Multivariate Time Series Analysis

Econometrics

Notes draw heavily from the following textbooks: Time Series Analysis (Hamilton, 1994), Econometric Analysis (Greene, 7th edition), and Analysis of Financial Time Series (Tsay, 3rd edition)

F-test review

- Suppose we want to run the regression

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \epsilon_t$$

- and test the joint hypothesis
  - $H_0 : \beta_2 = 4,\ \beta_3 = -2$
  - $H_a : \beta_2 \neq 4,\ \beta_3 \neq -2$
- One idea: variance-weighted squared distance of $(\widehat{\beta}_2, \widehat{\beta}_3)$ from (4,-2)
- Another idea: compare the variance of ϵ_t when $(\widehat{\beta}_2, \widehat{\beta}_3)$ are forced to be (4,-2) vs. when they are freely estimated.
- These turn out to be the same thing: an F-test.

F-test review

Some board work.

Forecasting with one equation

- Suppose we want to forecast y_{1t} based on
  - $(y_{1,t-1}, ..., y_{1,t-p})$
  - $(y_{2,t-1}, ..., y_{2,t-p})$
  - ...
  - $(y_{n,t-1}, ..., y_{n,t-p})$
  - and maybe some deterministic functions of the trend (time, time-squared, a sinusoid in time, seasonal dummies, etc.)
- Let $\mathbf{y_t} = (y_{1t}, y_{2t}, ..., y_{nt})'$ be an (n x 1) vector
  - $\mathbf{x_t} = (1, \mathbf{y'_{t-1}}, ..., \mathbf{y'_{t-p}})'$ is a (k x 1) vector, k=np+1
  - $\mathbf{y_t}, \mathbf{x_t}$ are covariance stationary

Forecasting with one equation

- Our forecast is $\hat{y}_{1t|t-1} = \beta'_{OLS}\mathbf{x_t}$
- For example: natural gas prices, oil prices, and drilling activity. Trying to forecast gas price returns.
  - y_{1t} =natural gas price return
  - y_{2t} =oil prices return
  - y_{3t} =oil & gas drilling activity

$$
\begin{aligned}
\hat{y}_{1t|t-1} \quad = \quad & \beta_0 + \beta_{1,ng}y_{1,t-1} + \ldots + \beta_{p,ng}y_{1,t-p} \\
& + \beta_{1,oil}y_{2,t-1} + \ldots + \beta_{p,oil}y_{2,t-p} + \beta_{1,dr}y_{3,t-1} + \ldots + \beta_{p,dr}y_{3,t-p}
\end{aligned}
$$

Forecasting with one equation

- How to determine lag length
  - AIC as before
  - "Test down the model": Successive F-tests of the p-th lag
  - Ho: coefficients on $y_{1,t-p}, \ldots, y_{n,t-p}$ are all jointly zero
- How to deal with MA terms
  - We will ignore them for now
  - We could model them explicitly
  - Tend to be captured by lags of y_2, y_3, etc.

Forecasting with one equation

R Example: lag length in Henry Hub, WTI, drilling equation.

Forecasting with one equation

- Granger causality
  - Ho: coefficients on $y_{2,t-1}, ..., y_{2,t-p}$ are all zero (Granger causality of y_2)
  - We would say that oil price changes “Granger cause” natural gas price changes
  - This does not necessarily mean that oil price changes **cause** natural gas price changes. It means they have *forecasting information* about gas price changes.
  - In some cases, Granger causality can be the opposite of true causality.
    - Stock returns Granger-cause (predict) GDP growth, but *are caused by it*.
  - Likely no variable would Granger-cause oil prices, for example, but oil prices Granger-cause (and actually cause) many things.

Forecasting with one equation

R Example: Granger causality of WTI price and drilling activity on Henry Hub price.

Forecasting with one equation

- Structural stability
- Chow test:
  1. Estimate model over entire sample, save residual sum of squares RSS_0
  2. Estimate model with different coefficients before and after a date of suspected break t_1:
     2.1 $y_{1t} = \mathbf{x_t}'\beta_\mathbf{1}(\delta_{[t \leq t_1]}) + \mathbf{x_t}'\beta_\mathbf{2}(\delta_{[t > t_1]}) + \epsilon_t$, save RSS_1
  3. Calculate $F(t_1) = \frac{(T-2k)(RSS_0 - RSS_1)}{kRSS_1}$, compare to $F(k, T-2k)$ critical value, null is that $\beta_1 = \beta_2$ (stability)

Forecasting with one equation

- Andrews 1993 test:
  - Do Chow test for all t_1 between first 15% and last 15% of the sample.
  - Compare **largest** $kF(t_1)$ to critical value in Table 1 of Andrews (1993)
    - (nonstandard "Andrews Distribution" when comparing multiple dates - the largest draw from a sequence of Chi-squareds. Like an extreme value distribution).
  - Bai & Perron (1998) extend to multiple break points, select optimal number and timing of breaks.
    - R package "strucchange".

Forecasting with one equation

R Example: structural break in the Brent oil price and LNG Asia price relationship

Vector Autoregression

- Suppose we are interested in more than just y_{1t}, or in potential feedbacks from y_{1t} through y_{2t} and y_{3t}.
- We have been estimating $y_{1t} = \pi_1' \mathbf{x_t} + \epsilon_{1t}$
  - $\mathbf{x_t} = (1, \mathbf{y}'_{\mathbf{t-1}}, ..., \mathbf{y}'_{\mathbf{t-p}})'$ is a (k x 1) vector, k=np+1
  - Here ϵ_{1t} is the forecast error for variable y_{1t}.
- We could run the analogous regression for $y_{2t} = \pi_2' \mathbf{x_t} + \epsilon_{2t}$ and so on.

Vector Autoregression

- Stack them in a vector system:
- $\begin{bmatrix} y_{1t} \\ \vdots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} \pi_{\mathbf{1}}' \\ \vdots \\ \pi_{\mathbf{n}}' \end{bmatrix} \mathbf{x_t} + \begin{bmatrix} \epsilon_{1t} \\ \vdots \\ \epsilon_{nt} \end{bmatrix}$
- $\mathbf{y_t} = \Pi' \mathbf{x_t} + \epsilon_{\mathbf{t}}$ where Π' is (n x k)
- $\Pi' \mathbf{x_t} = \begin{bmatrix} c & \Phi_1 & \ldots & \Phi_p \end{bmatrix} \begin{bmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix}$
- $\mathbf{y_t} = \mathbf{c} + \Phi_1 \mathbf{y_{t-1}} + \ldots + \Phi_p \mathbf{y_{t-p}} + \epsilon_{\mathbf{t}}$ is called a vector autoregression (VAR).

Vector Autoregression

- Example: Henry Hub, WTI, drilling
- $$\begin{bmatrix} y_{ng,t} \\ y_{oil,t} \\ y_{dr,t} \end{bmatrix} = \begin{bmatrix} \beta_0^{ng} & \beta_{1,ng}^{ng} & \cdots & \beta_{3,dr}^{ng} \\ \beta_0^{oil} & \beta_{1,ng}^{oil} & \cdots & \beta_{3,dr}^{oil} \\ \beta_0^{dr} & \beta_{1,ng}^{dr} & \cdots & \beta_{3,dr}^{dr} \end{bmatrix} \begin{bmatrix} 1 \\ y_{ng,t-1} \\ \vdots \\ y_{dr,t-3} \end{bmatrix} + \begin{bmatrix} \epsilon_{ng,t} \\ \epsilon_{oil,t} \\ \epsilon_{dr,t} \end{bmatrix}$$
- $$\begin{bmatrix} y_{ng,t} \\ y_{oil,t} \\ y_{dr,t} \end{bmatrix} = \begin{bmatrix} \beta_0^{ng} \\ \beta_0^{oil} \\ \beta_0^{dr} \end{bmatrix} + \begin{bmatrix} \beta_{1,ng}^{ng} & \beta_{1,oil}^{ng} & \beta_{1,dr}^{ng} \\ \beta_{1,ng}^{oil} & \beta_{1,oil}^{oil} & \beta_{1,dr}^{oil} \\ \beta_{1,ng}^{dr} & \beta_{1,oil}^{dr} & \beta_{1,dr}^{dr} \end{bmatrix} \begin{bmatrix} y_{ng,t-1} \\ y_{oil,t-1} \\ y_{dr,t-1} \end{bmatrix}$$
$$+ \ldots + \begin{bmatrix} \beta_{3,ng}^{ng} & \beta_{3,oil}^{ng} & \beta_{3,dr}^{ng} \\ \beta_{3,ng}^{oil} & \beta_{3,oil}^{oil} & \beta_{3,dr}^{oil} \\ \beta_{3,ng}^{dr} & \beta_{3,oil}^{dr} & \beta_{3,dr}^{dr} \end{bmatrix} \begin{bmatrix} y_{ng,t-3} \\ y_{oil,t-3} \\ y_{dr,t-3} \end{bmatrix} + \begin{bmatrix} \epsilon_{ng,t} \\ \epsilon_{oil,t} \\ \epsilon_{dr,t} \end{bmatrix}$$

Vector Autoregression

- We can estimate this with conditional MLE like we learned before (conditional on the first p observations).
- Assume $\epsilon_{\mathbf{t}} \sim N(\mathbf{0}, \Omega)$
- Treat the sample size as if its $T + p$ observations
- $logL = -\frac{Tn}{2} log(2\pi) - \frac{T}{2} log|\Omega| - \frac{1}{2} \sum(\mathbf{y_t} - \Pi'\mathbf{x_t})'\Omega^{-1}(\mathbf{y_t} - \Pi'\mathbf{x_t})$
- Some really nice things happen when you maximize this:
- The i^{th} row of the MLE of Π' is the OLS estimate of the coefficients of the i^{th} equation.
  - We could estimate OLS equation by equation and get the exact same answers.

Vector Autoregression

- The residuals from the i^{th} row/equation are the OLS residuals from that equation, even if the system is estimated by MLE.
  - The MLE estimate of the variance Ω is equivalent to those OLS residuals in a straightforward way:
  - $\hat{\Omega} = \frac{1}{T} \sum \hat{\epsilon}_t \hat{\epsilon}_t'$

Vector Autoregression

- However testing joint hypotheses across equations is much easier using likelihood ratio test with Ω than F-tests.
  - The maximized value of the log likelihood is $-\frac{Tn}{2}(1 + log2\pi) - \frac{T}{2}log|\hat{\Omega}|$
  - Now if we want to test lag length for every equation in the whole system (p versus p-1 lags of each variable in each equation, in both cases using the last T of p+T observations because we burned the first p observations already), we have two estimates of Ω for two different lag lengths:
  - $\hat{\Omega}(p-1)$ and $\hat{\Omega}(p)$ that can be constructed from the OLS residuals from the two regressions with p and p-1 lags.
  - The likelihood ratio test statistic (twice the log likelihood ratio) is:
  - $T\left[log|\hat{\Omega}(p-1)| - log|\hat{\Omega}(p)|\right] \sim \chi^2(n^2)$
  - the small sample correction multiplies this by $T-k$, not T.

Vector Autoregression

- Other criteria for lag length selection include
  - Akaike information criterion (AIC): choose the specification with smallest AIC
    - $log|\hat{\Omega}(p)| + 2p\frac{n^2}{T}$ penalizes the number of lags and equations per observation, rewards sample size
  - Schwarz or Schwarz-Bayes Criterion
    - $log|\hat{\Omega}(p)| + \left(\frac{pn^2}{T}\right) logT$, larger penalty on larger numbers of parameters relative to sample size.
  - Rules of thumb for macro data:
    - $p \geq 4$ for quarterly data
    - use lags 1-6 and 11-13 for monthly data

Impulse Response Functions, Forecasts

- We can rewrite the whole VAR $\underset{nX1}{\mathbf{y_t}} = \mathbf{c} + \Phi_1 \mathbf{y_{t-1}} + \ldots + \Phi_p \mathbf{y_{t-p}} + \epsilon_\mathbf{t}$ in companion form:
- $$\underset{(npX1)}{\begin{bmatrix} \mathbf{y_t} \\ \mathbf{y_{t-1}} \\ \vdots \\ \mathbf{y_{t-p+1}} \end{bmatrix}} = \begin{bmatrix} \mathbf{c} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} + \underset{(npXnp)}{\begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_p \\ \mathbf{I_n} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & & \vdots \\ \mathbf{0} & \cdots & \mathbf{I_n} & \mathbf{0} \end{bmatrix}} \begin{bmatrix} \mathbf{y_{t-1}} \\ \mathbf{y_{t-2}} \\ \vdots \\ \mathbf{y_{t-p}} \end{bmatrix} + \begin{bmatrix} \epsilon_\mathbf{t} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

Impulse Response Functions, Forecasts

- $\xi_t = \gamma + \mathbf{F}\xi_{t-1} + \mathbf{v}_t$
- $\xi_{t+2} = \gamma + \mathbf{F}\xi_{t+1} + \mathbf{v}_{t+2}$
- $\xi_{t+2} = (\mathbf{I}_{np} + \mathbf{F})\gamma + \mathbf{F}^2\xi_t + \mathbf{v}_{t+2} + \mathbf{F}\mathbf{v}_{t+1}$
- $$\xi_{t+s} = (\mathbf{I}_{np} + \mathbf{F} + \mathbf{F}^2 + \cdots + \mathbf{F}^{s-1})\gamma + \mathbf{F}^s\xi_t + \mathbf{v}_{t+s} + \mathbf{F}\mathbf{v}_{t+s-1} + \cdots + \mathbf{F}^{s-1}\mathbf{v}_{t+1}$$
- $\frac{\partial \xi_{t+s}}{\partial \xi_t'} = \mathbf{F}^s$

Impulse Response Functions, Forecasts

- We are really interested in the first n rows of $\xi_{\mathbf{t+s}}$: $\frac{\partial \mathbf{y_{t+s}}}{\partial \mathbf{y_t'}}$
- $\mathbf{y_{t+s}} = \mathbf{c_s} + \epsilon_{\mathbf{t+s}} + \Psi_1 \epsilon_{\mathbf{t+s-1}} + \cdots + \Psi_{s-1} \epsilon_{\mathbf{t+1}} + \Psi_s \mathbf{y_t} + \mathbf{F_{12}^{(s)}} \mathbf{y_{t-1}} + \cdots + \mathbf{F_{1p}^{(s)}} \mathbf{y_{t-p+1}}$
- $\Psi_s = \mathbf{F_{11}^{(s)}} = \frac{\partial \mathbf{y_{t+s}}}{\partial \mathbf{y_t'}} = \frac{\partial \mathbf{y_{t+s}}}{\partial \epsilon_{\mathbf{t}}'}$
- Plotting Ψ_s as a function of s is called the "non-orthogonalized impulse response function."

Impulse Response Functions, Forecasts

- Why "non-orthogonalized"? Errors across equations in the system are correlated.
  - $Cov(\epsilon_{it}\epsilon_{jt}) \neq 0$ in general.
  - Therefore a shock to ϵ_{it} contains information about all the other ϵ_{jt}'s, and so is not orthogonal to them in reality.
  - When we calculate the simple derivative, however, we hold the other ϵ_{jt}'s constant even though they are not.
  - We will learn another method to account for the additional information shortly.
- The row is the outcome being affected, and the column is the shock that's changing

Impulse Response Functions & Forecasts

- Row i, column j of Ψ_s gives $\frac{\partial y_{i,t+s}}{\partial y_{jt}} = \frac{\partial y_{i,t+s}}{\partial \epsilon_{jt}}$ holding constant ϵ_{kt} $(k \neq j)$ and ϵ_{t+m}.
  - (even though ϵ_{kt} $(k \neq j)$ not likely constant because of correlated shocks across equations).
- Notice that if our VAR(p) is stationary, it can be written as a VMA(∞):
  - $\mathbf{y_{t+s}} = \mu + \epsilon_{\mathbf{t+s}} + \Psi_1 \epsilon_{\mathbf{t+s-1}} + \Psi_2 \epsilon_{\mathbf{t+s-2}} + \ldots$
  - The Ψ's have to be absolutely summable, so $\Psi_s \rightarrow \mathbf{0}$
  - Plots of these should have a decay
  - If not stationary, impulse response functions will have permanent effects.

Orthogonalized Impulse Response Functions

- If we are using impulse response functions to forecast, then a shock of ϵ_{1t} has information about contemporaneous shocks $\epsilon_{2t}, \ldots \epsilon_{nt}$
- We might want to take account of this extra information if we learn ϵ_{1t} but have not yet learned the value of the other shocks.
- Another way to ask the question:
  - Suppose we have only period $t-1$ information $(\mathbf{y_{t-1}}, \mathbf{y_{t-2}}, \ldots)$.
  - Then suddenly we learn the value of one variable, say y_{1t}.
  - How does learning only y_{1t} change our forecast of $\mathbf{y_{t+s}}$?

Orthogonalized Impulse Response Functions

- Our VAR says:
- $\mathbf{y_{t+s}} = \mathbf{c_s} + \epsilon_{\mathbf{t+s}} + \Psi_1 \epsilon_{\mathbf{t+s-1}} + \cdots + \Psi_{s-1} \epsilon_{\mathbf{t+1}} + \Psi_s \mathbf{y_t} + \mathbf{F_{12}^{(s)}} \mathbf{y_{t-1}} + \cdots + \mathbf{F_{1p}^{(s)}} \mathbf{y_{t-p+1}}$
- For forecasting, $E(\epsilon_{\mathbf{t+k}}) = \mathbf{0}$, linear projection of $\mathbf{y_{t+s}}$ on period t information is
- $\hat{\mathbf{y}}_{\mathbf{t+s|t}} = \mathbf{c_s} + \Psi_s \mathbf{y_t} + \mathbf{F_{12}^{(s)}} \mathbf{y_{t-1}} + \cdots + \mathbf{F_{1p}^{(s)}} \mathbf{y_{t-p+1}}$
- Row i of this is our forecast of $y_{i,t+s}$ based on period t information.

Orthogonalized Impulse Response Functions

- What happens if we only knew period $t-1$ information and then learned ϵ_{1t},
  - ϵ_{1t} is correlated with $\epsilon_{2t}, \ldots, \epsilon_{nt}$ through the first column of Ω.
- Suppose our linear projection of $\mathbf{y_{t+s}}$ is based only on knowing ϵ_{1t} and all the $t-1$ information:
- $\hat{E}(\mathbf{y_{t+s}}|\epsilon_{1t}, \mathbf{y_{t-1}}, \mathbf{y_{t-2}}, \ldots) =$
  $\mathbf{c_s} + \Psi_s \hat{E}(\epsilon_\mathbf{t}|\epsilon_{1t}, \mathbf{y_{t-1}}, \mathbf{y_{t-2}}, \ldots) + \mathbf{F_{12}^{(s)}} \mathbf{y_{t-1}} + \cdots + \mathbf{F_{1p}^{(s)}} \mathbf{y_{t-p+1}}$
  - We use ϵ_{1t} to guess the rest of the period t information.

Orthogonalized Impulse Response Functions

- Let $\mathbf{a_1}\epsilon_{1t} = \hat{E}(\epsilon_\mathbf{t}|\epsilon_{1t})$
- $\hat{E}(\mathbf{y_{t+s}}|\epsilon_{1t}, \mathbf{y_{t-1}}, \mathbf{y_{t-2}}, \ldots) = \mathbf{c_s} + \Psi_s \mathbf{a_1}\epsilon_{1t} + \mathbf{F_{12}^{(s)}}\mathbf{y_{t-1}} + \cdots + \mathbf{F_{1p}^{(s)}}\mathbf{y_{t-p+1}}$
- So the effect of y_{1t} on our forecast of $\mathbf{y_{t+s}}$ if y_{1t} is all we know at period t is
  - $\frac{\partial \hat{E}(\mathbf{y_{t+s}}|\epsilon_{1t}, \mathbf{y_{t-1}}, \mathbf{y_{t-2}}, \ldots)}{\partial y_{1t}} = \Psi_s \mathbf{a_1}$
  - $\mathbf{a_1} = \frac{\partial \hat{E}(\epsilon_\mathbf{t}|\epsilon_{1t})}{\partial \epsilon_{1t}}$
- How do we estimate this?

Orthogonalized Impulse Response Functions

- Notice that $\mathbf{a_1}$ accounts for how ϵ_{1t} is correlated with $\epsilon_{2t}, \ldots, \epsilon_{nt}$ through the first column of Ω.
- Recall $\hat{\Omega} = \frac{1}{T} \sum \hat{\epsilon}_\mathbf{t} \hat{\epsilon}_\mathbf{t}'$
- We can factor this as $\hat{\Omega} = \hat{A}\hat{D}\hat{A}'$
- $\hat{A}$ is a lower triangular matrix with ones on the principle diagonal, and $\hat{D}$ is diagonal with all elements positive.
- $\hat{\mathbf{a}}_\mathbf{1}$ is the first column of $\hat{A}$

Orthogonalized Impulse Response Functions

- Alternatively, we could take the Cholesky factorization of $\hat{\Omega}$:
  - $\hat{\Omega} = \hat{P}\hat{P}' = \hat{A}\hat{D}\hat{A}' = \hat{A}\hat{D}^{\frac{1}{2}}\hat{D}^{\frac{1}{2}}\hat{A}'$
  - $\hat{P}$ is lower triangular with positive elements along principle diagonal.
  - Let $\hat{D}^{\frac{1}{2}}$ be the diagonal matrix whose diagonal is the diagonal of $\hat{P}$ (and all other elements of $\hat{D}$ are zero).
  - Then $\hat{A} = \hat{P}\hat{D}^{-\frac{1}{2}}$

Orthogonalized Impulse Response Functions

- The orthogonalized impulse response function is $\hat{\Psi}_s\hat{\mathbf{a}}_1$.
  - Many researchers simply report $\hat{\Psi}_s\hat{\mathbf{p}}_1$ where $\hat{\mathbf{p}}_1$ is the first column of $\hat{P}$
  - This only differs by the scale.
  - $\hat{\Psi}_s\hat{\mathbf{a}}_1$ is the effect of a one-unit increase in y_{1t}
  - $\hat{\Psi}_s\hat{\mathbf{p}}_1$ is the effect of a one-standard deviation increase in y_{1t}.

Orthogonalized Impulse Response Functions

- Now what happens once I've learned y_{1t}, if I then learn y_{2t} but not y_{3t}, etc.?
  - $\hat{\Psi}_s \hat{\mathbf{a}}_2$ where $\hat{\mathbf{a}}_2$ is the second column of $\hat{A}$.
  - or use $\hat{\Psi}_s \hat{\mathbf{p}}_2$ in terms of standard deviations.
- More generally, these are called "recursively orthogonalized impulse response functions"
  - $\hat{\blacksquare}_{\mathbf{s}} \hat{\mathbf{A}}$
- The value of this is different depending on the ordering of the equations.
  - How you order them depends on what question you want to answer.
  - Typically order them in the order in which you think information would be made known.

Forecast Error Variance Decomposition

- How much of the variation in our forecast is coming from a particular variable?
- How much of the variance in our *s*-period ahead coal export forecast is coming from variance in the natural gas price?
- Define $\mathbf{u_t} = \mathbf{A}^{-1}\epsilon_\mathbf{t}$.
- $E(\mathbf{u_t}\mathbf{u_t'}) = \mathbf{A}^{-1}\Omega(\mathbf{A}^{-1})' = \mathbf{A}^{-1}\mathbf{ADA}'(\mathbf{A}^{-1})' = \mathbf{D}$
  - The variance covariance matrix of $\mathbf{u_t}$ is diagonal.
  - $\mathbf{u_t}$ are called "orthogonalized innovations"

Forecast Error Variance Decomposition

- Interpretation:
  - $\epsilon_{\mathbf{t}} = \mathbf{y_t} - \hat{\mathbf{y}}_{\mathbf{t|t-1}}$
  - $u_{1t} = \epsilon_{1t} = y_{1t} - \hat{E}(y_{1t}|\mathbf{y_{t-1}}, \mathbf{y_{t-2}}, \ldots)$
  - $u_{jt} = y_{jt} - \hat{E}(y_{jt}|y_{j-1,t}, y_{j-2,t}, \ldots, y_{1t}, \mathbf{y_{t-1}}, \mathbf{y_{t-2}} \ldots)$

Forecast Error Variance Decomposition

- s-period-ahead forecast error:
  - $\mathbf{y_{t+s}} - \mathbf{\hat{y}_{t+s|t}} = \epsilon_{\mathbf{t+s}} + \Psi_1 \epsilon_{\mathbf{t+s-1}} + \cdots + \Psi_{s-1} \epsilon_{\mathbf{t+1}}$
  - $= \mathbf{Au_{t+s}} + \Psi_1 \mathbf{Au_{t+s-1}} + \cdots + \Psi_{s-1} \mathbf{Au_{t+1}}$
  - We can break this down into the contribution from each u_{jt}:
  - $\mathbf{Au_{t+m}} = \mathbf{a_1} u_{1,t+m} + \mathbf{a_2} u_{2,t+m} + \cdots + \mathbf{a_n} u_{n,t+m}$

Forecast Error Variance Decomposition

- s-period-ahead mean squared error:
  - $E(\mathbf{y_{t+s}} - \hat{\mathbf{y}}_{\mathbf{t+s|t}})(\mathbf{y_{t+s}} - \hat{\mathbf{y}}_{\mathbf{t+s|t}})' =$
  - $= \Omega + \Psi_1 \Omega \Psi_1' + \Psi_2 \Omega \Psi_2' + \cdots + \Psi_{s-1} \Omega \Psi_{s-1}'$
  - $= \mathbf{ADA} + \Psi_1 \mathbf{ADA} \Psi_1' + \Psi_2 \mathbf{ADA} \Psi_2' + \cdots + \Psi_{s-1} \mathbf{ADA} \Psi_{s-1}'$
  - $= \sum_{m=0}^{s-1} \Psi_m (\mathbf{a_1 a_1'} d_1 + \cdots + \mathbf{a_n a_n'} d_n) \Psi_m'$
  - The k'th term in paranthesis represents how much of our mean squared forecast error comes from not knowing $u_{k,t+1}, u_{k,t+2}, \ldots, u_{k,t+s}$:
  - $\sum_{m=0}^{s-1} \Psi_m (\mathbf{a_k a_k'} d_k) \Psi_m'$
  - You can plot this as a function of s, similar to an impulse response function.
  - Sometimes called “fevd” (which is the Stata subcommand for it) for “forecast error variance decomposition”.

Structural VAR

- VARs we have looked at so far have been "reduced form": each equation in the system is a function of lags only: of its own value, and lags of other variables
- Maybe we can use/test structural information about contemporaneous relationships.
- This increases the number of parameters we need to estimate - only okay if the number of parameters is not too large to be recovered from reduced form estimates.
  - If the number of parameters is too large, we may impose some structural parameters in order to estimate others.

Structural VAR

- Consider a macroeconomic relationship
  - y_{1t} =real GDP growth
  - y_{2t} =inflation
  - y_{3t} =fed funds rate
  - y_{4t} =rate of growth of M2

Structural VAR

- Current spending (GDP growth) depends only on past shocks
  - $y_{1t} = k_1 + B_1^{(1,.)}\mathbf{y_{t-1}} + \cdots + u_{1t}$
- Inflation has a Phillips curve relation between spending and inflation:
  - $y_{2t} = k_2 + b_0^{(2,1)}y_{1t} + B_1^{(2,.)}\mathbf{y_{t-1}} + \cdots + u_{2t}$
- Federal reserve responds to current output growth and inflation:
  - $y_{3t} = k_3 + b_0^{(3,1)}y_{1t} + b_0^{(3,2)}y_{2t} + B_1^{(3,.)}\mathbf{y_{t-1}} + \cdots + u_{3t}$
- Money demand depends on current output, inflation, and interest rates:
  - $y_{3t} = k_3 + b_0^{(4,1)}y_{1t} + b_0^{(4,2)}y_{2t} + b_0^{(4,3)}y_{3t} + B_1^{(3,.)}\mathbf{y_{t-1}} + \cdots + u_{4t}$

Structural VAR

- We can stack these equations into a vector dynamic structural model:
  - $\mathbf{B_0 y_t} = \mathbf{k} + \mathbf{B_1 y_{t-1}} + \cdots + \mathbf{B_p y_{t-p}} + \mathbf{u_t}$
  - Any dynamic structural model that has a linear approximation can be written in this way.
  - $\mathbf{B_0} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ b_0^{(2,1)} & 1 & 0 & 0 \\ b_0^{(3,1)} & b_0^{(3,2)} & 1 & 0 \\ b_0^{(4,1)} & b_0^{(4,2)} & b_0^{(4,3)} & 1 \end{bmatrix}$
  - The lower triangular form is a special case for this model

Structural VAR

- If we premultiply by $\mathbf{B_0^{-1}}$, we get the reduced form:
  - $\underset{nX1}{\mathbf{y_t}} = \mathbf{c} + \Phi_1 \mathbf{y_{t-1}} + \ldots + \Phi_p \mathbf{y_{t-p}} + \epsilon_\mathbf{t}$
  - $\mathbf{c} = \mathbf{B_0^{-1} k}$
  - $\Phi_j = \mathbf{B_0^{-1} B_j}$
  - $\epsilon_\mathbf{t} = \mathbf{B_0^{-1} u_t}$

Structural VAR

- The reduced form VAR is just another representation of a linearized dynamic structural model.
  - If we knew the structural parameters **B** and shocks **u** we could calculate the VAR parameters Φ and ϵ and the associated impulse response functions.
  - $\frac{\partial \mathbf{y_{t+m}}}{\partial \mathbf{u_t'}} = \frac{\partial \mathbf{y_{t+m}}}{\partial \epsilon_\mathbf{t}'} \frac{\partial \epsilon_\mathbf{t}}{\partial \mathbf{u_t'}} = \Psi_m \mathbf{B_0^{-1}}$

Structural VAR

- So can we estimate the reduced form and recover the structural parameters?
  - Reduced form VAR parameters: $n + n(n+1)/2 + n^2 p$
  - Structural VAR parameters: $n + n(n+1)/2 + n^2 p + n^2$
  - Need enough restrictions on $\mathbf{B_0}$ (e.g., zeros for some elements) and $E(\mathbf{u_t u_t'})$ (e.g., diagonal matrix) in order for their to be a one-to-one mapping.
  - The earlier lower triangular form, with diagonal $E(\mathbf{u_t u_t'})$, is one such example.

Structural VAR

- If there is a one-to-one mapping, then estimate reduced form and figure out the appropriate transformation of Φ and Ω
  - For example in the case above:
  - $\epsilon_t = \mathbf{B_0^{-1} u_t}$
  - $\Omega = \mathbf{B_0^{-1} D (B_0^{-1})'}$
  - Find triangular factorization of $\hat{\Omega}$ as before.
  - When the structural equation is lower triangular, this is the same as the recursively orthogonalized impulse response function.

Structural VAR

- For non-recursive structural models:
  - Let $E(\mathbf{u_t u_t'}) = \mathbf{D}$
  - Estimate reduced form by OLS, put enough restrictions on $\mathbf{D}, \mathbf{B_0}, \ldots, \mathbf{B_p}$ so that there is a unique mapping from $\Omega, \Phi_1, \ldots, \Phi_p$.
  - Typical approach: let $\mathbf{B_1}, \ldots, \mathbf{B_p}$ be unrestricted.
  - $E(\epsilon_{\mathbf{t}}\epsilon_{\mathbf{t}}') = \Omega = \mathbf{B_0^{-1} D (B_0^{-1})'}$
  - $\hat{\Omega}$ has $n(n+1)/2$ distinct elements. so we can have just that many unknowns in $\mathbf{B_0}$ and $\mathbf{D}$
  - Need to impose as many structural restrictions from theory, long run equilibrium, etc., to reduce parameters to a number that can be recovered from the reduced form (or estimated directly using MLE).

Scalar Case

- $y_t = \phi y_{t-1} + w_t$
- $y_{t+1} = \phi y_t + w_{t+1} = \phi^2 y_{t-1} + \phi w_t + w_{t+1}$
- $y_{t+m} = \phi^{m+1} y_{t-1} + \sum_{s=0}^{m} \phi^s w_{t+m-s}$
- If shock w_t is larger, but w_{t+j} is unchanged, what happens to y_{t+m} at date $t+m$?
- $\frac{\partial y_{t+m}}{\partial w_t} = \phi^m$

Scalar Case

- $\frac{\partial y_{t+m}}{\partial w_t} = \phi^m$
- Plot this derivative as a function of m, should be decaying in m if $\phi < 1$.
  - This is an “impulse response function”.
  - If $\phi < 0$ we have oscillation.
  - If $|\phi| > 1$, we have explosive unstable effect.
  - If $\phi = 1$, we have a permanent effect.

Vector Case

- $\mathbf{y_t}$ =vector of n variables observed at t (system of n equations).
- $\mathbf{y_t} = \mathbf{F y_{t-1}} + \mathbf{w_t}$, $\mathbf{F}$ is an nXn matrix of coefficients.
- $\mathbf{y_{t+1}} = \mathbf{F y_t} + \mathbf{w_{t+1}} = \mathbf{F^2 y_{t-1}} + \mathbf{F w_t} + \mathbf{w_{t+1}}$, where $\mathbf{F^2} = \mathbf{F} \cdot \mathbf{F}$
- If $\mathbf{F} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}$, then
  $\mathbf{F^2} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \phi_1^2 + \phi_2 & \phi_1 \phi_2 \\ \phi_1 & \phi_2 \end{bmatrix}$
- $\mathbf{y_{t+m}} = \mathbf{F^{m+1} y_{t-1}} + \mathbf{F^m w_t} + \mathbf{F^{m-1} w_{t+1}} + \ldots + \mathbf{F w_{t+m-1}} + \mathbf{w_{t+m}}$
  - $= \mathbf{F^{m+1} y_{t-1}} + \sum_{\mathbf{s=0}}^{\mathbf{m}} \mathbf{F^s w_{t+m-s}}$

Vector Case

- What happens if j^{th} component of $\mathbf{w_t}$ (w_{jt}) increases by 1 unit, with all other w fixed? What would happen to i^{th} component of $\mathbf{y_{t+m}}$ ($y_{i,t+m}$)?
- $\frac{\partial y_{i,t+m}}{\partial w_{jt}}$ = the row i, column j element of the nXn matrix $\frac{\partial \mathbf{y_{t+m}}}{\partial \mathbf{w_t'}} = \mathbf{F^m}$.
- For example, to find the effect of w_{1t} on $y_{1,t+3}$, find the (1,1) element of
  - $\mathbf{F^3} = \begin{bmatrix} \phi_1^2 + \phi_2 & \phi_1\phi_2 \\ \phi_1 & \phi_2 \end{bmatrix} \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}$
  - $\implies F^3_{1,1} = \phi_1^3 + 2\phi_1\phi_2$

Vector Case

- We might want to know how $\mathbf{F}^{\mathbf{m}}$ behaves as a function of $\mathbf{m}$ generally.
  - Depends on the eigenvalues of $\mathbf{F}$, or values of λ that make $|\mathbf{F}-\lambda\mathbf{I_n}| = 0$
  - or $\begin{vmatrix} f_{11}-\lambda & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nn}-\lambda \end{vmatrix} = 0$

Vector Case

- Calculating this determinant using cofactor expansion gives an n^{th}-order polynomial in λ
  - $a_0 + a_1\lambda + a_2\lambda^2 + ... + a_n\lambda^n = 0$
  - Suppose this polynomial has n distinct roots. Then there exists a nonsingular nXn matrix $\mathbf{T}$ such that we can diagonalize
    - $\mathbf{F} = \mathbf{T}\Lambda\mathbf{T}^{-1}$, where $\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \\ \vdots & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$
    - If there are not n distinct roots, we can use the Jordan normal form of $\mathbf{F}$ and get similar results.

Vector Case

- Now $\mathbf{F^2} = (\mathbf{T}\Lambda\mathbf{T}^{-\mathbf{1}})(\mathbf{T}\Lambda\mathbf{T}^{-\mathbf{1}}) = \mathbf{T}\Lambda^2\mathbf{T}^{-\mathbf{1}}$ or $\mathbf{F^m} = \mathbf{T}\Lambda^m\mathbf{T}^{-\mathbf{1}}$, where
  - $\blacksquare^{\mathbf{m}} = \begin{bmatrix} \lambda_1^m & \cdots & 0 \\ \vdots & \ddots & \\ 0 & & \lambda_n^m \end{bmatrix}$
  - We want to say something about the nature of the impulse response coefficients $\frac{\partial y_{i,t+m}}{\partial w_{jt}}$ as m grows.
    - This is now the i, j element of $\mathbf{T}\Lambda^m\mathbf{T}^{-\mathbf{1}}$, which because of $\mathbf{T}$ has the form $a_{ij}^{(1)}\lambda_1^m + ... + a_{ij}^{(n)}\lambda_n^m$
    - Knowing the eigenvalues of $\mathbf{F}$ and the elements of $\mathbf{T}$ and $\mathbf{T}^{-\mathbf{1}}$ tells us the impulse response.

Vector Case

- If the $\lambda' s$ are all real:
  - the impulse response function is a weighted average of the possible impulse response functions for the first-order scalar case.
  - If $|\lambda_s| < 1 \ \forall\ s$, then effects will die out over time (system is stable)
    - $lim_{m \to \infty} \frac{\partial y_{i,t+m}}{\partial w_{jt}} = 0 \ \forall\ i, j$
  - If $|\lambda_s| > 1$ for some s, then the system is explosive.
  - If $\lambda_n = 1$ and $|\lambda_s| < 1$ for $s = 1, ..., n-1$, then $\frac{\partial y_{i,t+m}}{\partial w_{jt}} \to a_{ij}^{(n)}$ and there is a permanent effect (even if w_n is not the shock).

Vector Case

- If the $\lambda' s$ are complex:
  - For example, $\mathbf{F} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}$
  - $\implies |\mathbf{F} - \lambda \mathbf{I}| = \begin{vmatrix} \phi_1 - \lambda & \phi_2 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - \phi_1 \lambda - \phi_2 = 0$
  - $\lambda = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$, so the roots can be complex conjugates if $\phi_1^2 + 4\phi_2 < 0$
  - $\lambda_1 = a + bi$, $\lambda_2 = a - bi$ for some a, b.

Vector Case

- If the $\lambda' s$ are complex:
  - $\frac{\partial \mathbf{y_{t+m}}}{\partial \mathbf{w_t'}} = \mathbf{F^m}$ is still real, but the matrices in the decomposition $\mathbf{T}\Lambda^m\mathbf{T}^{-1}$ has complex elements.
  - The $\sqrt{-1}$ will multiply and cancel out.
    - Complex numbers are only part of the intermediate analysis to get some insight.
    - Want to ask how does λ^m behave as a function of m for $\lambda = a + bi$

Vector Case, complex roots

- Recall that we can write $a + bi = R(cos\theta + i \cdot sin\theta) = Re^{i\theta}$, where the last equality comes from Euler's formula.
  - $R = \sqrt{a^2 + b^2}$
    - $\lambda^m = (a + bi)^m = R^m e^{i\theta m} = R^m (cos(\theta m) + i \cdot sin(\theta m))$
    - λ^m behaves like R^m multiplied by a sinusoidal function.
  - Magnitude of effect depends on size or $R = \sqrt{a^2 + b^2}$
    - $R < 1$ means effects die out over time (as m increases)
    - $R > 1$ means effects explode
    - $R = 1$ means there are permanent effects, and permanent oscillations in this case.

Summary

- stability of $\mathbf{y_t} = \mathbf{F}\mathbf{y_{t-1}} + \mathbf{w_t}$ depends on eigenvalues of $\mathbf{F}$.
  - If $\|\lambda_s\| < 1$ for all $s = 1, ..., n$ (or the modulus - absolute value when real, R when complex)
    - $|\lambda_s| < 1$when real, $R_s = \sqrt{a^2 + b^2} < 1$ when complex, or if all λ_s inside the unit circle
    - Then the system is **stable**.

2nd-order scalar difference equation

- $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t$
- Define $\mathbf{y_t} = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}$
- $\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \end{bmatrix}$
- $\mathbf{y_t} = \mathbf{F y_{t-1}} + \mathbf{w_t}$
- Call this the “companion form”.
- We have already analyzed this as a first order vector system.

p^{th}-order scalar difference equation

- $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-1} + ... + \phi_p y_{t-p} + w_t$
- Companion form:
- $\mathbf{y_t} = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \mathbf{F} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & & & 0 \\ 0 & 1 & & & 0 \\ \vdots & & \ddots & & \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \mathbf{w_t} = \begin{bmatrix} w_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$
- $\mathbf{y_t} = \mathbf{F}\mathbf{y_{t-1}} + \mathbf{w_t}$
- $\frac{\partial y_{1,t+m}}{\partial w_{1t}} = c_1\lambda_1^m + ... + c_p\lambda_p^m$ if all $\lambda's$ distinct.

p^{th}-order scalar difference equation

- Eigenvalues of **F**satisfy
  - $\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - ... - \phi_p = 0$
- effects die out if $\|\lambda_s\| < 1$ for all $s = 1, ..., p$
- effects explode if $\|\lambda_s\| > 1$ for some s
- effects are permanent of $\|\lambda_s\| = 1$ for some s.

Spurious Regression

- Suppose we have two independent random walks
  - $\Delta y_{1t} = \epsilon_{1t}, \ \Delta y_{2t} = \epsilon_{2t}$
  - $\epsilon_{\mathbf{t}} \sim iid(\mathbf{0}, \Omega)$ where $\Omega = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$

Spurious Regression

- Suppose we regress one random walk on the other
  - $y_{1t} = \alpha + \gamma y_{2t} + u_t$
  - true $\alpha = \gamma = 0$
  - However, $\hat{\gamma}_{ols}$ and $\hat{\alpha}_{ols}$ have nonstandard limiting distributions and are not mean zero.
  - Also, the t-stats do not have a limiting distribution, so there are no correct critical values.
  - Also, $R^2 = 1 - \frac{SSE}{SST}$ which is a ratio of variances of random walks. This converges to a nonstandard distribution which doesn't tell us much about the true variation explained.
- spurious regression: estimated relationship looks awesome when there is really nothing.

Spurious Regression

- How to recognize and characterize a spurious regression:
  - The error in a spurious regression will **not** be I(0) stationary.
  - $y_t = x_t'\beta + u_t$
  - If I can find a β that makes $u_t \sim I(0)$, then this is **not** spurious.
  - If no such β exists, then it is spurious.
  - Spurious case:
  - $y_{1t} = \alpha + \gamma y_{2t} + u_t$
  - true $\alpha = \gamma = 0 \Longrightarrow u_t = y_{1t} = \epsilon_{11} + \epsilon_{12} + ... + \epsilon_{1t}$
  - $u_t = y_{1t} \sim I(1)$.

Spurious Regression

- How to avoid a spurious regression
  1. Difference all variables before regression, or otherwise make sure all series are stationary.
     - $\Delta y_{1t} = \alpha + \gamma \Delta y_{2t} + u_t$
     - $\hat{\alpha},\ \hat{\gamma}$ in this case converge normally.
  2. Include lags of both LHS and RHS variables as additional regressors (ADL model).
     - $y_{1t} = \alpha + \gamma y_{2t} + \phi y_{1,t-1} + \delta y_{2,t-1} + u_t$
     - true $\alpha = \gamma = \delta = 0$, $\phi = 1$, $u_t = \epsilon_t = \Delta y_{1t}$

Spurious Regression

- How to avoid a spurious regression
  1. Difference all variables before regression, or otherwise make sure all series are stationary.
  2. Include lags of both LHS and RHS variables as additional regressors (ADL model).
     - $y_{1t} = \alpha + \gamma y_{2t} + \phi y_{1,t-1} + \delta y_{2,t-1} + u_t$
     - To analyze this we need to **rotate** the regressors
     - $y_{1t} = \alpha + \gamma \Delta y_{2t} + \phi y_{1,t-1} + \delta^\star y_{2,t-1} + u_t$
     - $\delta^\star = \delta + \gamma$
     - Δy_{2t} is mean-zero, stationary, while $y_{1,t-1}$ and $y_{2,t-1}$ are not.

Spurious Regression

- How to avoid a spurious regression
  1. Difference all variables before regression, or otherwise make sure all series are stationary.
  2. Include lags of both LHS and RHS variables as additional regressors (ADL model).
     - In this setup all estimates are consistent (we're not in spurious case)
     - Now γ converges to a normal, can use standard t-test.
     - Other parameters converge to something non-normal, can still use t-stat tested against alternative distributions.

Spurious Regression

- How to avoid a spurious regression
  1. Difference all variables before regression, or otherwise make sure all series are stationary.
  2. Include lags of both LHS and RHS variables as additional regressors (ADL model).
     - In original regression, γ and δ are now normal, can use standard t-stats.
     - Intuition for δ: $\delta = \delta^\star - \gamma$. $\delta^\star$ was consistent in rotated regression, converges to itself, while γ converges to a Normal.
     - F-tests of joint hypotheses are **not** standard.
     - t-test of $\phi = 1$ is nonstandard in same way as Dickey-Fuller test.

  – Conclusion: Including lags can help fix nonstationarity, but if you know they're nonstationary, just difference them first.

Unit Roots in Vector Autoregressions

- true model: $\Delta \mathbf{y_t} = \zeta_1 \Delta y_{t-1} + ... + \zeta_{\mathbf{p-1}} \Delta \mathbf{y_{t-p+1}} + \epsilon_\mathbf{t}$
  - the data **should** have been differenced. (Case 2)
  - Stationarity assumption for this model in VAR looks like
  - $|\mathbf{I_n} - \zeta_1 z - \zeta_2 z^2 - ... - \zeta_{\mathbf{p-1}} z^{p-1}| = 0$ with $||z|| < 1 \Longrightarrow \Delta \mathbf{y_t} \sim I(0)$
    - VAR(p-1) with no drift.

Unit Roots in Vector Autoregressions

- Suppose we did not difference, estimated model VAR in levels with a constant.
- $\mathbf{y_t} = \alpha + \Phi_1 \mathbf{y_{t-1}} + \Phi_2 \mathbf{y_{t-2}} + ... + \Phi_p \mathbf{y_{t-p}} + \epsilon_\mathbf{t}$
- As in previous Dickey-Fuller case, this is a rotation of
- $\mathbf{y_t} = \alpha + \rho \mathbf{y_{t-1}} + \zeta_\mathbf{1} \Delta \mathbf{y_{t-1}} + ... + \zeta_\mathbf{p-1} \Delta \mathbf{y_{t-p+1}} + \epsilon_\mathbf{t}$
- True $\alpha = \mathbf{0}$ and $\rho = \mathbf{I_n}$
- As before, $\Phi_p = -\zeta_\mathbf{p-1}$,
- $\Phi_s = \zeta_\mathbf{s} - \zeta_\mathbf{s-1}$ for $s = 2, ..., p-1$
- $\Phi_1 = \rho + \zeta_\mathbf{1}$

Unit Roots in Vector Autoregressions

- As before, all the $\zeta_{\mathbf{s}}$'s are asymptotically normal, can use standard t-tests and F-tests.
- $\hat{\alpha}$ and $\hat{\rho}$will be nonstandard but consistent (converge to something).
- Implications for estimating the levels regression:
  - t-tests and F-tests of elements of Φ_s are asymptotically valid
  - confidence intervals of impulse-response functions are asymptotically valid.
  - What's not okay: testing joint hypotheses that involve ρ.
  - For example, in testing for Granger causality, we might want to test that a particular element in $\mathbf{y_{t-s}}$ for $s = 1, ..., p$ has predictive power, or joint test of $\Phi_1^{(2,1)} = \Phi_2^{(2,1)} = ... = \Phi_p^{(2,1)} = 0$.
    - But this involves ρ and ζ, so we can't do it.
- On the other hand, if we should NOT difference, and we do, our model is misspecified - you will wash out important relationships, including cointegration.

Cointegration

- Let's illustrate with a special case:
- Suppose
  - $y_{1t} \sim I(1)$
  - $\Delta y_{2t} = u_{2t}$ with u_{2t} white noise (so y_{2t} is difference stationary).
  - $y_{1t} = \gamma y_{2t} + u_{1t}$
  - $\mathbf{u_t} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$, $E(\mathbf{u_t}) = 0$, $E(\mathbf{u_t u_t'}) = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$ if $t = s$

Cointegration

- Then $\Delta y_{1t} = \gamma \Delta y_{2t} + \Delta u_{1t} = \gamma u_{2t} + u_{1t} - u_{1,t-1} \sim I(0)$
  - This is a first-order moving average, with u_{2t} stationary.
- In other words, $y_{1t}, y_{2t} \sim I(1)$ but $y_{1t} - \gamma y_{2t} \sim I(0)$.
  - $[1 \quad -\gamma]'$ is the **cointegrating vector.**

Cointegration

Definition: an $(nX1)$ vector $\mathbf{y_t}$ is said to be cointegrated if each element y_{it} is $I(1)$ but there is a linear combination $\alpha'\mathbf{y_t} \sim I(0)$, where in general α is called the cointegrating vector.

- This is not spurious regression, even though we're regressing random walk on a random walk.
  - In spurious case, there is **no** choice of α that can make the vector stationary.
  - Spurious: $u_{1t} \sim I(1)$
  - Cointegrated: $u_{1t} \sim I(0)$

Cointegration

- In our example, y_{1t} will inherit the random walk that y_{2t} follows, but the cointegrating relationship keeps them close together.
  - They deviate by u_{1t}, which if $I(0)$ will always return to a fixed mean.
  - This is an appealing model of long run market (or ecological?) relationships.

Cointegration

- Rewrite this as a vector system:
  - $\begin{bmatrix} 1 & -\gamma \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$
  - premultiply by $\begin{bmatrix} 1 & -\gamma \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix}$
  - $\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} + \gamma u_{2t} \\ u_{2t} \end{bmatrix}$
- Let $\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} = \begin{bmatrix} u_{1t} + \gamma u_{2t} \\ u_{2t} \end{bmatrix}$, which is white noise because the u's are white noise.

Cointegration

- $E(\underline{\epsilon}_t) = \mathbf{0}$ and $E(\mathbf{e_t e_t'}) = \begin{bmatrix} \sigma_1^2 + \gamma^2\sigma_2^2 & \gamma\sigma_2^2 \\ \gamma\sigma_2^2 & \sigma_2^2 \end{bmatrix}$ if $t = s$, $\mathbf{0}$ otherwise.
- So we could run this vector system in levels.
- $\mathbf{y_t} = \Phi\mathbf{y_{t-1}} + \mathbf{e_t}$, where $\Phi = \begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix}$

Cointegration

- Notice a few things
  1. the vector $\mathbf{y_t}$ has a unit root (is nonstationary).
     1.1 $|\mathbf{I_2} - \Phi z| = \begin{vmatrix} 1 & -\gamma z \\ 0 & 1 - z \end{vmatrix} = 1 - z = 0$ when $z = 1$.
  2. we could write the system in differences as (analogous to our rotated regression).
     - $\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} + \gamma u_{2t} \\ u_{2t} \end{bmatrix}$
     - $\Delta \mathbf{y_t} = \rho \mathbf{y_{t-1}} + \mathbf{e_t}$
     - this is a hybrid of changes and levels.

Cointegration

- $\Delta \mathbf{y_t} = \underline{\rho} \mathbf{y_{t-1}} + \mathbf{e_t}$

1. If we should have had a system (VAR) in changes, then $\underline{\rho}$ will be **0**.
2. If we should have had a system (VAR) in levels (y_{1t}, y_{2t} were stationary), then $\underline{\rho}$ is an arbitrary set of coefficients and has full rank.
3. If the system is cointegrated, $\underline{\rho}$ has rank 1 and can be written as an outer product of two vectors, one if which is the cointegrating vector.
   - $\underline{\rho} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & -\gamma \end{bmatrix} = \mathbf{b}\alpha'$

Cointegration

- Why would we want to write the system this way: $\Delta \mathbf{y_t} = \rho \mathbf{y_{t-1}} + \mathbf{e_t}$?
- Notice we have $\Delta \mathbf{y_t} = \mathbf{b}\alpha' \mathbf{y_{t-1}} + \mathbf{e_t}$, and $\alpha' \mathbf{y_{t-1}} = u_{1,t-1}$
  - So $\Delta \mathbf{y_t} = \mathbf{b} u_{1,t-1} + \mathbf{e_t}$,
  - Today's changes $\Delta \mathbf{y_t}$ depend on how far y_1 and y_2 deviated from their cointegrating relation last period.
  - Because y_1 and y_2 are random walks, $\Delta \mathbf{y_t}$ should be white noise.
  - Indeed it is, but with a special kind of white noise: u_{1t-1}

Cointegration

- In general: An (nX1) vector $\mathbf{y_t}$ is said to exhibit $h < n$ cointegrating relations if
    1. each element of $\mathbf{y_t} \sim I(1)$
    2. there exists an (nXh) matrix $\mathbf{A}$ of full rank h such that each element of the (hX1) vector $\mathbf{A}'\mathbf{y_t} \sim I(0)$

    - The rows of $\mathbf{A}'$ are called the cointegrating relations or vectors.
- In other words, if $n > 2$, there might be more than one linear combination which is cointegrated.

Cointegration

- Implications for general case:
  - If we can write $\mathbf{y_t}$ as a vector system (VAR) then that system will have a unit root:
    1. $\mathbf{y_t} = \alpha + \Phi_1 \mathbf{y_{t-1}} + ... + \Phi_p \mathbf{y_{t-p}} + \mathbf{e_t}$
    2. $|\mathbf{I_n} - \Phi_1 z - ... - \Phi_p z^p| = 0$ when $z = 1$

Cointegration

- Implications for general case:
  - If previous slide holds, then we can write it as a rotated regression:
    1. $\Delta \mathbf{y_t} = \alpha + \rho \mathbf{y_{t-1}} + \zeta_1 \Delta \mathbf{y_{t-1}} + \ldots + \zeta_{p-1} \Delta \mathbf{y_{t-p+1}} + \mathbf{e_t}$
    2. $\rho = \mathbf{BA}'$ where $\mathbf{B}$ is an nXh matrix of rank h, $\mathbf{A}'$ is an hXn matrix whose rows are the cointegrating vectors.
    3. ρ has rank $h < n$. As before, $\rho = \mathbf{0}$ implies the VAR should be in differences $\Delta \mathbf{y_t}$ and if ρ has rank n, the VAR should be in levels $\mathbf{y_t}$
    4. the $\Delta \mathbf{y}$'s are stationary, and the $\mathbf{y}$'s are nonstationary. The rows of $\mathbf{A}'$ take exactly the linear combinations of the levels to not ruin the stationarity of the remaining differences.

Cointegration

- Implications for general case:
  - If previous slide holds, then we can think of $\mathbf{A}'\mathbf{y_t} = \mathbf{z_t}$ as the stationary residual from the cointegrating relations of $\mathbf{y_t}$ which is often called the "error correction term" and write
    1. $\Delta \mathbf{y_t} = \alpha + \mathbf{B}\mathbf{z_{t-1}} + \zeta_1 \Delta \mathbf{y_{t-1}} + ... + \zeta_{p-1} \Delta \mathbf{y_{t-p+1}} + \mathbf{e_t}$
    2. Notice that this is just a VAR in changes with an extra term (that would have been in the error if we didn't include it).
    3. $\mathbf{z_{t-1}}$ was the degree of deviation from the cointegrating relation last period. **B** measures how quickly we return to it.
- Conclusion: with a cointegrated system, we can estimate it as a VAR in levels, or as a VAR in differences with the error correction term.

Estimation and testing of a single cointegrating relation

- Let $\underline{\alpha}' = \begin{bmatrix} 1 & -\alpha_2 & \dots & -\alpha_n \end{bmatrix}$ be the cointegrating vector.
- $\underline{\alpha}'\mathbf{y_t} \sim I(0)$ for the true α but $\sim I(1)$ for any other α.
- Consider OLS of y_{1t} on $y_{2t}, \dots y_{n,t}$
- $y_{1t} = \alpha_2 y_{2t} + \dots + \alpha_n y_{nt} + u_t$
- $\underset{\underline{\alpha}}{min} \sum (y_{1t} - \alpha_2 y_{2t} - \dots - \alpha_n y_{nt})^2$

Estimation and testing of a single cointegrating relation

- Recall from spurious regression that $\frac{1}{T}\sum u_t^2 \to V$ if u_t stationary, and $\to \infty$ if not.
  - If T is large enough, we will be able to tell the difference from something going to infinity.
- Conclusion: OLS of y_{1t} on remaining y's is a good way to estimate $\underline{\alpha}$ and find out if it's really cointegrated or not.
- If it's not cointegrated, this will be a spurious regression.
- To tell the difference, take
  - Ho: no cointegration (spurious regression, residuals from OLS regression have a unit root).
  - Notice this is the same null hypothesis as the DF test for stationarity, which we can apply to the residuals (however, testing distribution is different).
  - If they are stationary, reject the null and conclude the series are cointegrated.

Estimation and testing of a single cointegrating relation

- Procedure (for example):

1. Estimate by OLS
   1.1 $y_{1t} = \alpha + \gamma_2 y_{2t} + ... + \gamma_n y_{nt} + u_t$
      1.1.1 includes a constant to account for any possible drift.
   1.2 save the residuals $\hat{u}_t$
2. Estimate by OLS
   2.1 $\hat{u}_t = \rho \hat{u}_{t-1} + \zeta_1 \Delta \hat{u}_{t-1} + ... + \zeta_{p-1} \Delta \hat{u}_{t-p+1} + \nu_t$
   2.2 no constant (Case 1).
   2.3 If spurious, $\rho = 1$, fail to reject DF test null
   2.4 If cointegrated, $\rho < 1$, reject DF test null.
   2.5 **caveat**: because the null involves a spurious regression, the distribution of ρ will have a **different nonstandard distribution** here than it does in the DF test.
   2.6 This is called a Phillips-Ouliaris-Hansen test - same setup as Dickey-Fuller, including different cases for drift and trend, but with different distributions.

Estimation and testing of a single cointegrating relation

- Estimating and testing of more than one cointegrating relation can be done using an MLE-type procedure called Johansen's algorithm,
  - this can be executed using the "vec" command in Stata with option , rank(#) for number of cointegrating vectors.
  - To figure out number of cointegrating vectors, use statistics derived from the Johansen MLE algorithm
    - Johansen's maximum eigenvalue, and Johansen's trace method tell you the number of cointegrating vectors.
    - Or use Stata's "vecrank" command.