Reduced Form Vector Autoregression

Time Series Econometrics

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
  - $\mathbf{x_t} = (1, \mathbf{y}'_{\mathbf{t-1}}, ..., \mathbf{y}'_{\mathbf{t-p}})'$ is a (k x 1) vector, k=np+1
  - $\mathbf{y_t}, \mathbf{x_t}$ are covariance stationary

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 $y_{1t} = \pi_1' \mathbf{x_t} + \epsilon_{1t}$
  - $\mathbf{x_t} = (1, \mathbf{y'_{t-1}}, ..., \mathbf{y'_{t-p}})'$ 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
- $$\begin{bmatrix} y_{ng,t} \\ y_{oil,t} \\ y_{dr,t} \end{bmatrix} = \begin{bmatrix} \beta_0^{ng} & \beta_{1,ng}^{ng} & \cdots & \beta_{3,dr}^{ng} \\ \beta_0^{oil} & \beta_{1,ng}^{oil} & \cdots & \beta_{3,dr}^{oil} \\ \beta_0^{dr} & \beta_{1,ng}^{dr} & \cdots & \beta_{3,dr}^{dr} \end{bmatrix} \begin{bmatrix} 1 \\ y_{ng,t-1} \\ \vdots \\ y_{dr,t-3} \end{bmatrix} + \begin{bmatrix} \epsilon_{ng,t} \\ \epsilon_{oil,t} \\ \epsilon_{dr,t} \end{bmatrix}$$
- $$\begin{bmatrix} y_{ng,t} \\ y_{oil,t} \\ y_{dr,t} \end{bmatrix} = \begin{bmatrix} \beta_0^{ng} \\ \beta_0^{oil} \\ \beta_0^{dr} \end{bmatrix} + \begin{bmatrix} \beta_{1,ng}^{ng} & \beta_{1,oil}^{ng} & \beta_{1,dr}^{ng} \\ \beta_{1,ng}^{oil} & \beta_{1,oil}^{oil} & \beta_{1,dr}^{oil} \\ \beta_{1,ng}^{dr} & \beta_{1,oil}^{dr} & \beta_{1,dr}^{dr} \end{bmatrix} \begin{bmatrix} y_{ng,t-1} \\ y_{oil,t-1} \\ y_{dr,t-1} \end{bmatrix}$$
  $$+ \dots + \begin{bmatrix} \beta_{3,ng}^{ng} & \beta_{3,oil}^{ng} & \beta_{3,dr}^{ng} \\ \beta_{3,ng}^{oil} & \beta_{3,oil}^{oil} & \beta_{3,dr}^{oil} \\ \beta_{3,ng}^{dr} & \beta_{3,oil}^{dr} & \beta_{3,dr}^{dr} \end{bmatrix} \begin{bmatrix} y_{ng,t-3} \\ y_{oil,t-3} \\ y_{dr,t-3} \end{bmatrix} + \begin{bmatrix} \epsilon_{ng,t} \\ \epsilon_{oil,t} \\ \epsilon_{dr,t} \end{bmatrix}$$

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 $-\frac{Tn}{2}(1+log 2\pi) - \frac{T}{2} log|\hat{\Omega}|$
  - 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 $\underset{nX1}{\mathbf{y_t}} = \mathbf{c} + \Phi_1 \mathbf{y_{t-1}} + ... + \Phi_p \mathbf{y_{t-p}} + \epsilon_\mathbf{t}$ in companion form:
- $$\underset{(npX1)}{\begin{bmatrix} \mathbf{y_t} \\ \mathbf{y_{t-1}} \\ \vdots \\ \mathbf{y_{t-p+1}} \end{bmatrix}} = \begin{bmatrix} \mathbf{c} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} + \underset{(npXnp)}{\begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ \mathbf{I_n} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & & \vdots \\ \mathbf{0} & \dots & \mathbf{I_n} & \mathbf{0} \end{bmatrix}} \begin{bmatrix} \mathbf{y_{t-1}} \\ \mathbf{y_{t-2}} \\ \vdots \\ \mathbf{y_{t-p}} \end{bmatrix} + \begin{bmatrix} \epsilon_\mathbf{t} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

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(∞):
  - $\mathbf{y_{t+s}} = \mu + \epsilon_{\mathbf{t+s}} + \Psi_1 \epsilon_{\mathbf{t+s-1}} + \Psi_2 \epsilon_{\mathbf{t+s-2}} + \dots$
  - 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.

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

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 **m** generally.
  - Depends on the eigenvalues of **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 $\mathbf{F^2} = (\mathbf{T}\Lambda\mathbf{T}^{-1})(\mathbf{T}\Lambda\mathbf{T}^{-1}) = \mathbf{T}\Lambda^2\mathbf{T}^{-1}$ or $\mathbf{F^m} = \mathbf{T}\Lambda^m\mathbf{T}^{-1}$, where
  - $\Lambda^m = \begin{bmatrix} \lambda_1^m & \cdots & 0 \\ \vdots & \ddots & \\ 0 & & \lambda_n^m \end{bmatrix}$
  - 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 $\mathbf{T}\Lambda^m\mathbf{T}^{-1}$, 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 $\mathbf{T}^{-1}$ 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 **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 $\mathbf{y_t} = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}$
- $\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \end{bmatrix}$
- $\mathbf{y_t} = \mathbf{F}\mathbf{y_{t-1}} + \mathbf{w_t}$
- Call this the “companion form”.
- We have already analyzed this as a first order vector system.

p^{th}-order scalar difference equation

- $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-1} + \ldots + \phi_p y_{t-p} + w_t$
- Companion form:
- $\mathbf{y_t} = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}$, $\mathbf{F} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & & & 0 \\ 0 & 1 & & & 0 \\ \vdots & & \ddots & & \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$, $\mathbf{w_t} = \begin{bmatrix} w_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$
- $\mathbf{y_t} = \mathbf{F}\mathbf{y_{t-1}} + \mathbf{w_t}$
- $\frac{\partial y_{1,t+m}}{\partial w_{1t}} = c_1 \lambda_1^m + \ldots + c_p \lambda_p^m$ 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.