3. VAR: FORECASTING AND IMPULSE RESPONSE FUNCTIONS
1 Forecasting with VAR models

We define the Mean Square Error for a predictor of $Y_{t+h}$, $Y_t(h)$ as

$$MSE(Y_t(h)) = E[(Y_{t+h} - Y_t(h))(Y_{t+h} - Y_t(h))']$$

In what follows we will consider predictor which are optimal in the sense that they minimize the MSE, i.e. minimizes the MSE of each component of $Y_{t+h}$.

Let us consider the VAR($p$) in companion form

$$Y_t = AY_{t-1} + e_t$$  \hspace{1cm} (1)

where $e_t$ is white noise. The $h$-step ahead predictor of $Y_{t+h}$, $Y_t(h)$, conditional on the information available at time $t$ is given by

$$Y_t(h) = A^hY_t = AY_t(h - 1)$$  \hspace{1cm} (2)

where the first $n$ rows of $Y_t(h)$ represent the optimal forecast of $Y_{t+h}$. From (37) it is easy to compute recursively the forecast for $Y_{t+h}$ at any horizon, given given by the first $n$ elements of $Y_t(h)$. 

2
The predictor in the previous slide is optimal in the sense that it delivers the minimum MSE among those that are linear functions of $Y$. To get the intuition let us consider for instance a simple VAR(1) process

$$Y_t = A_1 Y_{t-1} + e_t$$

It follows that

$$Y_{t+h} = A_1^h Y_t + \sum_{i=0}^{h-1} A_1^i e_{t+h-i}$$

Thus for a predictor

$$Y_t(h) = B_0 Y_t + B_1 Y_{t-1} + \ldots$$

the forecast error is given by

$$Y_{t+h} - Y_t(h) = \sum_{i=0}^{h-1} A_1^i e_{t+h-i} + (A_1^i - B_0) Y_t - \sum_{i=1}^{\infty} B_i Y_{t-i}$$
The MSE is given by

\[
MSE(Y_t(h)) = E \left[ \left( \sum_{i=0}^{h-1} A_i e_{t+h-i} \right) \left( \sum_{i=0}^{h-1} A_i e_{t+h-i} \right)' \right] + \\
E \left[ \left( A^h - B_0 \right) Y_t - \sum_{i=1}^\infty B_i Y_{t-i} \right] \left[ \left( A^h - B_0 \right) Y_t - \sum_{i=1}^\infty B_i Y_{t-i} \right]'
\]

Clearly the MSE is minimal for \( B_0 = A^h_1 \) and \( B_i = 0 \).
Using
\[ Y_t(h) = A^h Y_t + \sum_{i=0}^{h-1} A^i e_{t+h-i} \]  
we get the forecast error

\[ Y_{t+h} - Y_t(h) = \sum_{i=0}^{h-1} A^i e_{t+h-i} \]  

From the forecast error it is easy to obtain the Mean Square Error, the covariance of the forecast error,

\[
MSE[Y_t(h)] = E[Y_{t+k} - Y_t(h)] (Y_{t+k} - Y_t(h))' = \tilde{\Sigma}(h) = \sum_{i=0}^{h-1} A^i \tilde{\Omega} A^{i'} \\
= \tilde{\Sigma}(h - 1) + A^{h-1} \tilde{\Omega} A^{h-1'}
\]

the MSE for will be the first upper left $n \times n$ matrix. Notice that the MSE is non decreasing and that as $h \to \infty$ will approach the variance of $Y_t$. 

5
If $e_t$ is Gaussian WN we can easily construct confidence bands for the point forecast. Actually

$$Y_{t+h} - Y_t(h) \sim N(0, \Sigma(h))$$

This implies that the forecast errors of the individual components are normal so that

$$\frac{Y_{k,t+h} - Y_{k,t}(h)}{\sqrt{\Sigma_{kk}(h)}} \sim N(0, 1)$$

Denoting $z_\alpha$ the upper $\alpha100$ percentage point of the normal distribution we get

$$1 - \alpha = PR\{-z_{\alpha/2} \leq \frac{Y_{k,t+h} - Y_{k,t}(h)}{\sqrt{\Sigma_{kk}}(h)} \leq z_{\alpha/2}\}$$

$$= PR\{Y_{k,t}(h) - z_{\alpha/2} \sqrt{\Sigma_{kk}} \leq Y_{k,t+h} \leq Y_{k,t}(h) + z_{\alpha/2} \sqrt{\Sigma_{kk}}\}$$

(5)
Hence a \((1 - \alpha)100\%\) interval forecast, \(h\) periods ahead, for the \(k\)th component of \(Y_t\) is

\[
\left[ Y_{k,t}(h) - z_{\alpha/2} \sqrt{\Sigma_{kk}}, \quad Y_{k,t}(h) + z_{\alpha/2} \sqrt{\Sigma_{kk}} \right] \quad (6)
\]

In case \(e_t\) is not Gaussian we can still construct confidence bands for point forecast using the methods described for impulse response functions.
• What is the difference between in-sample and out-of-sample forecasting?

• How do we perform out-of-sample forecast? Suppose we have data for a sample of length \( T \). We proceed as follows:

1. We use the first \( T_0 \) observation to estimate the VAR and we forecast \( h \) periods ahead.

2. We update the sample with one observation (the length of the sample is now \( T_0 + 1 \)) and we perform the \( h \) periods ahead forecast.

3. We repeat step 2 for all the forecasting sample period up to the last date in the sample with one observation (the length of the sample is now \( T_0 + 1 \)) and we perform the \( h \) periods ahead forecast.
1.1 Comparing Forecasting Performance

Out-of-sample forecast is particularly important in order to compare the forecasting performance of different model. One alternative in order to judge the forecasting ability of a model for variable $k$ is to compare the the mean square forecast error for horizon $h$ computed as

$$
\frac{1}{T - T_0} \sum_{t=T_0}^{T} (Y_{kt+h} - Y_{kt}(h))^2
$$

where $T$ is the length of the out of sample forecasting period. or the root mean square forecast error as

$$\sqrt{\frac{1}{T - T_0} \sum_{t=T_0}^{T} (Y_{kt+h} - Y_{kt}(h))^2}$$
1.2 Application: Stock and Watson Monetary VAR

- SW consider a trivariate VAR(4) model including inflation, unemployment and the short term interest rate.


- In addition, the authors estimate compute the RMSE for a univariate AR(4) forecast and a naive random walk (no change) forecast.
Table 2

<table>
<thead>
<tr>
<th>Forecast Horizon</th>
<th>Inflation Rate</th>
<th></th>
<th>Unemployment Rate</th>
<th></th>
<th>Interest Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RW</td>
<td>AR</td>
<td>VAR</td>
<td>RW</td>
<td>AR</td>
</tr>
<tr>
<td>2 quarters</td>
<td>0.82</td>
<td>0.70</td>
<td>0.68</td>
<td>0.34</td>
<td>0.28</td>
</tr>
<tr>
<td>4 quarters</td>
<td>0.73</td>
<td>0.65</td>
<td>0.68</td>
<td>0.62</td>
<td>0.52</td>
</tr>
<tr>
<td>8 quarters</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>1.12</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Notes: Entries are the root mean squared error of forecasts computed recursively for univariate and vector autoregressions (each with four lags) and a random walk ("no change") model. Results for the random walk and univariate autoregressions are shown in columns labeled RW and AR, respectively. Each model was estimated using data from 1960:I through the beginning of the forecast period. Forecasts for the inflation rate are for the average value of inflation over the period. Forecasts for the unemployment rate and interest rate are for the final quarter of the forecast period.
2 Impulse Response Functions

Impulse response functions represent the mechanisms through which shock spread over time. Let us consider the Wold representation of a covariance stationary VAR($p$),

$$Y_t = C(L)\epsilon_t$$

$$= \sum_{i=0}^{\infty} C_i \epsilon_{t-i}$$ \hspace{1cm} (7)

The matrix $C_j$ has the interpretation

$$\frac{\partial Y_t}{\partial \epsilon'_{t-j}} = C_j$$ \hspace{1cm} (8)

or

$$\frac{\partial Y_{t+j}}{\partial \epsilon'_t} = C_j$$ \hspace{1cm} (9)

That is, the row $i$, column $k$ element of $C_j$ identifies the consequences of a unit increase in the $k$th variable’s innovation at date $t$ for the value of the $i$th variable at time $t + j$ holding all other innovation at all dates constant.
Example 1 Let us assume that the estimated matrix of VAR coefficients is

\[
A = \begin{pmatrix} 0.8 & 0.1 \\ -0.2 & 0.5 \end{pmatrix}
\]  \quad (10)

with eigenvalues 0.8562 and 0.4438. We generate impulse response functions of the Wold representation

\[
C_j = A^j
\]
Figure 3: Impulse response functions
Example 2 Let us now assume that the second variable does not Granger cause the first one so that

\[
A = \begin{pmatrix}
0.8 & 0 \\
-0.2 & 0.5
\end{pmatrix}
\]  

with eigenvalues 0.8 and 0.5. Impulse response functions are plotted in the next figure.
Figure 4: Impulse response functions when the second variable does not Granger cause the first one.
2.1 Error Bands for Impulse Response Functions

To make inference about the effects of the shocks we need to quantify the uncertainty around the point estimate, we need the confidence bands for the impulse response functions. We will see three different methods to construct confidence bands

1. Asymptotic bands
2. Montecarlo bands
3. Bootstrap bands
2.2 Error Bands for Impulse Response Functions: Asymptotic

Hamilton (1994) shows that

\[ \sqrt{T}(\text{vec}(\hat{C}_s) - \text{vec}(C_s)) \xrightarrow{L} N \left( \mathbf{0}, G_s(\Omega \otimes Q^{-1})G'_s \right) \]

where

\[ G_s = \sum_{i=1}^{s} \left[ C_{i-1} \otimes (0_{n1} C'_{s-i} \ C'_{s-i-1} \ldots C'_{s-i-p+1}) \right] \]

is of dimension \( n^2 \times np \). Standard errors for an estimated impulse response coefficient are given by the square root of the associated diagonal element of

\[ \hat{G}_{s,T}(\hat{\Omega}_T \otimes Q_T^{-1})G'_{s,T} \]

where

\[ \hat{Q}_T = (1/T)\sum_{t=1}^{T} X_t X'_t \] and \( X'_t = [Y'_{t-1}, ..., Y'_{t-1}] \)
2.3 Error Bands for Impulse Response Functions: Montecarlo

Montecarlo method proceeds as follows.

1. Draw $\hat{\pi}^l$ from $N(\hat{\pi}, \hat{\Omega} \otimes \hat{\mathcal{Q}}^{-1})$.

2. Compute $C(L)^l$.

3. Repeat 1-2 $M$ (with $M$ big, i.e.1000) times.

4. For all the elements $C_{i,j,h}, i, j = 1, ..., n, h = 1, 2, ...$ of the impulse response functions collect the $\alpha$th and $1 - \alpha$th percentile across the $\ell$ draws as a confidence interval for $C_{i,j,h}$. 
2.4 Error Bands for Impulse Response Functions: Bootstrap

The idea behind bootstrapping (Runkle, 1987) is to obtain estimates of the small sample distribution for the impulse response functions without assuming that the shocks are Gaussian. Steps:

1. Estimate the VAR and save the $\hat{\pi}$ and the fitted residuals $\{\hat{u}_1, \hat{u}_2, ..., \hat{u}_T\}$.

2. Draw uniformly from $\{\hat{u}_1, \hat{u}_2, ..., \hat{u}_T\}$ and set $\tilde{u}_1^{(1)}$ equal to the selected realization and use this to construct

$$Y_1^{(1)} = \hat{A}_1 Y_0 + \hat{A}_2 Y_{-1} + ... + \hat{A}_p Y_{-p+1} + \tilde{u}_1^{(1)}$$  \hspace{1cm} (12)

3. Taking a second draw (with replacement) $\tilde{u}_1^{(2)}$ generate

$$Y_1^{(2)} = \hat{A}_1 Y_1 + \hat{A}_2 Y_0 + ... + \hat{A}_p Y_{-p+2} + \tilde{u}_1^{(2)}$$  \hspace{1cm} (13)

4. Proceeding in this fashion generate a sample of length $T \{Y_1^1, Y_2^1, ..., Y_T^1\}$ and use the sample to compute $\hat{\pi}^{(1)}$ and the implied impulse response functions $C^{(1)}(L)$.

5. Repeat steps (3) – (4) $M$ times and collect $M$ realizations of $C^{(l)}(L)$, $l = 1, ... M$ and takes for all the elements of the impulse response functions and
for all the horizons the $\alpha$th and $1 - \alpha$th percentile to construct confidence bands.
2.5 Error Bands for Impulse Response Functions: Bootstrap after Bootstrap

The bootstrap after bootstrap method (Kilian 1998) is a way to generate bias corrected confidence bands and it works as follows.

1. Estimate the VAR using OLS.

2. Generate 1000 draws for impulse response functions using bootstrap.

3. Correct the OLS estimator for the bias and get the bias corrected estimator
   \[ \hat{\beta}^* = \hat{\beta} - \text{Bias} \]
   where \( Bias = \bar{\beta}^\ell - \hat{\beta} \) where \( \bar{\beta}^\ell \) is the average of the parameter over the bootstrap replications.

4. Use \( \hat{\beta}^* \) to generate 2000 new bootstrap correcting each OLS estimate for the previously estimated bias.
3 Application 3: A Monetary VAR

We estimate the standard monetary VAR which includes real output growth, the inflation rate and the federal funds rate. These three variables are the core variables for monetary policy analysis in VAR models. Data are taken from the St.Louis Fed FREDII database.
3.1 Application 3: Bootstrap Bands
3.2 Application 3: Boostrap After Bootstrap Bands
4 Cumulated impulse response functions

Suppose $Y_t$ is a vector of trending variables (i.e. log prices and output) so we consider the first difference to reach stationarity. So the model is

$$\Delta Y_t = (1 - L)Y_t = C(L)\varepsilon_t$$

We know how to estimate, interpret, and conduct inference on $C(L)$. But suppose we are interested in the response of the levels of $Y_t$ rather than their first differences (the level of and prices rather than their growth rates). How can we find these responses? We transform the model

$$Y_t = Y_{t-1} + C(L)\varepsilon_t$$

The effect of $\varepsilon_t$ on $Y_t$ is $C_0$. Now substituting forward we obtain

$$Y_{t+1} = Y_{t-1} + C(L)\varepsilon_t + C(L)\varepsilon_{t+1}$$
$$= Y_{t-1} + C_0\varepsilon_{t+1} + (C_0 + C_1)\varepsilon_t + ...$$

and for two periods ahead

$$Y_{t+2} = Y_{t-1} + C(L)\varepsilon_t + C(L)\varepsilon_{t+1} + C(L)\varepsilon_{t+2}$$
\[ Y_{t-1} + C_0 \varepsilon_{t+2} + (C_0 + C_1) \varepsilon_{t+1} + (C_0 + C_1 + C_2) \varepsilon_{t+1} \ldots \]

so the effect of \( \varepsilon_t \) on \( Y_{t+1} \) are \( (C_0 + C_1) \) and on \( Y_{t+2} \) is \( (C_0 + C_1 + C_2) \). In general the effects of \( \varepsilon_t \) on \( Y_{t+j} \) are

\[ \tilde{C}_j = C_0 + C_1 + \ldots + C_j \]

defined as cumulated impulse response functions.