Lesson 18: Building a Vector Autoregressive Model

Umberto Triacca

Dipartimento di Ingegneria e Scienze dell’Informazione e Matematica
Università dell’Aquila,
umberto.triacca@ec.univaq.it
A simulated example

Consider the two series plotted in Figures 1 and 2.

Figure 1
A simulated example

Figure: 2
Before we can estimate a bivariate VAR model for the two series we must specify the order $p$.

The most common approach for model order selection involves selecting a model order that minimizes one or more information criteria evaluated over a range of model orders.
VAR order selection

Akaike Information Criterion (AIC):

\[ AIC(p) = \ln \left| \tilde{\Sigma} \right| + \frac{2K^2p}{T} \]

where \( K \) is the number of variables in the system, \( T \) is the sample size, and \( \tilde{\Sigma} \) is an estimate of the covariance matrix \( \Sigma \).

Bayesian Information Criterion (BIC):

\[ BIC(p) = \ln \left| \tilde{\Sigma} \right| + \frac{K^2p\ln T}{T} \]

Hannan-Quinn Criterion:

\[ HQC(p) = \ln \left| \tilde{\Sigma} \right| + \frac{K^2p2\ln\ln T}{T} \]

The key difference between the criteria is how severely each penalizes increases in model order (the second term).
VAR order selection

We pose $M = 4$ as upper bound for the VAR order. The asterisks below indicate the best (that is, minimized) values of the respective information criteria, $\text{AIC} = \text{Akaike criterion}$, $\text{BIC} = \text{Schwarz Bayesian criterion}$ and $\text{HQC} = \text{Hannan-Quinn criterion}$.

<table>
<thead>
<tr>
<th>lag</th>
<th>AIC</th>
<th>BIC</th>
<th>HQC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.794937*</td>
<td>5.844321*</td>
<td>5.814701*</td>
</tr>
<tr>
<td>2</td>
<td>5.816110</td>
<td>5.914877</td>
<td>5.855636</td>
</tr>
<tr>
<td>3</td>
<td>5.840238</td>
<td>5.988389</td>
<td>5.899528</td>
</tr>
<tr>
<td>4</td>
<td>5.854880</td>
<td>6.052415</td>
<td>5.933934</td>
</tr>
</tbody>
</table>

We choose $p = 1$. 
The estimated VAR model

\[
\begin{bmatrix}
\hat{y}_{1t} \\
\hat{y}_{2t}
\end{bmatrix} = 
\begin{bmatrix}
0.441184 & -0.855086 \\
-0.480172 & -0.315974
\end{bmatrix}
\begin{bmatrix}
y_{1t-1} \\
y_{2t-1}
\end{bmatrix}
\]
Once a VAR model has been estimated, it is of pivotal interest to see whether the residuals obey the models assumptions.

1. absence of serial correlation
2. absence of heteroscedasticity
3. if the error process is normally distributed
Graphical Diagnostics

A common overall diagnostic is the plot of the residuals. The two series of residuals are plotted in Figure.

Figure : 3
Figure: 3
Diagnostics

Figure: 3

ACF for uhat2

PACF for uhat2

Figure: 3
An Omnibus Test for Multivariate Normality

Doornik-Hansen test:

\[ \text{Chi-square}(4) = 8.42839 \ [0.0771] \]
A real example

Consider the bivariate system consisting of the quarterly U.S. unemployment rate \((y_1)\) and inflation rate \((y_2)\). The two series are plotted in Figures 1 and 2.

Figure 1
A real example

Figure: 2
A real example

The two series do not exhibit trend or regular seasonal patterns. Therefore we assume that they are the realization of a stationary VAR($p$) process.
VAR order selection

We pose $M = 4$ as upper bound for the VAR order.

<table>
<thead>
<tr>
<th>lag</th>
<th>AIC</th>
<th>BIC</th>
<th>HQC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.582769</td>
<td>5.682065</td>
<td>5.622957</td>
</tr>
<tr>
<td>2</td>
<td>4.999687</td>
<td>5.165179</td>
<td>5.066666</td>
</tr>
<tr>
<td>3</td>
<td>4.883780</td>
<td>5.115469</td>
<td>4.977550</td>
</tr>
<tr>
<td>4</td>
<td>4.857118</td>
<td>5.155005</td>
<td>4.977680</td>
</tr>
</tbody>
</table>

We choose $p = 3$. 
The estimated model

Equation 1: unemp

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>0.197067</td>
<td>0.0167</td>
</tr>
<tr>
<td>unemp(_t-1)</td>
<td>1.59939</td>
<td>0.0000</td>
</tr>
<tr>
<td>unemp(_t-2)</td>
<td>-0.751586</td>
<td>0.0000</td>
</tr>
<tr>
<td>unemp(_t-3)</td>
<td>0.100552</td>
<td>0.1491</td>
</tr>
<tr>
<td>infl(_t-1)</td>
<td>0.00931450</td>
<td>0.2854</td>
</tr>
<tr>
<td>infl(_t-2)</td>
<td>0.0128278</td>
<td>0.1562</td>
</tr>
<tr>
<td>infl(_t-3)</td>
<td>0.00227849</td>
<td>0.7958</td>
</tr>
</tbody>
</table>
### The estimated model

**Equation 2: infl**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>0.712422</td>
<td>0.2636</td>
</tr>
<tr>
<td>unemp$_{t-1}$</td>
<td>-1.5374</td>
<td>0.0064</td>
</tr>
<tr>
<td>unemp$_{t-2}$</td>
<td>1.65412</td>
<td>0.0860</td>
</tr>
<tr>
<td>unemp$_{t-3}$</td>
<td>-0.165631</td>
<td>0.7596</td>
</tr>
<tr>
<td>infl$_{t-1}$</td>
<td>0.336218</td>
<td>0.0000</td>
</tr>
<tr>
<td>infl$_{t-2}$</td>
<td>0.192954</td>
<td>0.0065</td>
</tr>
<tr>
<td>infl$_{t-3}$</td>
<td>0.343084</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
Collinearity is defined as correlation among the explanatory variables in a multiple regression model.

What does collinearity do in regression?

A consequence:

Collinearity $\Rightarrow$ increases SE's $\Rightarrow$ small $t$-ratio
An example

In a VAR model the regressors are likely to be highly colinear

⇒

the $t$-test on individual coefficients may not be reliable
Graphical Diagnostics

The two series of residuals are plotted in Figure 3.

Figure 3