Vector autoregression VAR

• So far we have focused mostly on models where \( y \) depends on past \( y \).
• More generally we might want to consider models for more than one variable.
• If we only care about forecasting one series but want to use information from another series we can estimate an ARMA model and include additional explanatory variables.

\[
01 1 1 0 1 1 \begin{align*}
\text{y}_t &= \beta_0 + \beta_1 \text{y}_{t-1} + \gamma \text{x}_{t-1} + \epsilon_t
\end{align*}
\]

• For example if \( y_t \) is the series of interest, but we think \( x_t \) might be useful we can estimate models like \( y_t = \beta_0 + \beta_1 y_{t-1} + \gamma x_{t-1} + \epsilon_t \).

• This model can be fit by least squares. Our dependent variable is \( y_t \) and the independent variables are \( y_{t-1} \) and \( x_{t-1} \).
• Once the model is fit, the one-step ahead forecast is given by:

\[ E(y_{t+1} \mid F_t) = \beta_0 + \beta_y E(y_t \mid F_t) + \gamma E(x_t \mid F_t) = \beta_0 + \beta_y y_t + \gamma x_t \]

• Just like the simple AR model, the one step ahead forecast variance is \( \sigma^2_\epsilon \).

• A joint model for \( x_t \) and \( y_t \) is required if we are interested in \emph{multiple step ahead forecasts}, or if we are interested in feedback effects from one process to the other.

\[ E(y_{t+2} \mid F_t) = \beta_0 + \beta_t E(y_{t+1} \mid F_t) + \gamma E(x_{t+1} \mid F_t) \]

What do we use here?

• Answer: We need a model for \( x \) as well.
The VAR(1) model

• Suppose that we have 2 variables that we observe at time period \( t \) and we consider the joint model:

\[
\begin{align*}
x_t &= \beta_0^x + \beta_1^x x_{t-1} + \beta_2^x y_{t-1} + \varepsilon_t^x \\
y_t &= \beta_0^y + \beta_1^y x_{t-1} + \beta_2^y y_{t-1} + \varepsilon_t^y
\end{align*}
\]

• Each equation is like an AR(1) model with one other explanatory variable.

• Each equation depends on its own lag and the lag of the other variable.

• We also now have two errors, one for each equation: \( \varepsilon_t^x \) and \( \varepsilon_t^y \)
• We can write the model in matrix notation:

\[
\begin{bmatrix}
  x_t \\
  y_t
\end{bmatrix} =
\begin{bmatrix}
  \beta_0^x \\
  \beta_0^y
\end{bmatrix} +
\begin{bmatrix}
  \beta_1^x & \beta_1^y \\
  \beta_2^x & \beta_2^y
\end{bmatrix}
\begin{bmatrix}
  x_{t-1} \\
  y_{t-1}
\end{bmatrix} +
\begin{bmatrix}
  \varepsilon_t^x \\
  \varepsilon_t^y
\end{bmatrix}
\]

• Since \( x \) depends on \( y \) and \( y \) depends on \( x \), a thorough understanding of dynamics and forecasting requires us to jointly consider \( x \) and \( y \) in the system of equations.

---

Simplify Notation

• By defining the following vectors and matrices we end up with a very simple form for the VAR(1) model. Let

\[
\begin{align*}
  y_t &= \begin{bmatrix} x_t \\ y_t \end{bmatrix}, \\
  v_t &= \begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^y \end{bmatrix}, \\
  \beta_0 &= \begin{bmatrix} \beta_0^x \\ \beta_0^y \end{bmatrix}, \text{ and } \\
  \beta_1 &= \begin{bmatrix} \beta_1^x & \beta_2^x \\ \beta_1^y & \beta_2^y \end{bmatrix}
\end{align*}
\]

\[
y_t = \beta_0 + \beta_1 y_{t-1} + v_t
\]
Assumptions on the errors

• Assumption 1: the errors are uncorrelated through time (i.e. \( \varepsilon_i^x \) is uncorrelated with \( \varepsilon_{i-j}^y \) and \( \varepsilon_i^x \) is uncorrelated with \( \varepsilon_{i-j}^x \) and \( \varepsilon_{i-j}^y \) for \( j>0 \)).

• Assumption 2: The \( \varepsilon \)'s are iid but may be contemporaneously correlated. \( \Omega \) denotes the variance covariance matrix of the error terms.

\[
\Omega = \begin{bmatrix}
\sigma_{\varepsilon_i^x}^2 & \sigma_{\varepsilon_i^x\varepsilon_i^y} \\
\sigma_{\varepsilon_i^x\varepsilon_i^y} & \sigma_{\varepsilon_i^y}^2
\end{bmatrix}
\]

Variance of \( x \)

Contemporaneous covariance

Variance of \( y \)

Trades and quotes example

• Let \( t \) denote the \( t \)th trade in a given asset.

• Let \( p_t \) denote the log transaction price associated with the \( t \)th trade and let \( r_t \) denote the return \( r_t = p_t - p_{t-1} \).

• Let \( x_t \) denote whether the trade was buyer or seller initiated (+1=buy, -1=sell).
Trades and quotes equations

• Then we stack $r_t$ and $x_t$ in a vector.
• Now at each point in time $t$ we have an observation for the return associated with the transaction price and a variable for whether the trade was buyer or seller initiated.
• Define:

$$
\begin{align*}
    y_t &= \begin{bmatrix} r_t \\ x_t \end{bmatrix}, \\
    v_t &= \begin{bmatrix} \varepsilon^r_t \\ \varepsilon^x_t \end{bmatrix}, \\
    \beta_0 &= \begin{bmatrix} \beta^r_0 \\ \beta^x_0 \end{bmatrix}, \text{ and } \\
    \beta_1 &= \begin{bmatrix} \beta^r_1 & \beta^r_2 \\ \beta^x_1 & \beta^x_2 \end{bmatrix}
\end{align*}
$$

A First order Vector Autoregression

VAR(1) is then defined as:

$$
y_t = \beta_0 + \beta_1 y_{t-1} + v_t \quad \text{where the elements of } v_t \text{ are iid}
$$

Normal with variance covariance matrix $E(v_t v'_t) = \Omega$

and $E(v_t v'_{t-j}) = 0$ for $j > 0$

• For this model, the dependence of each variable on the past is summarized by the matrix $\beta_1$, and (conditional on the past) the contemporaneous dependence is determined by the variance covariance matrix of $v_t$.
• In general, $y_t$ is an n-dimensional vector.
Interpreting the dynamics of a VAR(1)

- For the VAR model with one lag we write:
  \[ y_t = \beta_0 + \beta_1 y_{t-1} + v_t \]

- If we substitute in for \( y_{t-1} \) we get:
  \[ y_t = \beta_0 + \beta_1 (\beta_0 + \beta_1 y_{t-2} + v_{t-1}) + v_t \]
  \[ = \beta_0 + \beta_1 \beta_0 + \beta_1^2 y_{t-2} + \beta_1 v_{t-1} + v_t \]

- Doing this substitution over and over, \( k-1 \) times, we get: \[ y_t = \beta_0^* + \beta_1^* y_{t-k} + \sum_{j=0}^{k-1} \beta_j^* v_{t-j} \] where \[ \beta_0^* = \left( \sum_{j=0}^{k-1} \beta_j \right)^{-1} \]
\[ y_t = \beta_0^* + \beta_t^* y_{t-k} + \sum_{j=0}^{k-1} \beta_j^* v_{t-j} \text{ where } \beta_0^* = \left( \sum_{j=0}^{k-1} \beta_j^* \right) \]

- Now \( y_t \) is a function of \( y_{t-k} \) and a weighted sum of the intervening values of the error vector \( v_{t-j} \)
- How much does the value of \( y_t \) change when we increase one element of \( y_{t-k} \) by one unit, holding all previous \( y_{t-j} \) \( j>0 \) prices and trades fixed?
- The answer is obtained by taking the derivative of \( y_t \) with respect to \( y_{t-k} \) above.

\[ \frac{dy_{t,i}}{dy_{t-k,j}} = \left[ \beta_i^* \right]_{i,j} \]

This means \((i,j)\) element of matrix

- When we plot \( \left[ \beta_i^* \right]_{i,j} \) as a function of \( k \), we see how future values of variable \( i \) are impacted by a one unit change in variable \( j \), \( k \) periods in the past.
- This is called the impulse response function of variable \( i \) to a change in variable \( j \).
- This is the primary method used to understand the implied dynamics of a VAR model.
- It answers the basic question of how a change in one variable affects the system in the future.
• So powers of the matrix $\beta_1$ determine how a change in one variable today effects the future values.

• Taking powers accommodates the “feedback” effects from one equation to the other in the right way.

• A matrix algebra trick makes the dynamics and decay of these effects a little more transparent. The decay is determined by the eigenvalues. (more in a few slides)

• Notice that if we write the VAR in its infinite MA representation by continuing the substitution we get:

$$y_t = \beta_0 + \sum_{j=0}^{\infty} \beta_j v_{t-j}$$ where $\beta_0 = \mu = (I - \beta_1)^{-1} \beta_0$

• The derivative of $y_t$ with respect to elements of past values of $v_{t-k}$ is the same as the derivative with respect to $y_{t-k}$ obtained before:

$$\frac{dy_{t,i}}{dv_{t-k,j}} = [\beta_i^k]_{i,j}$$
The reason is that if we hold all past values of $y$ fixed and change an element of $y_t$ by one unit, that is equivalent to making the time $t$ error epsilon one unit larger.

\[ y_{t-k} = \beta_0 + \beta_1 y_{t-k-1} + v_{t-k} \]

Recall our eigenvalue decomposition

\[ \beta_1 = T\Lambda T^{-1} \quad \beta_1' = T\Lambda' T^{-1} \]

Where $\Lambda$ is the diagonal matrix \[ \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \] with $\lambda_1$ and $\lambda_2$ the eigenvalues of the matrix $\beta_1$.

- $T$ is a matrix with corresponding eigenvectors $x_1$ and $x_2$.

\[ y_t = \beta_1^* y_0 + \beta_0^* + \sum_{j=0}^{t-1} T\Lambda' T^{-1} v_{t-j} \]
The response of element \( i \), to a shock in element \( j \), \( k \) periods ago is given by:

\[
\frac{dy_{t,i}}{dv_{t-k,j}} = \begin{bmatrix} \beta^k_i \end{bmatrix}_{i,j} = \begin{bmatrix} T \Lambda^k T^{-1} \end{bmatrix}_{i,j} = a_{ij} \lambda^k_1 + b_{ij} \lambda^k_2
\]

Where \( a_{ij} \) and \( b_{ij} \) are determined by the elements of \( T \) and \( T^{-1} \)

Using the lag operator notation, we can write

\[
(I - \beta_i L) y_t = \beta_0 + v_t
\]

If the eigenvalues of \( \beta_1 \) lie inside the unit circle, then \( I - \beta_i L \) has an inverse given by:

\[
(I - \beta_i L)^{-1} = (I + \beta_i L + \beta_i^2 L^2 + \beta_i^3 L^3 \cdots) \text{ since}\n\]

\[
(I + \beta_i L + \beta_i^2 L^2 + \beta_i^3 L^3 \cdots) (I - \beta_i L) = I - \beta_i^{k+1} L^{k+1}
\]

\[
\lim_{\lambda \to 1} [I - \beta_i^{k+1} L^{k+1}] = I
\]
• Problem, \( \begin{pmatrix} \frac{dr_i}{d c_{i,t}} \end{pmatrix} = \begin{bmatrix} T \Lambda_i T^{-1} \end{bmatrix}_{i,2} \) assumes that when you change one innovation (error) you hold the other contemporaneous values of the innovations fixed.
• Changes in one error will be correlated with changes in the other errors in the same time period
• if the variance covariance matrix is not diagonal the errors are correlated so that movements in one innovation tend to be associated with movements in the others.

• For example if I tell you that the price went up more than expected (a positive error) that might also make us more likely to think that the trade was a buy (a positive error).
• We have to make a decision about how to handle this contemporaneous correlation.
  – Ignore? Assume that we change one error without changing others. That may be the question you want to ask in a policy situation where you believe you can control one variable.
  – Make some other assumption?
• **Common solution:** Take a stand on the way that the shocks propagate. That is, allocate the correlation to one direction only.

• Do trades contemporaneously cause prices to change, or the other way around?

• The answer to this question cannot be assessed purely with statistics.

• We must bring economic ideas to bear on the question.

• Choosing an order that shocks propagate is equivalent to a choice of orthogonalizing.

• **Market structure of trades and quotes:** There is a natural ordering in the market. The bid and ask are posted. The trader trades at the *prevailing* bid or ask price.

• Prices influence trades, but because prices are set prior to the trade, trades don’t contemporaneously affect prices.
• How do we impose the one directional “causality” in the innovations?
• We create a vector of orthonormal innovations and then pre-multiply by a lower triangular matrix.
• We want $v_t = Pu_t$ where $E(u_t'u_t') = I$ and $P$ is lower triangular.
• Since $P$ is lower triangular, a movement in the first element of $u_t$ changes all elements of $v_t$ while movements in the second element of $u_t$ only affect the second element of $v_t$.

• Any variance covariance matrix can be decomposed into $\Omega = PP'$ where $P$ is a unique lower triangular matrix (Choleski Decomposition). Notice that

$$P^{-1}\Omega P'^{-1} = I_n$$

• Next, we can get the vector $u_t$ by premultiply the error vectors by $P^{-1}$. $u_t = P^{-1}v_t$. Let’s verify:

$$E(u_t'u_t') = E(P^{-1}v_t'v_t') = P^{-1}\Omega P'^{-1} = P^{-1}\Omega P'^{-1} = I_n$$

• So the $u_t$ are now a vector of shocks with uncorrelated elements with unit variance.
• Finally, $u_t = P^{-1}v_t$ so $v_t = Pu_t$

• So $v_t = \begin{bmatrix} \varepsilon'_x \\ \varepsilon'_y \end{bmatrix} = \begin{bmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{bmatrix}\begin{bmatrix} u'_x \\ u'_y \end{bmatrix} = \begin{bmatrix} p_{11}u'_x \\ p_{21}u'_x + p_{22}u'_y \end{bmatrix}$

• $u_1$ and $u_2$ are uncorrelated.

• If we move $u_1$ by one unit, $\varepsilon^x$ changes by $p_{11}$ and $\varepsilon^y$ changes by $p_{21}$.

• If we move $u_2$ by one unit, $\varepsilon^x$ doesn’t change at all and $\varepsilon^y$ changes by $p_{22}$.

• Implicitly, we have imposed that the reason that $\varepsilon^x$ and $\varepsilon^y$ are correlated is because changes in $\varepsilon^x$ “cause” changes in $\varepsilon^y$, but changes in $\varepsilon^y$ do not cause changes in $\varepsilon^x$.

• Hence the lower triangular form of $P$ means that changes in the variables that appear in the top of the vector contemporaneously affect variables below.

• We can choose the ordering and therefore the order in which the shocks propagate.
• In Eviews, the ordering of the variables in the VAR equation definition determines the ordering of the shocks. The first variables affect the last variables.

\[ y_t = \beta_1' y_0 + \beta_0^* + \sum_{j=0}^{t-1} \beta_1^j v_{t-j} \Leftrightarrow y_t = \beta_1' y_0 + \beta_0^* + \sum_{j=0}^{t-1} \beta_1^j P u_{t-j} \]

• So P determines how moving one variable in period t affects other variables contemporaneously.

• Powers of \( \beta_1 \), determine how future values of y will change.
Back to our trade Example

\[
\begin{bmatrix}
  r_t \\
  x_t
\end{bmatrix} = \begin{bmatrix}
  .00017 \\
  .0128
\end{bmatrix} + \begin{bmatrix}
  -.27 \\
  -6.00
\end{bmatrix} \begin{bmatrix}
  r_{t-1} \\
  x_{t-1}
\end{bmatrix} + \begin{bmatrix}
  e^r_t \\
  e^x_t
\end{bmatrix}
\]

\[
\hat{\Omega} = \begin{bmatrix}
  0.0013 & 0.0062 \\
  0.0062 & 0.8393
\end{bmatrix}
\]

\[
P = \begin{bmatrix}
  .0361 & 0 \\
  0.172 & 0.899
\end{bmatrix}
\]

Expected response of \( y_t \) given a one s.d. increase in \( r_{t-k} \)

Recall \( y_t = \beta_t'y_0 + \beta_0' + \sum_{j=0}^{t} \beta_j' \frac{P}{u_t} \)

\[
y_t = \beta_t'y_0 + \beta_0' + \sum_{j=0}^{t} \begin{bmatrix}
  -.27 \\
  -6.00
\end{bmatrix} \begin{bmatrix}
  .0057 \\
  .39
\end{bmatrix}^j \begin{bmatrix}
  0.036 & 0 \\
  0.172 & 0.899
\end{bmatrix} \begin{bmatrix}
  u_{t-j}
\end{bmatrix}
\]

\[
E \left[ \frac{dy_t}{du_{t-1}} \mid F_t \right] = \begin{bmatrix}
  -.27 & .0057^k \\
  -6.00 & .39
\end{bmatrix} \begin{bmatrix}
  .036 & 0 \\
  0.172 & 0.899
\end{bmatrix} \begin{bmatrix}
  1
\end{bmatrix}
\]

\[
= \begin{bmatrix}
  -.27 & .0057^k \\
  -6.00 & .39
\end{bmatrix} \begin{bmatrix}
  .036 \\
  .172
\end{bmatrix}
\]

When we make the return one standard deviation larger, we also make the trade indicator larger.

How that shock affects \( y_t \) \( k \)-periods ahead. Initial 1 std shock to \( v_t \).
\[
\begin{bmatrix}
-0.27 & 0.0057 \\
-6.00 & 0.39
\end{bmatrix} = \begin{bmatrix}
0.081 & 0.011 \\
0.99 & 0.99
\end{bmatrix} \begin{bmatrix}
-0.22 & 0 \\
0.33 & 0.99
\end{bmatrix} \begin{bmatrix}
0.081 & 0.011
\end{bmatrix}
\]

Eigenvalues of $\hat{\Omega}$

Expected response of $y_t$ given a one s.d. increase in $x_{t-k}$

\[
y_t = \beta_0^* + \sum_{j=0}^{t} \begin{bmatrix}
-0.27 & 0.0057 \\
-6.00 & 0.39
\end{bmatrix} \begin{bmatrix}
0.036 & 0 \\
0.171 & 0.899
\end{bmatrix} u_{t-j}
\]

\[
E \left[ \frac{dy_t}{du_{t-k}} \middle| F_{t-k} \right] = \begin{bmatrix}
-0.27 & 0.0057^k \\
-6.00 & 0.39
\end{bmatrix} \begin{bmatrix}
0.036 & 0 \\
0.172 & 0.899
\end{bmatrix} \begin{bmatrix}
0 \\
1
\end{bmatrix}
\]

\[
E \left[ \frac{dy_t}{du_{t-k}} \middle| F_{t-k} \right] = \begin{bmatrix}
-0.27 & 0.0057^k \\
-6.00 & 0.39
\end{bmatrix} \begin{bmatrix}
0 \\
1
\end{bmatrix}
\]

where

\[
\begin{bmatrix}
-0.27 & 0.0057 \\
-6.00 & 0.39
\end{bmatrix} = \begin{bmatrix}
0.081 & 0.012 \\
0.99 & 0.99
\end{bmatrix} \begin{bmatrix}
-0.22 & 0 \\
0.33 & 0.99
\end{bmatrix} \begin{bmatrix}
0.081 & 0.012
\end{bmatrix}
\]

When we make the trade indicator larger we don’t change the value of the return.
When is the VAR(1) model weakly stationary?

• Let \((I - \beta_1 L)^{-1} = (I + \beta_1 L + \beta_1^2 L + \beta_1^3 L + \ldots)\)
  which exists as long as the eigenvalues of \(\beta_1\) lie inside the unit circle.

\[ \beta_1^k = T \Lambda^k T^{-1} \]

• The variance of each series will exist if all the eigenvalues lie inside the unit circle.
• So the VAR(1) model is stationary as long as the eigenvalues of \(\beta_1\) lie inside the unit circle.
VAR(p) model

\[ y_t = \beta_0 + \sum_{j=1}^{p} \beta_j y_{t-j} + v_t \]

where the elements of \( v_t \) are Normal with variance covariance matrix \( E(v_t v_t') = \Omega \).

In general, \( y_t \) is an n-dimensional vector and \( \Omega \) is nxn.

Dynamics of the VAR(p)

- Rewrite the VAR(p) model as:

\[ (I - \beta_1 L - \beta_2 L^2 - \beta_3 L^3 - \ldots - \beta_p L^p) y_t = \beta_0 + v_t \]

- Factor the lag polynomial:

\[ (I - B_1 L - B_2 L^2 - B_3 L^3 - \ldots - B_p L^p) \]
\[ = (I - \Phi_1 L)(I - \Phi_2 L)\ldots(I - \Phi_p L) \]

where \( \Phi_j \) are a nxn matrices. n is the dimension of \( y \)

(2 in the trades and quotes example)
• As long as every matrix $\Phi$ has eigenvalues inside the unit circle, the VAR(p) model is weakly stationary.

• We can re-write the VAR(p) model as an infinite order vector MA model as in:

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j y_{t-j} \quad \text{where} \quad \mu = (I - \beta_1 - \beta_2 \ldots \beta_p)^{-1}$$

Impulse Response Functions for VAR(p) models

• We still must take a stand on the order the shocks propagate...

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j P u_{t-j}$$
How do we estimate VAR’s?

- Estimation is straight forward.
- First, pick $p$. Can be the same for all variables in $y$ or we can have a different $p$ for each variable. Usually stat packages use the same value of $p$ for all equations.
- Estimate the model equation by equation using OLS just as we did for the univariate models.

Standard errors

- For a stationary VAR($p$) model, under the usual regularity conditions (see Proposition 11.1 Hamilton).
- Let $\Pi’ = [\beta_0, \beta_1, \ldots, \beta_p]$ and $\pi_{n(p+1) \times 1} = \text{vec} \left( \Pi_{np+1 \times n} \right)$
- Then $\sqrt{T} \left( \hat{\pi} - \pi \right) \sim N \left( 0, \Omega \otimes Q^{-1} \right)$

where $\otimes$ is the Kronecker operator
- Where $Q = E(x_i x_i’)$ and $x_i’ = [1, y’_{t-1}, \ldots, y’_{t-p}]$
• If $\pi_i$ are the parameters of the model for $y_i$ then we get the usual least squares result:

$$\sqrt{T} (\hat{\pi}_i - \pi_i) \approx N \left( 0, \hat{\sigma}_{i,i} \left[ \sum_{t=1}^{T} x_t x'_t \right]^{-1} \right)$$

• Where

$$\hat{\sigma}_{i,i} = \frac{1}{T - np - 1} \sum_{t=1}^{T} e_{it}$$

**Impulse response standard errors**

• The impulse responses are read off the elements of the $MA(\infty)$ representation.

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j v_{t-j}$$

• The elements of the $\psi_j$ are just (nonlinear) functions of the parameters in $\pi$. 
• The delta method (proposition 7.4 Hamilton) tells us

\[
\sqrt{T} \text{vec}(\hat{\psi}'_j - \psi'_j) \approx N \left( 0, \hat{G}_j \left( \hat{\Omega} \otimes \hat{\Omega}^{-1} \right) \hat{G}'_j \right)
\]

where

\[
\hat{G}_j = \frac{d \text{vec}(\psi'_j(\pi))}{d \pi} \bigg| \pi = \hat{\pi}
\]

\[
\hat{\Omega} = \frac{\sum_{j=1}^{n^2 \times n^2} \left( \hat{G}_j \left( \hat{\Omega} \otimes \hat{\Omega}^{-1} \right) \hat{G}'_j \right)}{n^2 \times n^2}
\]

VAR(p) example for trades and prices

• Step one estimate model for transaction prices using lagged values of transaction prices and lagged trades.

• Step two estimate model for trades using lagged values of transaction prices and lagged trades.
Covariance matrix of errors is estimated by:

- We have the estimated model:

\[ r_t = b_0 + \sum_{j=1}^{p} b_j r_{t-j} + \sum_{j=1}^{p} b_j^* x_{t-j} + e_t^r \]

\[ x_t = c_0 + \sum_{j=1}^{p} c_j r_{t-j} + \sum_{j=1}^{p} c_j^* x_{t-j} + e_t^x \]

\[ \hat{\Omega} = \frac{1}{T - 2p - 1} \sum_{t=1}^{T} e_t e_t^r \]

Where \( b \) and \( c \) correspond to estimated parameters.

Does the model fit the data?

- If the model is well specified, the residuals should be uncorrelated.
- We can examine the residuals of each equation and check if they are uncorrelated with their own past.
- We can also check to see that they are uncorrelated with the residuals of the other equations.
• Examine plots of sample correlations of the residual vector:

$$\hat{\Gamma}_l = \frac{1}{T-np-1} \sum_{t=1}^{T} e_t e_{t-l}.$$ 

Note that this contains "cross correlations".

• Critical value is given by

$$\frac{2}{\sqrt{T-np-1}}$$

---

**Forecasting VAR’s**

• Let $E_t(y_{t+k})$ denote the k-step ahead forecast of $y_{t+k}$ i.e.

• Let $\beta_j$ denote the estimate of the $j^{th}$ matrix and $\beta_0$ denote the estimate of the intercept vector.

• The one-step ahead forecast is given by:

$$E_t(y_{t+1}) = \beta_0 + \sum_{j=1}^{p} \beta_j E_t(y_{t+1-j})$$

$$E_t(y_{t+1}) = \beta_0 + \sum_{j=1}^{p} \beta_j y_{t+1-j}$$
• The 2-step ahead forecast is given by:

\[ E_t(y_{t+2}) = \beta_0 + \sum_{j=1}^{p} \beta_j E_t(y_{t+2-j}) \]

\[ E_t(y_{t+2}) = \beta_0 + \beta_1 E_t(y_{t+1}) + \sum_{j=2}^{p} \beta_j y_{t+2-j} \]

• The k-step ahead forecast is given by:

\[ E_t(y_{t+k}) = \beta_0 + \sum_{j=1}^{p} \beta_j E_t(y_{t+k-j}) \]

\[ E_t(y_{t+k}) = \beta_0 + \sum_{j=1}^{k-1} \beta_j E_t(y_{t+k-j}) + \sum_{j=k}^{p} \beta_j y_{t+k-j} \]

---

**Granger Causality**

• Causality is difficult to define. Granger Causality is one way of rigorously defining causality in a statistical setting.

• Built on two assumptions:
  - The future cannot cause the past.
  - A causes B if A has unique information that explains B.
• From a general point of view then, let $F_t$ denote the universe ALL information available at time $t$.
• Let $F_t-Z_t$ denote all information at time $t$ except $Z_t$.
• Then $Z$ fails to cause $Y$ if for all $s>0$:
  $$\Pr(Y_{t+s} \mid F_t) = \Pr(Y_{t+s} \mid F_t - Z_t).$$

• To make this feasible to implement, we must define the universe in a way we can handle. To this end we usually specify a vector of information.
• We evaluate differences in forecasts using a mean squared error criteria.
• Consider the bivariate system of $x_t$ and $y_t$.

• We say that $x_t$ fails to Granger Cause $y_t$ if:

\[ MSE \left[ E \left( y_t \mid y_{t-1}, y_{t-2}, \ldots \right) \right] = MSE \left[ E \left( y_t \mid y_{t-1}, y_{t-2}, \ldots, x_{t-1}, x_{t-2}, \ldots \right) \right] \]

• A natural \textit{feasible} test is therefore to estimate two models for $y_t$, one that includes lagged $x$ and $y$ and one that only includes lagged $y$. Then compare the MSE of the two models. This is the usual GC test.

---

**GC procedure**

• Restricted model: $y_i = \beta_0 + \sum_{j=1}^{p} \beta_j y_{i-j} + \varepsilon_i$

• Unrestricted model $y_i = \beta_0 + \sum_{j=1}^{p} \beta_j y_{i-j} + \sum_{j=1}^{q} \gamma_j x_{i-j} + \eta_i$

• Let $RSS_0 = \sum_{i=1}^{T} \hat{\varepsilon}_i^2$ and $RSS_i = \sum_{i=1}^{T} \hat{\eta}_i^2$

• Then the test stat is: $S = \frac{T(RSS_0 - RSS_i)}{RSS_i} \text{ asy.}\chi^2(p)$