

11.2 Vector autoregressions

One limitation of the models that we have considered so far is that they impose a unidirectional relationship — the forecast variable is influenced by the predictor variables, but not vice versa. However, there are many cases where the reverse should also be allowed for — where all variables affect each other. In Section 9.2, the changes in personal consumption expenditure (C_t) were forecast based on the changes in personal disposable income (I_t). However, in this case a bi-directional relationship may be more suitable: an increase in I_t will lead to an increase in C_t and vice versa.

An example of such a situation occurred in Australia during the Global Financial Crisis of 2008–2009. The Australian government issued stimulus packages that included cash payments in December 2008, just in time for Christmas spending. As a result, retailers reported strong sales and the economy was stimulated. Consequently, incomes increased.

Such feedback relationships are allowed for in the vector autoregressive (VAR) framework. In this framework, all variables are treated symmetrically. They are all modelled as if they all influence each other equally. In more formal terminology, all variables are now treated as “endogenous.” To signify this, we now change the notation and write all variables as y s: $y_{1,t}$ denotes the t th observation of variable y_1 , $y_{2,t}$ denotes the t th observation of variable y_2 , and so on.

A VAR model is a generalisation of the univariate autoregressive model for forecasting a vector of time series.²² It comprises one equation per variable in the system. The right hand side of each equation includes a constant and lags of all of the variables in the system. To keep it simple, we will consider a two variable VAR with one lag. We write a 2-dimensional VAR(1) as

$$y_{1,t} = c_1 + \phi_{11,1}y_{1,t-1} + \phi_{12,1}y_{2,t-1} + e_{1,t} \quad (11.1)$$

$$y_{2,t} = c_2 + \phi_{21,1}y_{1,t-1} + \phi_{22,1}y_{2,t-1} + e_{2,t}, \quad (11.2)$$

where $e_{1,t}$ and $e_{2,t}$ are white noise processes that may be contemporaneously correlated. The coefficient $\phi_{ii,\ell}$ captures the influence of the ℓ th lag of variable y_i on itself, while the coefficient $\phi_{ij,\ell}$ captures the influence of the ℓ th lag of variable y_j on y_i .

If the series are stationary, we forecast them by fitting a VAR to the data directly (known as a “VAR in levels”). If the series are non-stationary, we take differences of the data in order to make them stationary, then fit a VAR model (known as a “VAR in differences”). In both cases, the models are estimated equation by equation using the principle of least squares. For each equation, the parameters are estimated by minimising the sum of squared $e_{i,t}$ values.

The other possibility, which is beyond the scope of this book and therefore we do not explore here, is that the series may be non-stationary but cointegrated, which means that there exists a linear combination of them that is stationary. In this case, a VAR specification that includes an error correction mechanism (usually referred to as a vector error correction model) should be included, and alternative estimation methods to least squares estimation should be used.²³

Forecasts are generated from a VAR in a recursive manner. The VAR generates forecasts for *each* variable included in the system. To illustrate the process, assume that we have fitted the 2-dimensional VAR(1) described in Equations (11.1)–(11.2), for all observations up to time T . Then the one-step-ahead forecasts are generated by

$$\begin{aligned}\hat{y}_{1,T+1|T} &= \hat{c}_1 + \hat{\phi}_{11,1}y_{1,T} + \hat{\phi}_{12,1}y_{2,T} \\ \hat{y}_{2,T+1|T} &= \hat{c}_2 + \hat{\phi}_{21,1}y_{1,T} + \hat{\phi}_{22,1}y_{2,T}.\end{aligned}$$

This is the same form as (11.1)–(11.2), except that the errors have been set to zero and parameters have been replaced with their estimates. For $h = 2$, the forecasts are given by

$$\begin{aligned}\hat{y}_{1,T+2|T} &= \hat{c}_1 + \hat{\phi}_{11,1}\hat{y}_{1,T+1} + \hat{\phi}_{12,1}\hat{y}_{2,T+1} \\ \hat{y}_{2,T+2|T} &= \hat{c}_2 + \hat{\phi}_{21,1}\hat{y}_{1,T+1} + \hat{\phi}_{22,1}\hat{y}_{2,T+1}.\end{aligned}$$

Again, this is the same form as (11.1)–(11.2), except that the errors have been set to zero, the parameters have been replaced with their estimates, and the unknown values of y_1 and y_2 have been replaced with their forecasts. The process can be iterated in this manner for all future time periods.

There are two decisions one has to make when using a VAR to forecast, namely how many variables (denoted by K) and how many lags (denoted by p) should be included in the system. The number of coefficients to be estimated in a VAR is equal to $K + pK^2$ (or $1 + pK$ per equation). For example, for a VAR with $K = 5$ variables and $p = 3$ lags,

there are 16 coefficients per equation, giving a total of 80 coefficients to be estimated. The more coefficients that need to be estimated, the larger the estimation error entering the forecast.

In practice, it is usual to keep K small and include only variables that are correlated with each other, and therefore useful in forecasting each other. Information criteria are commonly used to select the number of lags to be included.

VAR models are implemented in the **vars package** in R. It contains a function `VARselect()` for selecting the number of lags p using four different information criteria: AIC, HQ, SC and FPE. We have met the AIC before, and SC is simply another name for the BIC (SC stands for Schwarz Criterion, after Gideon Schwarz who proposed it). HQ is the Hannan–Quinn criterion, and FPE is the “Final Prediction Error” criterion.²⁴ Care should be taken when using the AIC as it tends to choose large numbers of lags. Instead, for VAR models, we prefer to use the BIC.

A criticism that VARs face is that they are atheoretical; that is, they are not built on some economic theory that imposes a theoretical structure on the equations. Every variable is assumed to influence every other variable in the system, which makes a direct interpretation of the estimated coefficients difficult. Despite this, VARs are useful in several contexts:

1. forecasting a collection of related variables where no explicit interpretation is required;
2. testing whether one variable is useful in forecasting another (the basis of Granger causality tests);
3. impulse response analysis, where the response of one variable to a sudden but temporary change in another variable is analysed;
4. forecast error variance decomposition, where the proportion of the forecast variance of each variable is attributed to the effects of the other variables.

Example: A VAR model for forecasting US consumption

```
library(vars)
```

```
VARselect(uschange[,1:2], lag.max=8,
  type="const")["selection"]
#> AIC(n)  HQ(n)  SC(n) FPE(n)
#>      5      1      1      5
```

The R output shows the lag length selected by each of the information criteria available in the **vars** package. There is a large discrepancy between the VAR(5) selected by the AIC and the VAR(1) selected by the BIC. This is not unusual. As a result we first fit a VAR(1), as selected by the BIC.

```
var1 <- VAR(uschange[,1:2], p=1, type="const")
serial.test(var1, lags.pt=10, type="PT.asymptotic")
var2 <- VAR(uschange[,1:2], p=2, type="const")
serial.test(var2, lags.pt=10, type="PT.asymptotic")
```

In similar fashion to the univariate ARIMA methodology, we test that the residuals are uncorrelated using a Portmanteau test²⁵. Both a VAR(1) and a VAR(2) have some residual serial correlation, and therefore we fit a VAR(3).

```
var3 <- VAR(uschange[,1:2], p=3, type="const")
serial.test(var3, lags.pt=10, type="PT.asymptotic")
#>
#> Portmanteau Test (asymptotic)
#>
#> data: Residuals of VAR object var3
#> Chi-squared = 34, df = 28, p-value = 0.2
```

The residuals for this model pass the test for serial correlation. The forecasts generated by the VAR(3) are plotted in Figure 11.10.

```
forecast(var3) %>%
  autoplot() + xlab("Year")
```

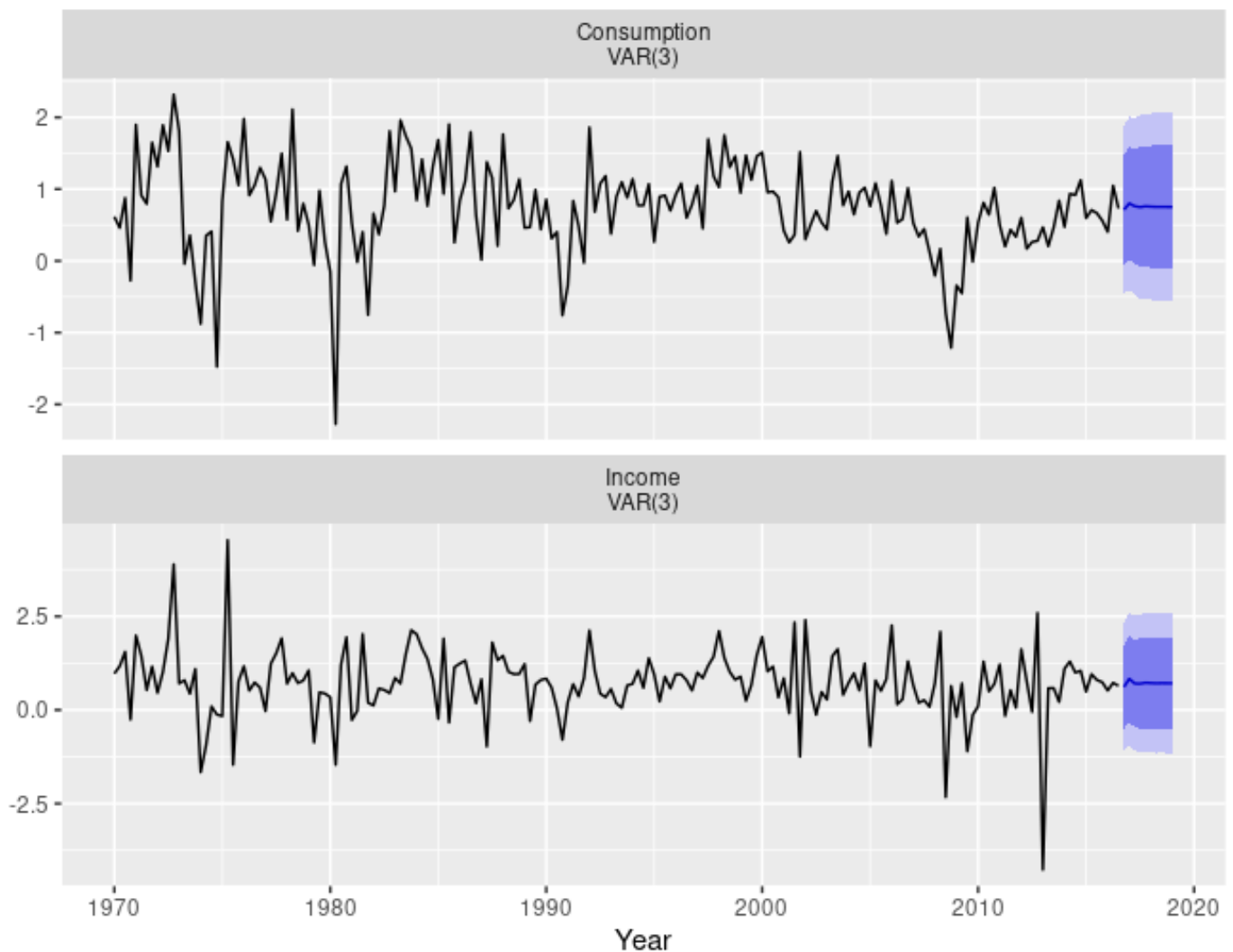


Figure 11.10: Forecasts for US consumption and income generated from a VAR(3).

22. A more flexible generalisation would be a Vector ARMA process. However, the relative simplicity of VARs has led to their dominance in forecasting. Interested readers may refer to [Athanasopoulos, Poskitt, & Vahid \(2012\)](#).↵
23. Interested readers should refer to [Hamilton \(1994\)](#) and [Lütkepohl \(2007\)](#).↵
24. For a detailed comparison of these criteria, see Chapter 4.3 of [Lütkepohl \(2005\)](#).↵
25. The tests for serial correlation in the “vars” package are multivariate generalisations of the tests presented in Section 3.3.↵