Lecture 6 - Forecasting with VAR Models

Pedro Valls[1]

[1]Sao Paulo School of Economics - FGV
CEQEF-FGV

July 22, 2024

FGV EESP

CENTRO DE ESTUDOS QUANTITATIVOS EM ECONOMIA E FINANÇAS

Table of Contents

Outline

- Introduction
- Representation
- Specification of the model
- Estimation
- Diagnostic checks
- Forecasting
- Impulse response functions
- Forecast error variance decomposition
- Structural VARs with long-run restrictions

Table of Contents

Introduction

- A natural extension of the univariate time series models is the vector (multivariate) $ARMA$ class ($VARMA$).
- The main advantage of these models is their generality and they do not impose any strong a priori restrictions on the cross-variable relationships.
- The cost is that they are heavily parameterized .
- The seminal work by [Sims, 1980] led to a widespread use of VAR models.
- Since it is difficult to model and estimate a multivariate MA component, usually only Vector Autoregressive (VAR) models will be considered.

Table of Contents

Representation I

- A $VAR(p)$ model for the set of m variables $y_{1t}, \cdots, y_{mt}$ grouped in the $m \times 1$ vector $\mathbf{y}_t = (y_{1t}, \cdots, y_{mt})'$ is given by

$$\underset{m\times 1}{\mathbf{y}_t} = \underset{m\times 1}{\mathbf{c}} + \underset{m\times m}{\Phi_1}\mathbf{y}_{t-1} + \cdots + \Phi_p \mathbf{y}_{t-p} + \underset{m\times 1}{\epsilon_t} \text{ and } \epsilon_t \sim WN\left(0, \underset{m\times m}{\Sigma}\right) \tag{1}$$

- In a $VAR(p)$ each variable depends on:
  - up to p of its own lags and up to p lags of each other variables
  - an intercept or other deterministic components - seasonal dummies or a time trend
  - error term of each equation has zero mean and is serially uncorrelated and homoskedastic but it can be contemporaneously correlated with the erros in other equations

- The number of parameters for n a $VAR(p)$ is

$$\underset{\text{intercepts}}{m} + \underset{\text{coeficients of lagged variable}}{m^2 p} + \underset{\text{var-cov for errors}}{m(m+1)/2}$$

Representation II

- for example if $m = 10$ and $p = 2$ we have 265 parameters
- As in the univariate case, weak stationarity is an important property. It requires that

$$\begin{aligned} E(\mathbf{y}_t) &= \mathbf{c}^* \\ Var(\mathbf{y}_t) &= \Gamma_0 < \infty \\ Cov(\mathbf{y}_t, \mathbf{y}_{t+k}) &= \Gamma_k \text{ depends on } k \text{ but not on } t \end{aligned} \tag{2}$$

  where $\mathbf{c}^* = (\mathbf{I} - \Phi_1 - \ldots \Phi_p)^{-1} \times \mathbf{c}$.
- For a $VAR(p)$ process, weak stationarity is verified when all the roots z of $\det(\mathbf{I} - \Phi_1 \mathbf{z} - \cdots - \Phi_p \mathbf{z}^p) = 0$ are larger than one in absolute value.

Representation III

- To theoretically justify the use of VAR models we should use the Wold representation theorem for the multivariate case that for a $\mathbf{y}_t$ $m-$dimensional weakly stationary process, that is, it admits the representation

$$\mathbf{y}_t = \mathbf{C}(L)\epsilon_t \tag{3}$$

- where $\mathbf{C}(L)$ is a $m \times m$ matrix polynomial in the lag operator L, that is

$$\mathbf{C}(L) = \mathbf{I} + \mathbf{C}_1 L + \mathbf{C}_2 L^2 + \cdots$$

- Under mild conditions $\mathbf{C}(L)$ can be approximated by $\mathbf{A}^{-1}(L)\mathbf{B}(L)$ and (3) can be rewrite as

$$\mathbf{A}(L)\mathbf{y}_t = \mathbf{B}(L)\epsilon_t \tag{4}$$

- with $\mathbf{A}(L) = \mathbf{I} + \mathbf{A}_1 L + \cdots + \mathbf{A}_p L^p$, and $\mathbf{B}(L) = \mathbf{I} + \mathbf{B}_1 L + \cdots + \mathbf{B}_q L^q$

Representation IV

- Under slightly more stringent condition we can assume $q = 0$ and (4) becomes a $VAR(p)$

Table of Contents

Specification of the model

- For the specification of the number of lags use multivariate versions of the information criteria.
- The most common multivariate IC are

$$
\begin{aligned}
AIC(j) &= \ln\left|\widehat{\Sigma}_j\right| + \frac{2}{T} j m^2 \\
BIC(j) &= \ln\left|\widehat{\Sigma}_j\right| + \frac{\ln(T)}{T} j m^2 \quad j = 1, \cdots, p_{\max}
\end{aligned}
$$

- Or, in terms of the log-likelihood:

$$
\begin{aligned}
AIC(j) &= \frac{(-2\widehat{\ell} + 2jm^2)}{T} \\
BIC(j) &= \frac{(-2\widehat{\ell} + \ln(T) j m^2)}{T} \quad j = 1, \cdots, p_{\max}
\end{aligned}
$$

- For a discussion of the use of these and related scalar measures to choose between alternative models in a class, see [Lütkepohl, 2007] .

Table of Contents

Estimation

- Since the regressors are the same in each equation of the VAR, OLS equation by equation is equivalent to OLS for the system and also equivalent to GLS for the system (see my lecture notes in Time Series Econometrics avaliable here VAR&VEC)
- Also under normality OLS and MLE are equivalent

Table of Contents

Diagnostic checks

- It is important to asses whether the underlying assumptions are suported by the data.
- In particular, if the errors are multivariate white noise, namely, uncorrelated and homoskedastic
- These properties can be tested using the multivariate versions of the statistics for no serial correlation and homoskedasticity and normality

Table of Contents

- The optimal forecast in the MSFE sense for a VAR(p) is obtained as a simple extension of the formula for the univariate case:

$$\widehat{\mathbf{y}}_{T+h} = \Phi_1 \widehat{\mathbf{y}}_{T+h-1} + \cdots + \Phi_p \widehat{\mathbf{y}}_{T+h-p} \tag{5}$$

- where $\widehat{\mathbf{y}}_{T+h-j} = \mathbf{y}_{T+h-j}$ for $h-j \leq 0$, can possibily add an intercept $\mathbf{c}$ or any other deterministic component
- Hence to compute the forecast for $\widehat{\mathbf{y}}_{T+h}$ we calculate $\widehat{\mathbf{y}}_{T+1}$, use it to obtain $\widehat{\mathbf{y}}_{T+2}$ and keep iterating until we obtain $\widehat{\mathbf{y}}_{T+h}$. This approach is usually defined as "iterated forecasting" .

Forecasting II

- For example if $\mathbf{y}_t$ follows a $VAR(1)$ then

$$\begin{aligned}
\widehat{\mathbf{y}}_{T+1} &= \Phi_1 \mathbf{y}_T \\
\widehat{\mathbf{y}}_{T+2} &= \Phi_1 \widehat{\mathbf{y}}_{T+1} = \Phi_1^2 \mathbf{y}_T \\
&\vdots \\
\widehat{\mathbf{y}}_{T+h} &= \Phi_1 \widehat{\mathbf{y}}_{T+h-1} = \Phi_1^h \mathbf{y}_T
\end{aligned}$$

- In the $VAR(p)$ context, the direct model takes the form

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-h} + \cdots + \Phi_p \mathbf{y}_{t-h-p} + \epsilon_t$$

- with forecast

$$\widetilde{\mathbf{y}}_{T+h} = \Phi_1 \mathbf{y}_T + \cdots + \Phi_p \mathbf{y}_{T-p} \tag{6}$$

Forecasting III

- Under correct specification $\widehat{\mathbf{y}}_{T+h}$ is more efficient than $\widetilde{\mathbf{y}}_{T+h}$. However in the presence of model mis-specifications, $\widetilde{\mathbf{y}}_{T+h}$ can be more robust than $\widehat{\mathbf{y}}_{T+h}$.
- In the context of forecasting the notion of Granger non-causality is important. Given a VAR written in the form

$$\begin{bmatrix} a_{11}(L) & a_{12}(L) \\ a_{21}(L) & a_{22}(L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \epsilon_{y_t} \\ \epsilon_{x_t} \end{bmatrix} \tag{7}$$

- x_t does not Granger-cause y_t if $a_{12}(L) = 0$.
- Similarly y_t does not Granger-cause x_t if $a_{21}(L) = 0$.
- We can also write the optimal forecast using the $MA(\infty)$ representation (3) in the following way

$$\widehat{\mathbf{y}}_{T+h} = \sum_{j=0}^{\infty} \mathbf{C}_{j+h} \epsilon_{T-j} \tag{8}$$

- with associated forecast error

$$\mathbf{e}_{T+h} = \sum_{j=0}^{h-1} \mathbf{C}_{j+h}\epsilon_{T+h-j} \tag{9}$$

- Hence the variance-covariance matrix of the forecast error is

$$V(\mathbf{e}_{T+h}) = \Sigma + \mathbf{C}_1\Sigma\mathbf{C}_1' + \cdots + \mathbf{C}_{h-1}\Sigma\mathbf{C}_{h-1}' \tag{10}$$

- The elements on the diagonal of $V(\mathbf{e}_{T+h})$ can be used to construct interval forecast for variables $\widehat{y}_{j,T+h} \quad j = 1, \cdots, m$.
- The formula for the optimal forecast in (8) can be rearranged into

$$\begin{aligned} \widehat{\mathbf{y}}_{T+h} &= \mathbf{C}_h\epsilon_T + \sum_{j=0}^{\infty} \mathbf{C}_{j+h+1}\ \epsilon_{T-1-j} \\ &= \widehat{\mathbf{y}}_{T+h|T-1} + \mathbf{C}_h\epsilon_T \end{aligned} \tag{11}$$

Forecasting V

- and using (9) gives

$$\widehat{\mathbf{y}}_{T+h} = \widehat{\mathbf{y}}_{T+h|T-1} + \mathbf{C}_h \left(\underbrace{\mathbf{y}_T - \widehat{\mathbf{y}}_{T|T-1}}_{\text{forecast error}} \right) \quad (12)$$

- The left hand side of equation (12) is the optimal forecast for period $T + h$ made in period T and the right hand side is optimal forecast for period $T + h$ made in period $T - 1$ and one-step ahead forecast error made in forecasting y_T in period $T - 1$
- It was assumed that the parameters were known
- When the parameters are estimated, the expressions for the forecast error variance must be modified to take into account parameter estimation uncertainty.
- See [Lütkepohl, 2007] for a detailed deviation of these expressions

Table of Contents

Forecasting - Lütkepohl I

3.5 Forecasting with Estimated Processes

3.5.1 General Assumptions and Results

In Chapter 2, Section 2.2, we have seen that the optimal h-step forecast of the process (3.1.1) is

$$y_t(h) = \nu + A_1 y_t(h-1) + \cdots + A_p y_t(h-p), \tag{3.5.1}$$

where $y_t(j) = y_{t+j}$ for $j \leq 0$. If the true coefficients $B = (\nu, A_1, \ldots, A_p)$ are replaced by estimators $\widehat{B} = (\widehat{\nu}, \widehat{A}_1, \ldots, \widehat{A}_p)$, we get a forecast

$$\widehat{y}_t(h) = \widehat{\nu} + \widehat{A}_1 \widehat{y}_t(h-1) + \cdots + \widehat{A}_p \widehat{y}_t(h-p), \tag{3.5.2}$$

where $\widehat{y}_t(j) = y_{t+j}$ for $j \leq 0$. Thus, the forecast error is

$$\begin{aligned} y_{t+h} - \widehat{y}_t(h) &= [y_{t+h} - y_t(h)] + [y_t(h) - \widehat{y}_t(h)] \\ &= \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} + [y_t(h) - \widehat{y}_t(h)], \end{aligned} \tag{3.5.3}$$

where the Φ_i are the coefficient matrices of the canonical MA representation of y_t (see (2.2.9)). Under quite general conditions for the process y_t, the forecast errors can be shown to have zero mean, $E[y_{t+h} - \widehat{y}_t(h)] = 0$, so that the forecasts are unbiased even if the coefficients are estimated. Because we do not need this result in the following, we refer to Dufour (1985) for the details and a proof. All the u_s in the first term on the right-hand side of the last equality sign in (3.5.3) are attached to periods $s > t$, whereas all the y_s in the second term correspond to periods $s \leq t$, if estimation is done with observations from periods up to time t only. Therefore, the two terms are uncorrelated. Hence, the MSE matrix of the forecast $\widehat{y}_t(h)$ is of the form

$$\begin{aligned} \Sigma_{\widehat{y}}(h) &:= \text{MSE}[\widehat{y}_t(h)] = E\{[y_{t+h} - \widehat{y}_t(h)][y_{t+h} - \widehat{y}_t(h)]'\} \\ &= \Sigma_y(h) + \text{MSE}[y_t(h) - \widehat{y}_t(h)], \end{aligned} \tag{3.5.4}$$

where

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

(see (2.2.11)). In order to evaluate the last term in (3.5.4), the distribution of the estimator $\widehat{B}$ is needed. Because we have not been able to derive the small sample distributions of the estimators considered in the previous sections but we have derived the asymptotic distributions instead, we cannot hope for more than an asymptotic approximation to the MSE of $y_t(h) - \widehat{y}_t(h)$. Such an approximation will be derived in the following.

There are two alternative assumptions that can be made in order to facilitate the derivation of the desired result:

Forecasting - Lütkepohl II

(1) Only data up to the forecast origin are used for estimation.
(2) Estimation is done using a realization (time series) of a process that is independent of the process used for prediction and has the same stochastic structure (for instance, it is Gaussian and has the same first and second moments as the process used for prediction).

The first assumption is the more realistic one from a practical point of view because estimation and forecasting are usually based on the same data set. In that case, because the sample size is assumed to go to infinity in deriving asymptotic results, either the forecast origin has to go to infinity too or it has to be assumed that more and more data at the beginning of the sample become available. Because the forecast uses only p vectors y_s prior to the forecast period, these variables will be asymptotically independent of the estimator $\widehat{B}$ (they are asymptotically negligible in comparison with all the other observations going into the estimate). Thus, asymptotically the first assumption implies the same results as the second one. In the following, for simplicity, the second assumption will therefore be used. Furthermore, it will be assumed that for $\boldsymbol{\beta} = \mathrm{vec}(B)$ and $\widehat{\boldsymbol{\beta}} = \mathrm{vec}(\widehat{B})$ we have

$$\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\boldsymbol{\beta}}}). \tag{3.5.5}$$

Samaranayake & Hasza (1988) and Basu & Sen Roy (1986) give a formal proof of the result that the MSE approximation obtained in the following remains valid under assumption (1) above.

With the foregoing assumptions it follows that, conditional on a particular realization $Y_t = (y_t', \dots, y_{t-p+1}')'$ of the process used for prediction,

$$\sqrt{T}\,[\widehat{y}_t(h) - y_t(h)|Y_t] \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial y_t(h)}{\partial \boldsymbol{\beta}'} \Sigma_{\widehat{\boldsymbol{\beta}}} \frac{\partial y_t(h)'}{\partial \boldsymbol{\beta}}\right) \tag{3.5.6}$$

because $y_t(h)$ is a differentiable function of $\boldsymbol{\beta}$ (see Appendix C, Proposition C.15(3)). Here T is the sample size (time series length) used for estimation. This result suggests the approximation of MSE $[\widehat{y}_t(h) - y_t(h)]$ by $\Omega(h)/T$, where

$$\Omega(h) = E\left[\frac{\partial y_t(h)}{\partial \boldsymbol{\beta}'} \Sigma_{\widehat{\boldsymbol{\beta}}} \frac{\partial y_t(h)'}{\partial \boldsymbol{\beta}}\right]. \tag{3.5.7}$$

In fact, for a Gaussian process y_t,

$$\sqrt{T}\,[\widehat{y}_t(h) - y_t(h)] \xrightarrow{d} \mathcal{N}(0, \Omega(h)). \tag{3.5.8}$$

Hence, we get an approximation

$$\Sigma_{\widehat{y}}(h) = \Sigma_y(h) + \frac{1}{T}\Omega(h) \tag{3.5.9}$$

for the MSE matrix of $\widehat{y}_t(h)$.

Forecasting - Lütkepohl III

From (3.5.7) it is obvious that $\Omega(h)$ and, thus, the approximate MSE $\Sigma_{\hat{y}}(h)$ can be reduced by using an estimator that is asymptotically more efficient than $\widehat{\boldsymbol{\beta}}$, if such an estimator exists. In other words, efficient estimation is of importance in order to reduce the forecast uncertainty.

3.5.2 The Approximate MSE Matrix

To derive an explicit expression for $\Omega(h)$, the derivatives $\partial y_t(h)/\partial \boldsymbol{\beta}'$ are needed. They can be obtained easily by noting that

$$y_t(h) = J_1 \mathbf{B}^h Z_t, \tag{3.5.10}$$

where $Z_t := (1, y_t', \dots, y_{t-p+1}')'$,

$$\mathbf{B} := \underset{[(Kp+1)\times(Kp+1)]}{\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ \nu & A_1 & A_2 & \dots & A_{p-1} & A_p \\ 0 & I_K & 0 & \dots & 0 & 0 \\ 0 & 0 & I_K & & 0 & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & I_K & 0 \end{bmatrix}} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ & B & & \\ 0 & I_{K(p-1)} & & 0 \end{bmatrix}$$

and

$$J_1 := [\underbrace{0}_{(K\times 1)} : I_K : \underbrace{0 : \dots : 0}_{(K\times K(p-1))}] \quad [K \times (Kp+1)].$$

The relation (3.5.10) follows by induction (see Problem 3.8). Using (3.5.10), we get

$$\begin{aligned}
\frac{\partial y_t(h)}{\partial \boldsymbol{\beta}'} &= \frac{\partial \operatorname{vec}(J_1 \mathbf{B}^h Z_t)}{\partial \boldsymbol{\beta}'} = (Z_t' \otimes J_1) \frac{\partial \operatorname{vec}(\mathbf{B}^h)}{\partial \boldsymbol{\beta}'} \\
&= (Z_t' \otimes J_1) \left[\sum_{i=0}^{h-1} (\mathbf{B}')^{h-1-i} \otimes \mathbf{B}^i \right] \frac{\partial \operatorname{vec}(\mathbf{B})}{\partial \boldsymbol{\beta}'} \\
&\qquad \text{(Appendix A.13, Rule (8))} \\
&= (Z_t' \otimes J_1) \left[\sum_{i=0}^{h-1} (\mathbf{B}')^{h-1-i} \otimes \mathbf{B}^i \right] (I_{Kp+1} \otimes J_1') \\
&\qquad \text{(see the definition of } \mathbf{B}) \\
&= \sum_{i=0}^{h-1} Z_t' (\mathbf{B}')^{h-1-i} \otimes J_1 \mathbf{B}^i J_1' \\
&= \sum_{i=0}^{h-1} Z_t' (\mathbf{B}')^{h-1-i} \otimes \Phi_i,
\end{aligned} \tag{3.5.11}$$

Forecasting - Lütkepohl IV

where $\Phi_i = J_1 \mathbf{B}^i J_1'$ follows as in (2.1.17). Using the LS estimator $\widehat{\boldsymbol{\beta}}$ with asymptotic covariance matrix $\Sigma_{\boldsymbol{\beta}} = \Gamma^{-1} \otimes \Sigma_u$ (see Proposition 3.1), the matrix $\Omega(h)$ is seen to be

$$\begin{aligned}
\Omega(h) &= E\left[\frac{\partial y_t(h)}{\partial \boldsymbol{\beta}'} (\Gamma^{-1} \otimes \Sigma_u) \frac{\partial y_t(h)'}{\partial \boldsymbol{\beta}} \right] \\
&= \sum_{i=0}^{h-1} \sum_{j=0}^{h-1} E(Z_t' (\mathbf{B}')^{h-1-i} \Gamma^{-1} \mathbf{B}^{h-1-j} Z_t) \otimes \Phi_i \Sigma_u \Phi_j' \\
&= \sum_{i} \sum_{j} E[\mathrm{tr}(Z_t' (\mathbf{B}')^{h-1-i} \Gamma^{-1} \mathbf{B}^{h-1-j} Z_t)] \Phi_i \Sigma_u \Phi_j' \\
&= \sum_{i} \sum_{j} \mathrm{tr}[(\mathbf{B}')^{h-1-i} \Gamma^{-1} \mathbf{B}^{h-1-j} E(Z_t Z_t')] \Phi_i \Sigma_u \Phi_j' \\
&\models \sum_{i=0}^{h-1} \sum_{j=0}^{h-1} \mathrm{tr}[(\mathbf{B}')^{h-1-i} \Gamma^{-1} \mathbf{B}^{h-1-j} \Gamma] \Phi_i \Sigma_u \Phi_j', \qquad (3.5.12)
\end{aligned}$$

provided y_t is stable so that

$$\Gamma := \mathrm{plim}(ZZ'/T) = E(Z_t Z_t').$$

Here $Z := (Z_0, \dots, Z_{T-1})$ is the $((Kp+1) \times T)$ matrix defined in (3.2.1).
For example, for $h = 1$,

$$\Omega(1) = (Kp+1)\Sigma_u.$$

Hence, the approximation

$$\Sigma_{\hat{y}}(1) = \Sigma_u + \frac{Kp+1}{T}\Sigma_u = \frac{T+Kp+1}{T}\Sigma_u \qquad (3.5.13)$$

of the MSE matrix of the 1-step forecast with estimated coefficients is obtained. This expression shows that the contribution of the estimation variability to the forecast MSE matrix $\Sigma_{\hat{y}}(1)$ depends on the dimension K of the process, the VAR order p, and the sample size T used for estimation. It can be quite substantial if the sample size is small or moderate. For instance, considering a three-dimensional process of order 8 which is estimated from 15 years of quarterly data (i.e., $T = 52$ plus 8 presample values needed for LS estimation), the 1-step forecast MSE matrix Σ_u for known processes is inflated by a factor $(T + Kp + 1)/T = 1.48$. Of course, this approximation is derived from asymptotic theory so that its small sample validity is not guaranteed. We will take a closer look at this problem shortly. Obviously, the inflation factor $(T+Kp+1)/T \to 1$ for $T \to \infty$. Thus the MSE contribution due to sampling variability vanishes if the sample size gets large. This result is a consequence of estimating the VAR coefficients consistently. An expression for $\Omega(h)$ can also be derived on the basis of the mean-adjusted form of the VAR process (see Problem 3.9).

Forecasting - Lütkepohl V

In practice, for $h > 1$, it will not be possible to evaluate $\Omega(h)$ without knowing the AR coefficients summarized in the matrix B. A consistent estimator $\widehat{\Omega}(h)$ may be obtained by replacing all unknown parameters by their LS estimators, that is, $\mathbf{B}$ is replaced by $\widehat{\mathbf{B}}$ which is obtained by using $\widehat{B}$ for B, Σ_u is replaced by $\widehat{\Sigma}_u, \Phi_i$ is estimated by $\widehat{\Phi}_i = J_1 \widehat{\mathbf{B}}^i J_1'$, and Γ is estimated by $\widehat{\Gamma} = ZZ'/T$. The resulting estimator of $\Sigma_{\widehat{y}}(h)$ will be denoted by $\widehat{\Sigma}_{\widehat{y}}(h)$ in the following.

The foregoing discussion is of importance in setting up interval forecasts. Assuming that y_t is Gaussian, an approximate $(1-\alpha)100\%$ interval forecast, h periods ahead, for the k-th component $y_{k,t}$ of y_t is

$$\widehat{y}_{k,t}(h) \pm z_{(\alpha/2)} \widehat{\sigma}_k(h) \tag{3.5.14}$$

or

$$\left[\widehat{y}_{k,t}(h) - z_{(\alpha/2)} \widehat{\sigma}_k(h),\ \widehat{y}_{k,t}(h) + z_{(\alpha/2)} \widehat{\sigma}_k(h)\right], \tag{3.5.15}$$

where $z_{(\alpha)}$ is the upper $\alpha 100$-th percentile of the standard normal distribution and $\widehat{\sigma}_k(h)$ is the square root of the k-th diagonal element of $\widehat{\Sigma}_{\widehat{y}}(h)$. Using Bonferroni's inequality, approximate joint confidence regions for a set of forecasts can be obtained just as described in Section 2.2.3 of Chapter 2.

3.5.3 An Example

To illustrate the previous results, we consider again the investment/income/consumption example of Section 3.2.3. Using the VAR(2) model with the coefficient estimates given in (3.2.22) and

$$y_{T-1} = y_{72} = \begin{bmatrix} .02551 \\ .02434 \\ .01319 \end{bmatrix} \quad \text{and} \quad y_T = y_{73} = \begin{bmatrix} .03637 \\ .00517 \\ .00599 \end{bmatrix}$$

results in forecasts

$$\begin{aligned} \widehat{y}_T(1) &= \widehat{\nu} + \widehat{A}_1 y_T + \widehat{A}_2 y_{T-1} = \begin{bmatrix} -.011 \\ .020 \\ .022 \end{bmatrix}, \\ \widehat{y}_T(2) &= \widehat{\nu} + \widehat{A}_1 \widehat{y}_T(1) + \widehat{A}_2 y_T = \begin{bmatrix} .011 \\ .020 \\ .015 \end{bmatrix}, \end{aligned} \tag{3.5.16}$$

and so on.

The estimated forecast MSE matrix for $h = 1$ is

$$\begin{aligned} \widehat{\Sigma}_{\widehat{y}}(1) &= \frac{T+Kp+1}{T} \widehat{\Sigma}_u = \frac{73+6+1}{73} \widehat{\Sigma}_u \\ &= \begin{bmatrix} 23.34 & .785 & 1.351 \\ .785 & 1.505 & .674 \\ 1.351 & .674 & .978 \end{bmatrix} \times 10^{-4}, \end{aligned} \tag{3.5.17}$$

Table of Contents

Impulse response functions I

- A stationary VAR can be written in $MA(\infty)$ form as

$$\mathbf{y}_t = \Phi^{-1}(L)\epsilon_t = \Theta(L)\epsilon_t \quad \text{with} \quad \epsilon_t \sim WN(0, \Sigma) \tag{13}$$

- Since Σ is positive definite, there exists a non-singular matrix P such that

$$P\Sigma P' = I \tag{14}$$

- Can rewrite (13) as

$$\begin{aligned} \mathbf{y}_t &= \Theta(L)P^{-1}P\epsilon_t = \Psi(L)\upsilon_t \qquad (15) \\ \upsilon_t &= P\epsilon_t \\ E(\upsilon_t) &= 0 \ \text{ and } \ E(\upsilon_t\upsilon_t') = P\Sigma P' = I \end{aligned}$$

- Equation (15) is the $MA(\infty)$ representation of the model

$$P\Phi(L)\mathbf{y}_t = \upsilon_t \tag{16}$$

Impulse response functions II

- which is a Structural VAR ($SVAR$) as there are contemporaneous relationships among the variables because of the P matrix
- The orthogonal errors υ_t are the structural (economic) shocks.
- How the variable under analysis will react to the structural shocks by writing the polynomial $\Psi(L)$ as

$$\begin{aligned} \Psi(L) &= P^{-1} - \Theta_1 P^{-1} L - \Theta_2 P^{-1} L^2 - \cdots \\ &= \Psi_1 - \Psi_2 L - \Psi_3 L^2 \end{aligned} \tag{17}$$

Impulse response functions III

- A shock is a vector with one element equal to one and all the others equal to zero, e.g.

$$\upsilon_{1,t+1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; \upsilon_{2,t+1} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; \cdots, \; \upsilon_{m,t+1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{18}$$

- The non-zero elements can also be set to any other constant such as the standard deviation of the corresponding element of ϵ_t
- Since the system is linear, the size of the shock is irrelevant because the response is just proportional to the size of the shock.

Impulse response functions IV

- The response in period $t+i$ of y to a shock in period $t+1$ will be:

$$
\begin{aligned}
\frac{\partial \mathbf{y}_{t+1}}{\partial \upsilon_{t+1}} &= P^{-1} = \Psi_1 \\
\frac{\partial \mathbf{y}_{t+2}}{\partial \upsilon_{t+1}} &= -\Psi_2 \\
\frac{\partial \mathbf{y}_{t+3}}{\partial \upsilon_{t+1}} &= -\Psi_3 \\
&\vdots
\end{aligned}
\tag{19}
$$

- where for example

$$
\Psi_1 = \begin{bmatrix} \frac{\partial y_{1,t+1}}{\partial \upsilon_{1,t+1}} & \cdots & \frac{\partial y_{1,t+1}}{\partial \upsilon_{m,t+1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_{m,t+1}}{\partial \upsilon_{1,t+1}} & & \frac{\partial y_{m,t+1}}{\partial \upsilon_{m,t+1}} \end{bmatrix} \tag{20}
$$

Impulse response functions V

- and the terms $\frac{\partial y_{k,t+i}}{\partial \upsilon_{j,t+1}}$ are known as the impulse response in period $t+i$ of variable k to the shock j with $k, j = 1, \ldots, m$ and $i = 1, 2, \cdots$
- The collection of all impulse responses is known as the impulse response function (**IRF**)
- The matrix P is not unique as a consequence the IRF is also not unique.
- Since Σ has $m(m+1)/2$ distinct elements this is the maximum number of unrestricted elements in P. So typically P is chosen to be a triangular matrix, i.e.

$$P = \begin{bmatrix} p_{11} & 0 & \ldots & 0 \\ p_{21} & p_{22} & \ldots & 0 \\ \vdots & & \ddots & \\ p_{m1} & p_{m2} & \ldots & p_{mm} \end{bmatrix}$$

Impulse response functions VI

- where the elements p_{ij} are such that $P\Sigma P' = I$

Table of Contents

Forecast error variance decomposition

- The $h-$steps ahead forecast error can be written as

$$\mathbf{e}_{T+h} = \epsilon_{T+h} + \Theta_1 \epsilon_{T+h-1} + \cdots + \Theta_{h-1}\epsilon_{T+1}$$

- so that

$$Var(\mathbf{e}_{T+h}) = \Sigma + \Theta_1 \Sigma \Theta_1' + \cdots + \Theta_{h-1}\Sigma\Theta_{h-1}'$$

- which can be rewritten as

$$\begin{aligned} Var(\mathbf{e}_{T+h}) &= P^{-1}P\Sigma P'(P^{-1})' + \Theta_1 P^{-1}P\Sigma P'(P^{-1})'\Theta_1' + \\ &\quad + \cdots + \Theta_{h-1}P^{-1}P\Sigma P'(P^{-1})'\Theta_{h-1}' \\ &= \Psi_1\Psi_1' + \Psi_2\Psi_2' + \cdots + \Psi_h\Psi_h' \end{aligned}$$

- It follows that

$$\Psi_{ij,1}^2 + \Psi_{ij,2}^2 + \cdots + \Psi_{ij,h}^2 \tag{21}$$

- represents the contribution of the innovations in the j^{th} variable in explaining the $h-$steps ahead forecast error variance for y_i with $i, j = 1, 2, \ldots, m$
- This is called forecast error variance decomposition (**FEVD**)

Table of Contents

Structural VARs with long-run restrictions I

- [Blanchard and Quah, 1989] explain how to identify $VAR's$ imposing long-run restrictions
- Consider the bivariate $VAR(1)$ for GDP growth and unemployment defined as

$$\mathbf{y}_t = A\mathbf{y}_{t-1} + B\upsilon_t \tag{22}$$

- where υ_t are the structural shocks and $B = P^{-1}$ as before.
- The MA representation is

$$\mathbf{y}_t = (I - AL)^{-1} B\upsilon_t \tag{23}$$

- the cumulative long-run response of $\mathbf{y}$ to shocks in υ is

$$(I - AL)^{-1} B = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \tag{24}$$

Structural VARs with long-run restrictions II

- The (i, j) element of the above matrix captures the long-run effects of shock j on variable i
- From the variance-covariance matrix Σ we obtain 3 parameters so in order to indentify B we need to impose one a priori restriction on its elements.
- One possibility is to impose $b_{21} = 0$ as in the Cholesky decomposition .
- Another possibility if $\mathbf{y}_t = [\text{growth}_t : \text{unemployment}_t]'$ then the demand shock υ_{1t} has no long-run effect on growth if

$$\pi_{11} b_{11} + \pi_{12} b_{21} = 0$$

- In fact this restriction implies

$$(I - AL)^{-1} B = \begin{bmatrix} 0 & * \\ * & * \end{bmatrix} \tag{25}$$

Structural VARs with long-run restrictions III

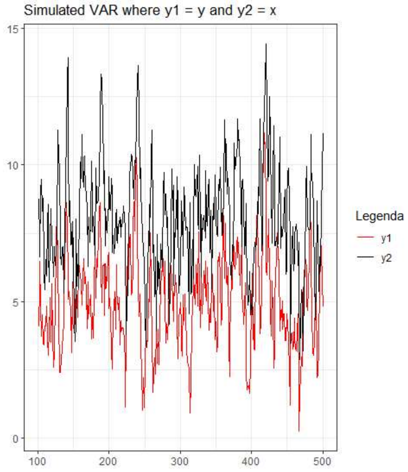
- rather than imposing $b_{21} = 0$, restriction (40) implies that b_{11} and b_{21} are linked by that linear relationship.

Table of Contents

Stationary VAR with simulated data I

- Use the following commands in R to generate a Stationary VAR
- Simulated VAR

Stationary VAR with simulated data II



Stationary VAR with simulated data III

- The VAR(1) model is given by:

$$\begin{aligned} y_t &= 1 + 0.8y_{t-1} + \epsilon_{y,t} \\ x_t &= 1 + 0.6y_{t-1} + 0.5x_{t-1} + \epsilon_{x,t} \end{aligned} \tag{26}$$

- with

$$\upsilon_t = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix} \epsilon_t \quad \text{and} \quad \upsilon_t \overset{i.i.d.}{\sim} N(0, I) \tag{27}$$

- Then $\epsilon_t \overset{i.i.d.}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} ; \begin{bmatrix} 1.25 & 0.5 \\ 0.5 & 1.0 \end{bmatrix} \right)$.
- Also note that both variables are stationary because the eigenvalues of the matrix $\begin{bmatrix} 0.8 & 0 \\ 0.6 & 0.5 \end{bmatrix}$ are $\lambda_1 = 0.8$ and $\lambda_2 = 0.5$.
- The sample size is $T = 600$ but the estimation sample will be $T = 101, 500$ and the forecast sample from $501, ..., 600$.

Stationary VAR with simulated data IV

- For the estimation period the best model, when BIC information criterion is used:
- selection AIC(n) HQ(n) SC(n) FPE(n)
- 1 1 1 1
- 1 2 3 4 5 6
- AIC(n) -0.02913672 -0.02089000 -0.004446732 0.001613109 0.02106193 0.03260397
- HQ(n) -0.00485097 0.01958625 0.052220010 0.074470349 0.11010967 0.13784220
- SC(n) 0.03211594 0.08119777 0.138476141 0.185371089 0.24565502 0.29803216
- FPE(n) 0.97128426 0.97932948 0.995570936 1.001631084 1.02131634 1.03319315

Stationary VAR with simulated data V

- 7 8 9 10 11 12
- AIC(n) 0.0414267 0.0566401 0.06334653 0.07343361 0.08358958 0.08967472
- HQ(n) 0.1628554 0.1942593 0.21715626 0.24343384 0.26978031 0.29205595
- SC(n) 0.3476900 0.4037385 0.45128004 0.50220223 0.55319331 0.60011355
- FPE(n) 1.0423771 1.0583937 1.06556303 1.07642502 1.08748523 1.09420960
- The estimated parameters for the best VAR(1) model are given by
- VAR(1)
- In order to compare the forecast of the best model with another model with mispecification, we also estimate a VAR(4)

Stationary VAR with simulated data VI

- VAR(4)
- Next test the specification of the VAR(1) model. We have the following results:
- Mispecification tests for VAR(1)
- And for the VAR(4)
- Mispecification tests for VAR(4)

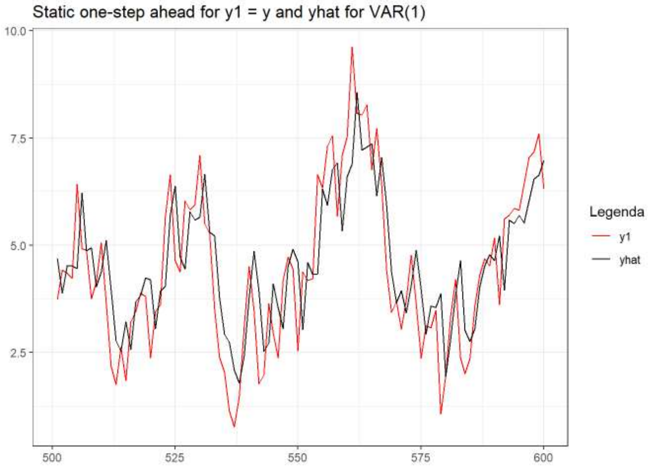
Table of Contents

- Use the following commands in R
- Forecasting VAR(1)
- RMSE and MAE for h-steps ahead VAR(1)
-
	y	x
RMSE	1.91	2.46
MAE	1.60	2.07

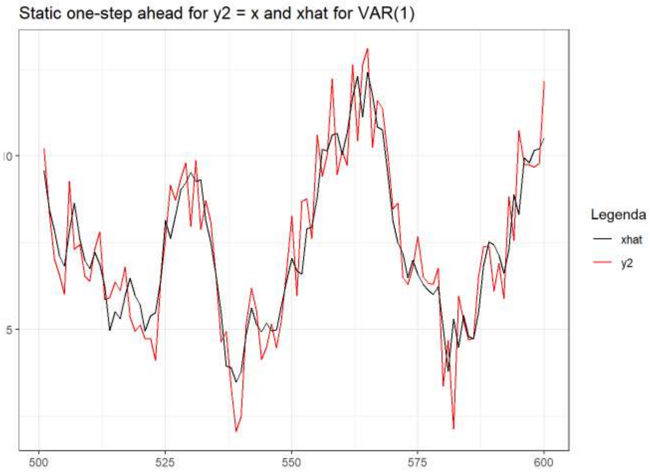
- RMSE and MAE for one-step ahead VAR(1)
-
	y	x
RMSE_dyn	1.11	0.97
MAE_dyn	0.89	0.79

- Comparison of one-step ahead and observations for VAR(1)

Stationary VAR with simulated data - Forecasting II



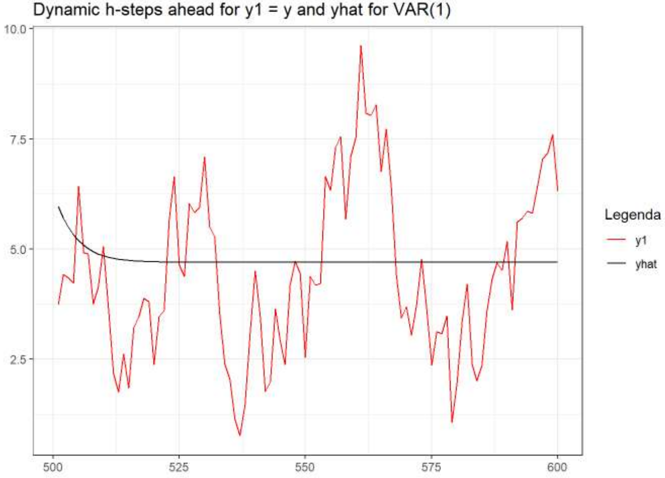
Stationary VAR with simulated data - Forecasting III

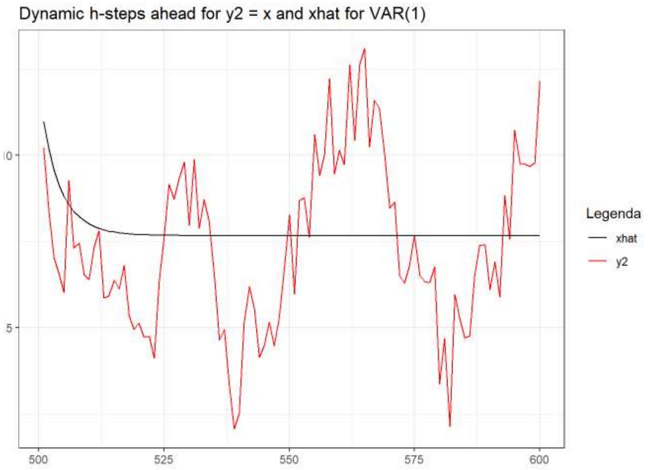


Stationary VAR with simulated data - Forecasting IV

- Comparison of h-steps ahead and observations for VAR(1)

Stationary VAR with simulated data - Forecasting V




Dynamic h-steps ahead for y2 = x and xhat for VAR(1)
10
5
500
525
550
575
600
Legenda
xhat
y2

- Estimate univariate ARMA models for y and x
- The following models are estimated

$$
\begin{aligned}
y_t &= \underset{(0.2360)}{4.9622} + u_t \\
u_t &= \underset{(0.0317)}{0.7684} u_{t-1} + \epsilon_t
\end{aligned}
$$

and

$$
\begin{aligned}
x_t &= \underset{(0.2374)}{8.0387} + u_t \\
u_t &= \underset{(0.0980)}{0.6841} u_{t-1} + \epsilon_t + \underset{(0.1082)}{0.0696} \epsilon_{t-1} + \\
&\quad \underset{(0.0893)}{0.0237} \epsilon_{t-2} + \underset{(0.0693)}{0.0189} \epsilon_{t-3}
\end{aligned}
$$

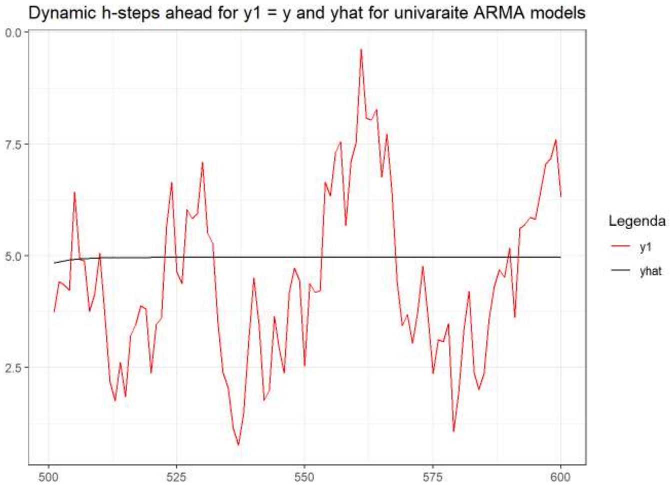
- RMSE and MAE for h-steps ahead univariate ARMA models for each variables

Stationary VAR with simulated data - Forecasting VIII

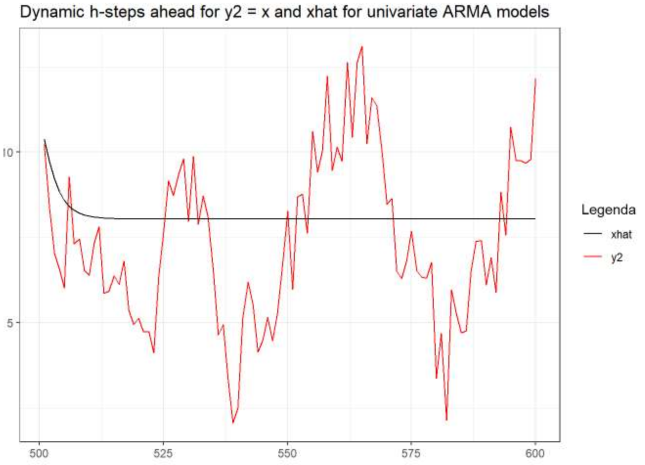
	y	x
RMSE	1.94	2.51
MAE	1.64	2.11

- Comparison of h-steps ahead and observations for unvariate ARIMA for each variables

Stationary VAR with simulated data - Forecasting IX



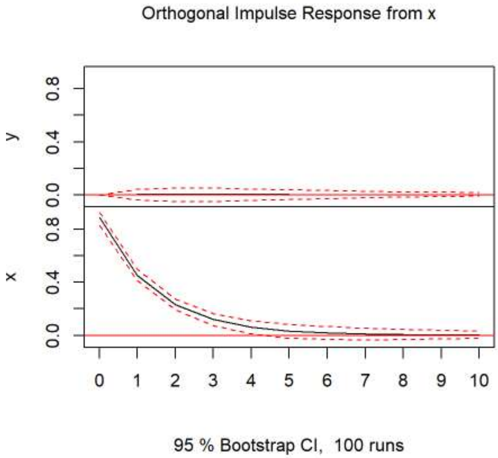
Stationary VAR with simulated data - Forecasting X

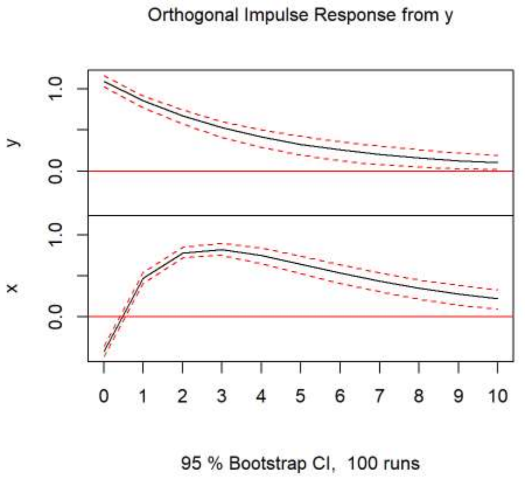


Stationary VAR with simulated data - Forecasting XI

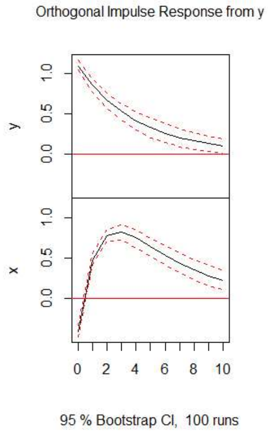
- Impulse Response Functions (IRF) for the system (y, x)

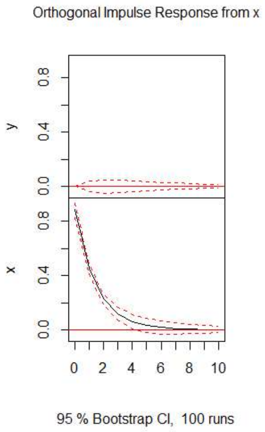
Stationary VAR with simulated data - Forecasting XII




Orthogonal Impulse Response from y
y
x
1.0
0.0
1.0
0.0
0
1
2
3
4
5
6
7
8
9
10
95 % Bootstrap CI, 100 runs

- Impulse Response Functions (IRF) for the system (x, y)


Orthogonal Impulse Response from y
y
x
1.0
0.5
0.0
1.0
0.5
0.0
0
2
4
6
8
10
95 % Bootstrap CI, 100 runs


Orthogonal Impulse Response from x
y
x
0.0
0.4
0.8
0.0
0.4
0.8
0
2
4
6
8
10
95 % Bootstrap CI, 100 runs

- RMSE and MAE for h-steps ahead VAR(4)

	y	x
RMSE4	1.92	2.47
MAE4	1.61	2.08

- RMSE and MAE for one-step ahead VAR(4)

	y	x
RMSE_dyn4	1.07	0.97
MAE_dyn4	0.88	0.78

Table of Contents

Reference I

Blanchard, O. J. and Quah, D. (1989).
The dynamic effetcs of aggregate demand and supply disturbances.
American Economic Review, 79:655–673.

Lütkepohl, H. (2007).
New Introduction to Multiple Time Series Analysis.
Spinger, 2nd edition edition.

Sims, C. A. (1980).
Macroeconomics and reality.
Econometrica, 48(1):1–48.