(.103)}{.331} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
	$\bar{Q}_{12} = 79.3\ [.937]^{**}$	$\bar{Q}_{12} = 85.5\ [.893]$
	$\bar{Q}_{20} = 144\ [.943]$	$\bar{Q}_{20} = 152\ [.898]$

*Estimated standard errors in parentheses.
**p-value.

Estimated residual autocorrelations

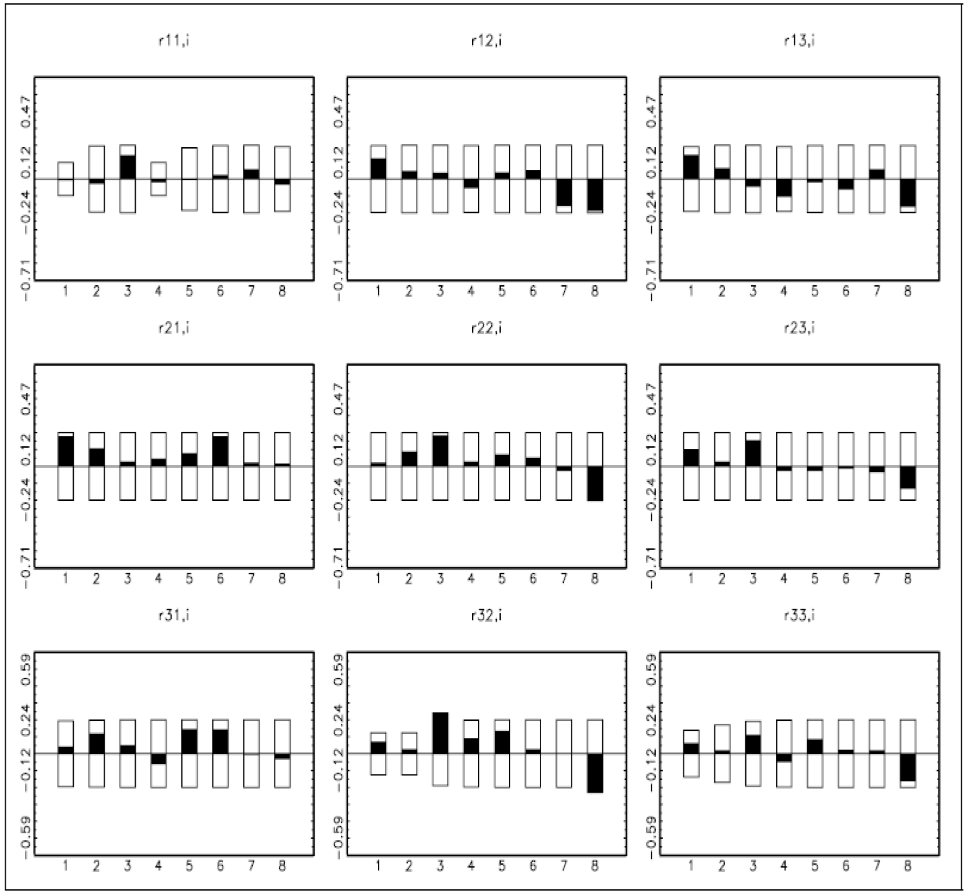

Fig. 5.1. Estimated residual autocorrelations of the investment/income/consumption HQ-SC subset VAR model with estimated asymptotic two-standard error bounds.

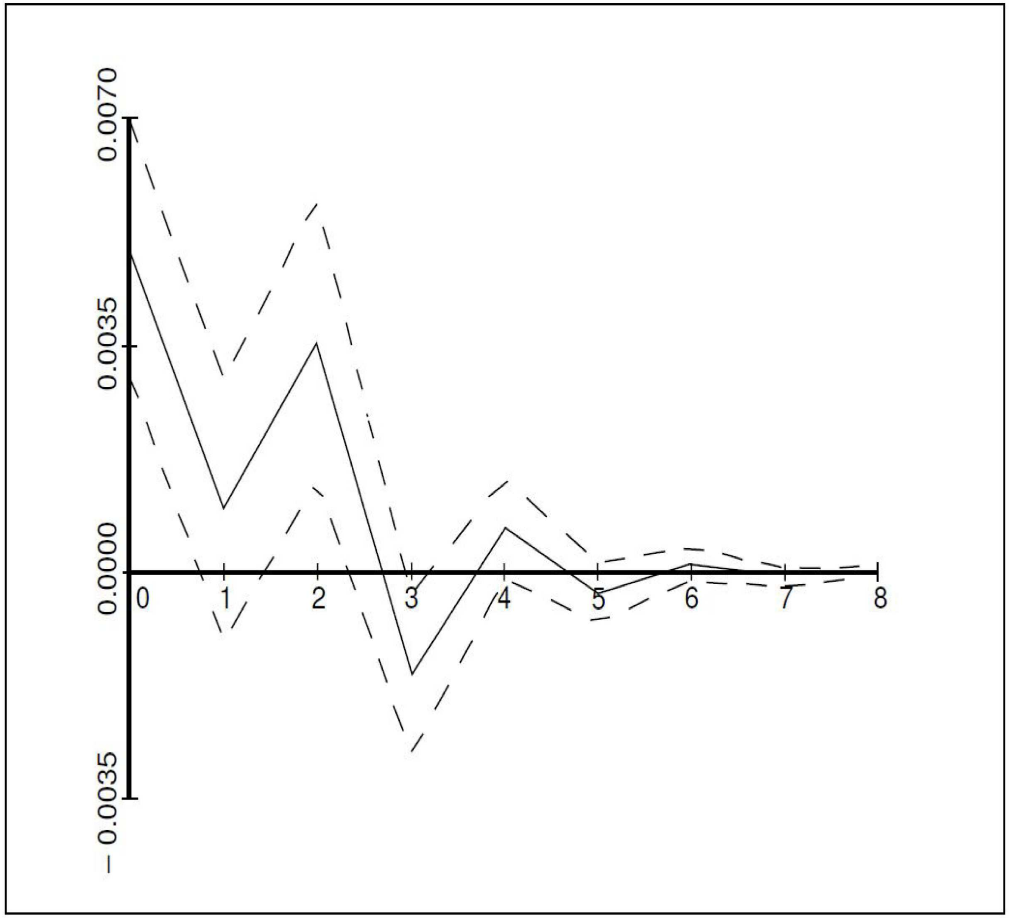

Fig. 5.2. Estimated responses of consumption to an orthogonalized impulse in income with two-standard error bounds based on the HQ-SC subset VAR model.

Table 5.2. Point and interval forecasts from full and subset VAR(4) models for the investment/income/consumption example

variable	forecast horizon	full VAR(4) point forecast	full VAR(4) 95% interval forecast	HQ-SC subset VAR(4) point forecast	HQ-SC subset VAR(4) 95% interval forecast
investment	1	.006	[−.091, .103]	.015	[−.074, .105]
	2	.025	[−.075, .125]	.023	[−.068, .115]
	3	.028	[−.071, .126]	.018	[−.073, .110]
	4	.026	[−.074, .125]	.023	[−.069, .115]
income	1	.021	[−.005, .047]	.020	[−.004, .044]
	2	.022	[−.004, .049]	.020	[−.004, .044]
	3	.017	[−.009, .043]	.020	[−.004, .044]
	4	.022	[−.004, .049]	.020	[−.004, .044]
consumption	1	.022	[.001, .042]	.023	[.004, .042]
	2	.015	[−.006, .036]	.013	[−.007, .033]
	3	.020	[−.004, .043]	.022	[.001, .044]
	4	.019	[−.004, .042]	.018	[−.004, .040]

Example - Interpretation

Causality checking has been built into the model selection procedure: income and consumption are not Granger causal for investment

Although theoretically, the more parsimonious model chosen by SC should be more precise than the one chosen by AIC that has more parameters if the restrictions are correct, this is not the case here. Lütkepohl argues that this is not a big issue here because the estimation has been made on the basis of a single realization of an unknown data generation process

Thank you for your attention!