# VAR Order Selection
# &
# Checking the Model Adequacy

**Mozhgan Raeisian**

**Shima Goudarzi**

**Mike Bronner**

**17 Jan 2012**

# Content

- **Introduction**
- **A Sequence of Tests for Determining the VAR Order**
  - The Impact of the Fitted VAR Order on the Forecast MSE
  - The Likelihood Ratio Test Statistic
  - A Testing Scheme for VAR Order Determination
- **Criteria for VAR Order Selection**
  - Minimizing the Forecast MSE
  - Consistent Order Selection
- **Checking the Whiteness of the Residuals**
- **Testing for Nonnormality**
  - Tests for Nonnormality of a Vector White Noise Process
  - Tests for Nonnormality of a VAR Process
- **Tests for Structural Change**
  - Chow Tests
  - Forecast Tests for Structural Change

# Introduction

- Assume a K-dimentional multiple time series $y_1, \dots, y_T$ with $y_t = (y_{1t}, \dots, y_{Kt})'$, known to be generated by VAR (p) process,

$$y_t = \nu + A_1 Y_{t-1} + \dots + A_P y_{t-p} + u_t \qquad (1)$$

Deriving the properties of the estimators, a number of assumptions were made. As it will rarely be known with certainty whether the conditions hold that are required to derive consistency and asymptotic normality of the estimators, some statistical tools should be used to check the validity of the assumptions.

- **What to do if the VAR order p is unknown?**

  **Choosing p unnecessarily large:**

  1) will reduce the forecast precision of the corresponding estimated VAR(p) model.

  2) also the estimation precision of the impulse response.

  3) Approximate MSE matrix will increase with the order p.

# A Sequence of Tests for Determining the VAR Order

- There is not just one correct VAR order for the mentioned process. if (1) is a correct summary of the characteristics of the process $y_t$, then the same is true for:

$$y_t = v + A_1 Y_{t-1} + \ \ldots + \ A_P y_{t-p} + A_{P+1} y_{t-p-1} + u_t$$

with $A_{P+1} = 0$. (As according to assumptions, the possibility of zero coefficient matrices is not excluded)

- It is practical to have a unique number that is called the order of the process. $Y_t$ is called a VAR(p) process if $A_P \neq 0$ and $A_i = 0$ for $i > p$ so that p is the smallest possible order. This unique number will be called the VAR order.

# The Impact of the Fitted VAR Order on the Forecast MSE

- If $y_t$ is a VAR(p) process, it is useful to fit a $VAR(p)$ model and not a $VAR(p+i)$ because, forecasts from the latter process will be inferior to those based on an estimated VAR(p) model. This result follows from the approximate forecast MSE matrix. Example: $h = 1$, approximate forecast MSE is:

$$\sum_{\hat{y}}(1) = \frac{T+Kp+1}{T}\Sigma_u$$

  $\sum_{\mathrm{y}}(1)$ is an increasing function of the order of the model fitted to the data.

- The approximate MSE matrix is derived from asymptotic theory.

Checking whether the result remains true in small samples:

- 1000 Gaussian bivariate time series generated with process

$$y_t = \begin{bmatrix} .02 \\ .03 \end{bmatrix} + \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} y_{t-2} + u_t,$$

$$\Sigma_u = \begin{bmatrix} .09 & 0 \\ 0 & .04 \end{bmatrix}$$

- $VAR(2)$, $VAR(4)$, and $VAR(6)$ fitted to the generated series.
- forecasts with the estimated models have been computed.
- These forecasts compared to generated post-sample values.
- Average squared forecasting errors result for different forecast horizons h and sample size T (Table 1).

Table 1. Average squared forecast errors for the estimated bivariate $VAR(2)$ process based on 1000 realizations

| Sample size | Forecast horizon | Average squared forecast errors | | | | | |
|---|---|---|---|---|---|---|---|
| | | VAR(2) | | VAR(4) | | VAR(6) | |
| T | h | $y_1$ | $y_2$ | $y_1$ | $y_2$ | $y_1$ | $y_2$ |
| | 1 | .111 | .052 | .132 | .062 | .165 | .075 |
| 30 | 2 | .155 | .084 | .182 | .098 | .223 | .119 |
| | 3 | .146 | .141 | .183 | .166 | .225 | .202 |
| | 1 | .108 | .043 | .119 | .048 | .129 | .054 |
| 50 | 2 | .132 | .075 | .144 | .083 | .161 | .093 |
| | 3 | .142 | .120 | .150 | .130 | .168 | .145 |
| | 1 | .091 | .044 | .095 | .046 | .098 | .049 |
| 100 | 2 | .120 | .064 | .125 | .067 | .130 | .069 |
| | 3 | .130 | .108 | .135 | .113 | .140 | .113 |

- Forecasts based on estimated $VAR(2)$ models are clearly superior to the $VAR(4)$ and $VAR(6)$ forecasts for sample sizes T = 30, 50, and 100.

- Another Example: 1000 three-dimensional time series with the $VAR(1)$ process

$$y_t = \begin{bmatrix} .01 \\ .02 \\ 0 \end{bmatrix} + \begin{bmatrix} .5 & 0 & 0 \\ .1 & .1 & .3 \\ 0 & .2 & .3 \end{bmatrix} y_{t-1} + u_t \quad \text{with} \quad \Sigma_u = \begin{bmatrix} 2.25 & 0 & 0 \\ 0 & 1.0 & .5 \\ 0 & .5 & .74 \end{bmatrix}$$

- $VAR(1)$, $VAR(3)$, and $VAR(6)$ models fitted to these data.
- Forecasts and forecast errors computed.
- Average squared forecast errors are (Table 2).

**Table 2. Average squared forecast errors for the estimated three-dimensional VAR(1) process based on 1000 realizations**

| Sample size T | Forecast horizon h | Average squared forecast errors | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | VAR(2) | | | VAR(4) | | | VAR(6) | | |
| | | $y_1$ | $y_2$ | $y_3$ | $y_1$ | $y_2$ | $y_3$ | $y_1$ | $y_2$ | $y_3$ |
| 30 | 1 | .87 | 1.14 | 2.68 | 1.14 | 1.52 | 3.62 | 2.25 | 2.78 | 6.82 |
| | 2 | 1.09 | 1.21 | 3.21 | 1.44 | 1.67 | 4.12 | 2.54 | 2.98 | 7.85 |
| | 3 | 1.06 | 1.31 | 3.32 | 1.35 | 1.58 | 4.23 | 2.59 | 2.79 | 8.63 |
| 50 | 1 | .81 | 1.03 | 2.68 | .96 | 1.22 | 2.97 | 1.18 | 1.53 | 3.88 |
| | 2 | 1.01 | 1.23 | 2.92 | 1.20 | 1.40 | 3.47 | 1.48 | 1.68 | 4.38 |
| | 3 | 1.01 | 1.29 | 3.11 | 1.12 | 1.44 | 3.48 | 1.42 | 1.77 | 4.66 |
| 100 | 1 | .73 | .93 | 2.35 | .77 | 1.00 | 2.62 | .86 | 1.12 | 2.91 |
| | 2 | .94 | 1.15 | 2.86 | 1.00 | 1.24 | 3.12 | 1.12 | 1.38 | 3.53 |
| | 3 | .90 | 1.15 | 3.02 | .93 | 1.20 | 3.23 | 1.03 | 1.35 | 3.51 |

Again forecasts from lower order models are clearly superior to higher order models.

# Question?

- What if the true order is unknown and an upperbound, say M, for the order is known only.

Set up a significance test. First,

$$H_0\colon A_M = 0$$

is tested. If this null hypothesis cannot be rejected, we test

$$H_0\colon A_{M-1} = 0$$

and so on until we can reject a null hypothesis.

# The Likelihood Ratio Test Statistic

Because we just need to test zero restrictions on the coefficients, we may use the Wald statistic. For this, it may be instructive to consider the likelihood ratio testing principle which based on comparing the maxima of the log-likelihood function over the unrestricted and restricted parameter space. The likelihood ratio statistic is

$$\lambda_{LR} = 2[\ln l(\tilde{\delta}) - \ln l(\tilde{\delta}_r)]$$

$\tilde{\delta}$: unrestricted ML estimator obtained by maximizing likelihood function over the feasible parameter space.

$\tilde{\delta}_r$: restricted ML estimator obtained by max the likelihood function over that part of the parameter space where the restrictions of interest are satisfied.

- In case of linear constraint for the coefficient of a VAR, $\lambda_{LR}$ can be shown to have an asymptotic $\chi^2$-distribution with as many degrees of freedom as there are distinct linear restrictions.

- To obtain this result, assume that $y_t$ is a stable Gaussian VAR(p) process. The log-likelihood function is

$$\ln l(\beta, \Sigma_u) = -\frac{KT}{2}\ln 2\pi - \frac{T}{2}\ln|\Sigma_u|$$
$$-\frac{1}{2}[y - (Z' \otimes I_K)\beta]'(I_T \otimes \Sigma_u^{-1})[y - (Z' \otimes I_K)\beta]$$
$$\frac{\partial \ln l}{\partial \beta} = (Z \otimes \Sigma_u^{-1})y - (ZZ' \otimes \Sigma_u^{-1})\beta$$

the unrestricted ML/LS estimator: $\tilde{\beta} = ((ZZ')^{-1}Z \otimes I_K)y$

Restriction for $\beta$: $C\beta = c$

Where $C$ is a known $\left(N \times (K^2 p + K)\right)$matrix and $c$ is a known $(N \times 1)$ vector.

The restricted ML estimator may be found by a Lagrangian approach

$$\mathcal{L}(\beta, \gamma) = \ln l(\beta) + \gamma'(C\beta - c)$$

Equating first order partial derivatives to zero and solving:

$$\tilde{\beta}_r = \tilde{\beta} + [(ZZ')^{-1} \otimes \Sigma_u]C'[C\left((ZZ')^{-1} \otimes \Sigma_u\right)C']^{-1}(c - C\tilde{\beta})$$

for any given coefficient matrix $B^0$ the maximum of $\ln l$ with respect to $\Sigma_u$ is obtained for

$$\Sigma_u^0 = \frac{1}{T}(Y - B^0 Z)(Y - B^0 Z)'$$

the maximum for the unrestricted case is attained for

$$\tilde{\Sigma}_u^r = \frac{1}{T}(Y - \tilde{B}Z)(Y - \tilde{B}Z)'$$

and for the restricted case we get $\tilde{\Sigma}_u = \frac{1}{T}(Y - \tilde{B}_r Z)(Y - \tilde{B}_r Z)'$

for this particular situation, the likelihood ratio statistic becomes

$$\lambda_{LR} = 2[\ln l\left(\tilde{\beta}, \tilde{\Sigma}_u\right) - \ln l(\tilde{\beta}_r, \tilde{\Sigma}_u^r)]$$

## Proposition 1. (Asymptotic Distribution of the LR Statistic)

- Let $y_t$ be a stationary, stable VAR(p) process with standard white noise $u_t$. Suppose the true parameter vector $\beta$ satisfies linear constraints $C\beta = c$, where C is an $\left(\mathrm{N} \times (\mathrm{K}^2\mathrm{p} + \mathrm{K})\right)$ matrix of rank N and c is an $(\mathrm{N} \times 1)$vector. Moreover, let $\ln l$ denote the Gaussian log likelihood function and let $\tilde{\beta}$ and $\tilde{\beta}_r$be the (quasi)ML and restricted (quasi) ML estimators, respectively, with corresponding estimators $\tilde{\Sigma}_u$and $\tilde{\Sigma}_{\mathrm{u}}^{r}$ of the white noise covariance matrix $\Sigma_u$.

$$\lambda_{\mathrm{LR}} = 2[\ln \mathrm{l}\left(\tilde{\beta}, \tilde{\Sigma}_{\mathrm{u}}\right) - \ln \mathrm{l}(\tilde{\beta}_{\mathrm{r}}, \tilde{\Sigma}_{\mathrm{u}}^{r})]$$

$$= T(ln\left|\tilde{\Sigma}_{\mathrm{u}}^{r}\right| - ln\left|\tilde{\Sigma}_{\mathrm{u}}\right|)$$

$$= (\tilde{\beta}_r - \tilde{\beta})'(\mathrm{ZZ}' \otimes \tilde{\Sigma}_{\mathrm{u}}^{-1})(\tilde{\beta}_r - \tilde{\beta})$$

$$= (\tilde{\beta}_r - \tilde{\beta})'(\mathrm{ZZ}' \otimes (\tilde{\Sigma}_{\mathrm{u}}^{r})^{-1})(\tilde{\beta}_r - \tilde{\beta}) + o_p(1)$$

$$= (C\tilde{\beta} - c)'\left[C\left((ZZ')^{-1} \otimes \tilde{\Sigma}_{\mathrm{u}}\right)C'\right]^{-1}(C\tilde{\beta} - c) + o_p(1)$$

$$= (C\tilde{\beta} - c)'\left[C\left((ZZ')^{-1} \otimes \tilde{\Sigma}_{\mathrm{u}}^{r}\right)C'\right]^{-1}(C\tilde{\beta} - c) + o_p(1)$$

and $\lambda_{LR} \rightarrow \chi^2(N)$. T: Sample size and $Z \coloneqq (Z_0, \ldots, Z_{T-1})$ with $Z'_t = \left(1, y'_t, \ldots y'_{t-p+1}\right)$.

# A Testing Scheme for VAR Order Determination

Assuming M as upper bound for the VAR order, the following sequence of null and alternative hypotheses may be tested using LR tests:

$H_0^1: A_M = 0$ versus $H_1^1: A_M \neq 0$

$H_0^2: A_{M-1} = 0$ versus $H_1^2: A_{M-1} \neq 0 \mid A_M = 0$

$H_0^i: A_{M-i+1} = 0$ versus $H_1^i: A_{M-i+1} \neq 0 \mid A_M = \cdots = A_{M-i+2} = 0$

$H_0^M: A_1 = 0$ versus $H_1^M: A_1 \neq 0 \mid A_M = \cdots = A_2 = 0$

The procedure terminates and the VAR order is chosen accordingly, if one of the null hypotheses is rejected. That is, if $H_0^i$ is rejected, $\hat{p} = \mathrm{M} - \mathrm{i} + 1$ will be chosen as estimate of the autoregressive order.

- The likelihood ratio statistic for testing the i-th null hypothesis is

$$\lambda_{\mathrm{LR}}(\mathrm{i}) = \mathrm{T}[\ln\left|\tilde{\Sigma}_{\mathrm{u}}(\mathrm{M}-\mathrm{i})\right| - \ln\left|\tilde{\Sigma}_{\mathrm{u}}(\mathrm{M}-\mathrm{i}+1)\right|]$$

  $\tilde{\Sigma}_{\mathrm{u}}$: ML estimator of $\Sigma_{\mathrm{u}}$

- By Proposition 1, this statistic has an asymptotic $\chi^2(\mathrm{K}^2)$ distribution if $\mathrm{H}_0^{\mathrm{i}}$ and all previous null hypotheses are true. As $\mathrm{K}^2$ parameters are set to 0, so we have to test $\mathrm{K}^2$restriction and use $\lambda_{\mathrm{LR}}(\mathrm{i})$ in conjunction with critical values from a $\chi^2(\mathrm{K}^2)$ distribution. Alternatively use $\lambda_{LR}(i)/_{K^2}$in conjunction with the $\mathrm{F}(\mathrm{K}^2, \mathrm{T}-\mathrm{K}(\mathrm{M}-\mathrm{i}+1)-1)$-distribution.

- The order chosen for a process depend on the significance levels used in the tests. It is important to realize that the significance levels of the individual tests must be distinguished from the Type I error of the whole procedure because rejection of $\mathrm{H}_0^{\mathrm{i}}$ means that $\mathrm{H}_0^{\mathrm{i}+1}, \ldots, \mathrm{H}_0^{\mathrm{M}}$ are automatically rejected too.

- $D_j$ : the event that $H_0^j$ is rejected in the j-th test when it is actually true. The probability of a Type I error for the i-th test in the sequence is

$$\epsilon_i = \Pr(D_1 \cup D_2 \cup \ldots \cup D_i)$$

- $\gamma_j = \Pr(D_j)$ is the significance level of the j-th individual test. Hence, for independent events in large samples:

$$\epsilon_i = \Pr(D_1 \cup \ldots \cup D_{i-1}) + \Pr(D_i) - \Pr\{(D_1 \cup \ldots \cup D_{i-1}) \cap D_i\} =$$
$$\epsilon_{i-1} + \gamma_i - \epsilon_{i-1}\gamma_i = \epsilon_{i-1} + \gamma_i(1 - \epsilon_{i-1}),\ i=2,3,\ldots,M$$

- $\epsilon_1 = \gamma_1$. Thus by induction:

$$\epsilon_i = 1 - (1 - \gamma_1) \ldots (1 - \gamma_i) \qquad i = 1,2, \ldots, M$$

In the literature another testing scheme was also suggested and used. In that, the first set of hypotheses $(i = 1)$ is as the first test and for $i > 1$ the following hypotheses are tested:

$$H_0^i: A_M = \cdots = A_{M-i+1} = 0$$

versus

$$H_1^i: A_M \neq 0 \; or \; \ldots . or \; A_{M-i+1} \neq 0$$

Here $H_0^i$ is not tested conditionally on the previous null hypotheses being true but it is tested against the full VAR(M) model. Unfortunately, the LR statistics to be used in such a sequence will not be independent so that the overall significance level (probability of Type I error) is difficult to determine.

# Example

Table 3. ML estimates of the error covariance matrix

| VAR order m | $\widetilde{\Sigma}_u(m) \times 10^4$ | $\left\|\widetilde{\Sigma}_u(m)\right\| \times 10^{11}$ |
|---|---|---|
| 0 | $\begin{bmatrix} 21.83 & .410 & 1.228 \\ . & 1.420 & .571 \\ . & . & 1.084 \end{bmatrix}$ | 2.473 |
| 1 | $\begin{bmatrix} 20.14 & .493 & 1.173 \\ . & 1.318 & .625 \\ . & . & 1.018 \end{bmatrix}$ | 1.782 |
| 2 | $\begin{bmatrix} 19.18 & .617 & 1.126 \\ . & 1.270 & .574 \\ . & . & .821 \end{bmatrix}$ | 1.255 |
| 3 | $\begin{bmatrix} 19.08 & .599 & 1.126 \\ . & 1.235 & .543 \\ . & . & .784 \end{bmatrix}$ | 1.174 |
| 4 | $\begin{bmatrix} 16.96 & .573 & 1.252 \\ . & 1.234 & .544 \\ . & . & .765 \end{bmatrix}$ | .958 |

- Variables $y_1, y_2, y_3$: first differences of the logarithms of the investment, income, and consumption data.
- sample size is T = 71 in each test and M=4.

- The corresponding $\chi^2$ and F-test values are summarized in Table 4. Because the denominator degrees of freedom for the F-statistics are quite large, the F-tests are qualitatively similar to the $\chi^2$ tests.
- Using individual significance levels of .05 in each test, $A_2 = 0$ is the first null hypothesis that is rejected. Thus, the estimated order from both tests is 2.

**Table 4. LR Statistic for the investment/income/consumption system**

| i | $H_0^i$ | VAR order under $H_0^i$ | $\lambda_{LR^a}$ | $\lambda_{LR^a}/9^b$ |
|---|---|---|---|---|
| 1 | $A_4 = 0$ | 3 | 14.44 | 1.60 |
| 2 | $A_3 = 0$ | 2 | 4.76 | .53 |
| 3 | $A_2 = 0$ | 1 | 24.90 | 2.77 |
| 4 | $A_1 = 0$ | 0 | 23.25 | 2.58 |

a: Critical value for individual 5% level test: $\chi 2(9).95=16.92$.
b: Critical value for individual 5% level test: $F(9,71\text{-}3(5\text{-}i)\text{-}1).95 \approx 2$

# Criteria for VAR order selection

Review of assumptions:

K-dimensional multiple time series $y_1, \ldots, y_T$, with $y_t = (y_{1t}, \ldots, y_{Kt})$, which is known to be generated by a VAR(p) process:

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$$

The aim is estimation of the parameters $\nu, A_1, \ldots, A_p$, and $\Sigma_u = E(u_t u'_t)$

In deriving the properties of the estimators, a number of assumptions were made. Therefore statistical tools should be used in order to check the validity of the assumptions made.

In this section, it will be discussed what to do if the VAR order p is unknown. As

In practice, the order will usually be unknown.

Till now we have not assumed that all the $A_i$ are nonzero. In particular $A_p$ may be zero. In other words, p is just assumed to be an upper bound for the VAR order. And we know that the approximate MSE matrix of the 1-step ahead predictor will increase with the order p. Thus, choosing p unnecessarily large will reduce the forecast precision of the corresponding estimated VAR(p) model.

Therefore it is useful to have procedures or criteria for choosing an adequate VAR order.

# 4.3. INFORMATION CRITERIA FOR VAR ORDER SELECTION

Information criteria (IC) are statistics that measure the distance between observations and model classes. If the IC value is small, the distance is small and the model class contains a good descriptor of the DGP.

These criteria consist of two additive parts:

- The first one is a goodness-of-fit measure, such as minus the maximized likelihood within a given model class or a residual variance matrix. It becomes smaller as the model becomes more sophisticated.
- The second part is a penalty that increases with the model's complexity.

In AIC criteria, the most popular information criteria, the second part is just a monotonic function of the number of estimated parameters and the first part is the ML estimate of the error variance matrix.

$$AIC(m) = \ln|\widetilde{\sum}_u(m)| + \frac{2m}{T}$$

m : the number of free parameters ,T : sample size

$\sum_u$ denotes the ML estimate of the error variance matrix based on using the given model class with m free parameters.

-VAR(p) models have $m = K^2p + K + K(K + 1)/2$ free parameters.

-As here finding p value is central, the dimension K is kept constant, without imposing any restrictions on the error variance matrix.
Acting as if the VAR(p) model have $m = p\ K^2$ parameters. The AIC for a VAR(p) process is defined as:

$$AIC(m) = \ln|\tilde{\sum}_u(m)| + \frac{2m}{T}\ ,\ m = p\ K^2 \Rightarrow AIC(p) = \ln|\tilde{\sum}_u(m)| + \frac{2p}{T}k^2$$

The main strength of the AIC is due to the forecasting properties of the selected model, in this definition of AIC the MSE of a VAR-based one step prediction is approximated.

AIC also implies a relatively good model choice in small samples under different aspects, but it is not consistent.

Paulsen has claimed that only criteria of the form

$$Cr(m)= \ln|\widetilde{\sum}_u (m)|+\frac{m}{T}c_T$$

can achieve consistency.

Cr(m) is consistence criteria, because when $T \to \infty$, the selected p will be the true p with probability one.

$c_T$ is a function of T and can not be constant as
$T \to \infty$ then . $\frac{C_T}{T} \to 0$ and $c_T \to \infty$

If $c_T$ be constant can not complete the former condition.

Hannan-Quinn (HQ) and Schwartz criterion fulfill the conditions ascribed to Paulsen

$$HQ(m)= \ln|\widetilde{\sum}_u (m)|+ \frac{2\,m}{T}\, ln(ln\ T)$$

Schwarz version

$$SC(m)= \ln|\widetilde{\sum}_u(m)\ |+ \frac{m}{T}\, lnT$$

- HQ imposes a milder penalty in relation to SC, then SC selects comparatively smaller lag orders.

- HQ is consistent for univariate process and strogly consistent for K>1.And SC(m) is strongly consistent for any K.

If applied to vector auto regressions of order p, SC has the form

$$SC(p) = \ln|\widetilde{\sum}_u(p)| + \frac{p}{T} k^2 \ln T$$

Notes:

- These criteria vary across authors and software.
- The estimate of $\widetilde{\sum}_u$ is supposed to be a ML estimate but is occasionally replaced by a bias-corrected estimate with a correction for degrees of freedom. With these modifications the name of AIC changes to AICC and in some cases is preferred due to better small-sample properties.

To illustrate these procedures for VAR order selection, using investment/income/consumption example.

| VAR order | AIC(m) | HQ(m) | SC(m) |
|---|---|---|---|
| 0 | -24.42 | -24.42 | -24.42 |
| 1 | -24.5 | -24.38 | -24.21 |
| 2 | -24.59 | -24.37 | -24.02 |
| 3 | -24.41 | -24.07 | -23.55 |
| 4 | -24.36 | -23.9 | -23.21 |

The estimated $\tilde{p}$ (HQ or SC or AIC) is the order that minimize HQ or SC or AIC for m=0,1,…,M

# 4.4. RESIDUAL ANALYSIS: AUTOCORRELATION

Chitturi 1974

Hosking 1980

Li&McLeod 1981

Ahn 1988

It is helpful to note that estimated residuals from VAR models are not white noise, while errors from the true VAR are white noise.

All tests for misspecification rely on these none WN estimated residuals, and then what is checked is whether the residuals are ‘too non-white Noise’.

$u_t$ : K-dimensional white noise process with nonsingular covariance matrix $\Sigma_u$.

$u_t$ represents the residuals of a VAR(p) process.

U := ($u_1$, . . . , $u_T$ )

The autocovariance matrices of $u_t$ are estimated as

$$C_i=\frac{1}{T}\sum_{t=i+1}^{T} u_t \; u'_{t-i}$$

Let $C_h$:= Vec($C_1$, $C_2$ ,…, $C_h$) and also $r_h$:=vec($\mathbf{R}_h$)

$r_h$ is also the vectorization of the first h ,ACF matrices .

Formally it can be proved that the asymptotic distribution of $r_h$ (white noise Auto covariance) and $C_h$(white noise Autocorrelation)

$$\sqrt{T}\ C_h \xrightarrow{Converges} N(0, I_h \otimes \sum_u \otimes \sum_u )$$

And

$$\sqrt{T}\ r_h \xrightarrow{Converges} N(0, I_h \otimes R_u \otimes R_u )$$

The Matrix $R_i$ that is assumed to be an estimator of the true correlation matrix corresponding to $\sum_u$ , and

$$R_j = D^{-1}\ C_j\ D^{-1}$$

The diagonal matrices D contain the estimate of standard deviation (square roots of the variance ) of the K component.

Of more concern here is that these equations are literally only valid for true white noise, and residuals are not white noise.

Therefore, it is important that similar results hold for the estimated autocovariances of residuals from a fitted valid VAR model. It can be shown that

$$\sqrt{T}\,\hat{C}_h \xrightarrow{Converges} N(0, I_h \otimes \sum_u \otimes \sum_u - (\bar{G}\,' \otimes I_k)(\Gamma_Y(0)^{-1} \otimes \sum_u)\,(\bar{G} \otimes I_k)$$

$\bar{G}$ Matrix , which is defined as

$$\bar{G} = \begin{bmatrix} \Sigma_u & \Phi_1\Sigma_u & \dots & \Phi_{h-1}\Sigma_u \\ 0 & \Sigma_u & & \Phi_{h-2}\Sigma_u \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \Phi_{h-p}\Sigma_u \end{bmatrix}$$

-This $\bar{G}$ Matrix has dimension Kp × Kh, not quadratic. Here it is assumed h > p.

-The elements $\Phi_j$ come from the infinite MA representation.

- Correction terms are non-negative definite, therefore estimated auto covariance and autocorrelation functions tend to be smaller than those defined from the unobserved errors

-A similar result with similar structure holds for the estimates of the errors autocorrelation function based on residuals $\hat{r}_h$.

Usually, the hypothesis of whiteness of errors is tested using portmanteau or Q tests.

Q statistics is defined as

$$Q_h = T \sum_{j=1}^{h} tr\ (\hat{R}'_j \hat{R}_u^{-1} \hat{R}_j \hat{R}_u^{-1})$$

- $\hat{R}_j$ are estimates of the autocorrelation function of the errors at lag j
- $\hat{R}_u$ is an estimate of the errors correlation matrix $R_u = R_0$.

- Under the $H_0 : R_1 = R_2 = \ldots = R_h = 0$

$$Q_h \sim \chi^2(K^2(h-p))$$

- Important: approximation will be invalid for small h.

# Brief recap

Statistical tools to check validity of assumptions regarding consistency and asymptotic normality of the estimators considered, *eg*...

...so far:

- Procedures or criteria for choosing adequate VAR order.
- Tools for checking whiteness of residuals.

...next:

- Tests for nonnormality (normality assumption used in setting up forecast intervals).
- Tests for structural change (system stationarity is important assumption). *

---

* Nonstationarities may have various forms; trends indicate deviations from stationarity, but also changes in the variability or variance of the system; exogenous shocks may affect various characteristics of the system.

# Residual analysis: testing for nonnormality (1)

Univariate models (vector white noise process):

- Most common check for nonnormality in univariate models is Jarque-Bera test.
- Checks if skewness (third moment) is zero and kurtosis (normalised fourth moment) is 3 → null hypothesis contains non-Gaussian distributions.
- Under null, statistic is asymptotically distributed as $\chi^2$ with 2 degrees of freedom.

Multivariate models (VAR process):

- Corresponding multivariate version is sum of *K* univariate statistics calculated from *K* orthogonalised VAR residuals (orthogonalisation via Cholesky split). *
- Statistic distributed as $\chi^2$ with 2*K* degrees of freedom under null.

* Only sum of uncorrelated $\chi^2$ random variables will yield a $\chi^2$ distribution.

# Residual analysis: testing for nonnormality (2)

*Univariate models (vector white noise process):*

**Proposition** (asymptotic distribution of skewness and kurtosis):
If $u_t$ is Gaussian white noise with nonsingular covariance matrix $\Sigma_u$ and expectation $\mu_u$, $u_t \sim N(\mu_u, \Sigma_u)$, then

$$\sqrt{T}\begin{bmatrix} b_1 \\ b_2 - 3_K \end{bmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} 6I_K & 0 \\ 0 & 24I_K \end{bmatrix}\right).$$

So $b_1$ and $b_2$ (estimators of vectors in next slide) are asymptotically independent and normally distributed.

# Residual analysis: testing for nonnormality (3)

The proposition implies that

$\lambda_s := T b_1' b_1/6 \xrightarrow{d} \chi^2(K)$, and $\lambda_k := T(b_2 - \mathbf{3}_K)'(b_2 - \mathbf{3}_K)/24 \xrightarrow{d} \chi^2(K)$.

$\lambda_s$ can be used to test $H_0: E\begin{bmatrix} w_{1t}^3 \\ \vdots \\ w_{Kt}^3 \end{bmatrix} = 0$ against $H_1: E\begin{bmatrix} w_{1t}^3 \\ \vdots \\ w_{Kt}^3 \end{bmatrix} \neq 0$

$\lambda_k$ can be used to test $H_0: E\begin{bmatrix} w_{1t}^4 \\ \vdots \\ w_{Kt}^4 \end{bmatrix} = \mathbf{3}_K$ against $H_1: E\begin{bmatrix} w_{1t}^4 \\ \vdots \\ w_{Kt}^4 \end{bmatrix} \neq \mathbf{3}_K$. *

Also, $\lambda_{sk} := \lambda_s + \lambda_k \xrightarrow{d} \chi^2(2K)$ can be used for joint test of null hypotheses above.

---

* Components of wt are independent standard normal random variables.

# Residual analysis: testing for nonnormality (4)

*Multivariate models (VAR process):*

Transform residual vector to make individual components independent; check if all have third and fourth moments corresponding to those of normal distributions.

More precisely, for given residuals $\hat{u}_t$ ($t = 1,\ldots,T$) of estimated process, determine residual covariance matrix $\widetilde{\Sigma}_u$ and compute corresponding square root matrix $\widetilde{\Sigma}_u^{1/2}$.

Test for nonnormality based on skewness and kurtosis of standardised residuals $\hat{u}_t^s = (\hat{u}_{1t}^s, \ldots, \hat{u}_{Kt}^s)' = \widetilde{\Sigma}_u^{-1/2}\hat{u}_t$: *

$\mathbf{b}_1 = (b_{11}, \ldots, b_{1K})'$ with $b_{1k} = T^{-1} \sum_{t=1}^{T} (\hat{u}_{kt}^s)^3$, and
$\mathbf{b}_2 = (b_{21}, \ldots, b_{2K})'$ with $b_{2k} = T^{-1} \sum_{t=1}^{T} (\hat{u}_{kt}^s)^4$

---

* Choleski decomposition of residual covariance matrix can be used to standardise residuals.

# Residual analysis: testing for nonnormality (5)

Standardised residuals should be mean-adjusted unless $\hat{u}_t$'s have zero sample mean.

Tests can be based on third moments only, using test statistic
$s_3^2 = T\mathbf{b}_1'\mathbf{b}_1/6$, or on fourth moments only, using
$s_4^2 = T(\mathbf{b}_2 - 3_K)'(\mathbf{b}_2 - \mathbf{3}_K)/24$. Here $\mathbf{3}_K = (3, \ldots, 3)'$ is a (*K x 1*) vector.

Both statistics have asymptotic $\chi^2$ (*K*)-distributions if underlying white noise process $u_t$ is normally distributed.

Moreover, they can be combined as $JB_K = s_3^2 + s_4^2$ to check third and fourth moments jointly. Statistic $JB_K$ has $\chi^2$ (2*K*) limiting distribution under normality of process.

# Residual analysis: tests for structural change (1)

Objective:

- Check stability or time invariance/detect structural breaks during sample period.
- Model defects may indicate poor representation of data generating process (DGP).
- Then find better representation (*eg*, add other variables or further lags, include nonlinear terms/change functional form, modify sampling period, get other data).

Two tests for alternative that some characteristics of VAR have changed over time:

- Multivariate version of univariate Chow test assumes given possible time point at which process may have changed.
- Stability test checks whether linear VAR forecast behaves as it should under assumption of normal errors and well-specified time-constant VAR.

# Residual analysis: tests for structural change (2)

*Chow test:*

Tests for significance of all VAR regressors ($pK^2 + K$) including intercept multiplied with a dummy that is 0 before a break point and one afterwards.

Chow statistic has an asymptotic $\chi^2$ distribution with $pK^2 + K$ degrees of freedom under null.

Potential practical problems in applying these tests:

- If possible break date is very close to sample beginning or end, LS/ML estimators of $B_i$ may not be available due to lack of degrees of freedom.
- Problems may also result from multiple structural breaks within sample period.
- Asymptotic theory may be poor guide for small sample properties of Chow tests, in particular, if models with many parameters are under consideration.

# Residual analysis: tests for structural change (3)

Suppose change in parameters of VAR($p$) process is suspected after period $T_1 < T$. Given sample $y_1,\ldots,y_T$ plus required presample values, model can be set up as follows for estimation purposes:

$$\left[Y_{(1)}: Y_{(2)}\right] = [B_1: B_2]\mathbf{Z} + \left[U_{(1)}: U_{(2)}\right] = \mathbf{BZ} + U,$$

where $Y_{(1)} := \left[y_1, \ldots, y_{T_1}\right], Y_{(2)} := \left[y_{T_1+1}, \ldots, y_T\right]$, $U$ is partitioned accordingly, $B_1 := \left[v_1, A_{11}, \ldots A_{p1}\right]$ and $B_2 := \left[v_2, A_{12}, \ldots A_{p2}\right]$ are ($K$ x ($pK$ +1)) dimensional parameter matrices associated with first ($t$ = 1,...,$T_1$) and last ($t$ = $T_1$ + 1,...,$T$) subperiods, respectively, $\mathbf{B} := [B_1: B_2]$ is ($K$ x 2($Kp$ + 1)) dimensional and

$$\mathbf{Z} := \begin{bmatrix} Z_{(1)} & 0 \\ 0 & Z_{(2)} \end{bmatrix}.$$

Here, $Z_{(1)} := \left[Z_0, \ldots, Z_{T_1-1}\right]$ and $Z_{(2)} := \left[Z_{T_1}, \ldots, Z_{T-1}\right]$ with $Z_t' := \left(1, y_t', \ldots, y_{t-p+1}'\right)$.

# Residual analysis: tests for structural change (4)

$$H_0 : B_1 = B_2 \text{ or } [I : -I]\text{vec}(\mathbf{B}) = 0 \text{ versus } \quad H_1 : B_1 \neq B_2.$$

The LS estimator of **B** is

$$\widehat{\mathbf{B}} = \left[Y_{(1)} : Y_{(2)}\right]\mathbf{Z}'(ZZ')^{-1} = \left[Y_{(1)}Z'_{(1)}\left(Z_{(1)}Z'_{(1)}\right)^{-1} : Y_{(2)}Z'_{(2)}\left(Z_{(2)}Z'_{(2)}\right)^{-1}\right]$$

Estimator has asymptotic normal distribution under certain assumptions (see proposition on next slide); it must be assured that $T^{-1}\mathbf{ZZ}'$ converges in probability to nonsingaluar matrix, *ie* $\text{plim}\frac{T_i}{T}\frac{1}{T_i}Z_{(i)}Z'_{(i)}$, $i = 1, 2$, must exist and be nonsingular.

$T_i/T$ must not go to zero when $T$ goes to $\infty$ → both subperiods before and after break assumed to increase with $T$.

If asymptotic normality assumptions justified, Wald test could be used to test stability hypothesis; or, LR or quasi LR test may be applied (also known as Chow test).

# Residual analysis: tests for structural change (5)

**Proposition** (asymptotic properties of LS estimator):
Let $y_t$ be a stable, *K*-dimensional VAR(*p*) process with standard white noise residuals, $\hat{B} = YZ'(ZZ')^{-1}$ is the LS estimator of the VAR coefficients *B* and all symbols correspondingly defined.

Then $\operatorname{plim} \hat{B} = B$, and

$$\sqrt{T}\left(\hat{\beta} - \beta\right) = \sqrt{T}\, vec\left(\hat{B} - B\right) \xrightarrow{d} N(0, \Gamma^{-1} \otimes \Sigma_u),$$

or, equivalently,

$$\sqrt{T}\left(\hat{\mathbf{b}} - \mathbf{b}\right) = \sqrt{T}\, vec\left(\hat{B}' - B'\right) \xrightarrow{d} N(0,\ \Sigma_u \otimes \Gamma^{-1}),$$

where $\Gamma = \operatorname{plim} ZZ'/T$.

# Residual analysis: tests for structural change (6)

*Forecast tests for structural change:*

Departs from observation that (theoretical, *ie* for known VAR coefficients) *h*-step forecast error from VAR model at time *T* has property $e_T(h) \sim N(0, \Sigma_y(h))$, where $\Sigma_y(h) = \Sigma_{j=0}^{h-1} \Phi_j \Sigma_u \Phi_j'$. *

$\Phi_j$ denotes matrices from infinite MA representation of VAR; entities on right side must be replaced by estimates; then may hope that $\bar{\tau}_h = e_T(h)' \widehat{\Sigma}_y(h)^{-1} e_T(h)$ ** is distributed with $\chi^2$ with *K* degrees of freedom und null. ***

---

* Forecast MSE matrix.

** Statistic used to test whether $y_{T+h}$ is generated by same Gaussian VAR(*p*) process that has generated $y_1,\ldots,y_T$ (null hypothesis).

*** But this is not quite true, as the naive estimate for $\Sigma_y$ underestimates variance of prediction error; it is recommended to add adjustment term to $\widehat{\Sigma}_y$ that cares for fact that VAR parameters must be estimated and are not known to forecaster.

**Thank you for your attention!**