Time Series Econometrics, 2ST111

Lecture 7. Vector Autoregressions

Yukai Yang

Department of Statistics, Uppsala University

Outline of Today's Lecture

- Chapter 11: Vector Autoregressions (not 11.6-11.7)
- *Vector Autoregressions* by Stock and Watson (2001) (unless you're interested, you may skip the parts about structural models)

Some history

- Christopher A. Sims, Princeton University
- Awarded the Nobel Prize in Economics in 2011 together with Thomas J. Sargent "for their empirical research on cause and effect in the macroeconomy"
- "Macroeconomics and Reality" (1980, in Econometrica) is a seminal paper in the field
- Many more important contributions

ADL (Autoregressive distributed lag)

$Y_{1t} = c + \beta_1 Y_{1,t-1} + \ldots$

ML Estimation

- Suppose that the model is

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1\mathbf{y}_{t-1} + \mathbf{\Phi}_2\mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

  where $\boldsymbol{\varepsilon}_t \sim$ i.i.d. $N(\mathbf{0}, \mathbf{\Omega})$.
- Assume we have a sample of length $T+p$, i.e. $\mathbf{y}_{-p+1}, \ldots, \mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_T$. The simplest method of estimation is then to condition upon $(\mathbf{y}_{-p+1}, \ldots, \mathbf{y}_0)$ and maximize the conditional likelihood

$$f_{\mathbf{Y}_T, \mathbf{Y}_{T-1}, \ldots, \mathbf{Y}_1 | \mathbf{Y}_0, \ldots, \mathbf{Y}_{-p+1}}(\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_1 | \mathbf{y}_0, \ldots, \mathbf{y}_{-p+1}; \boldsymbol{\theta})$$

  where $\boldsymbol{\theta}$ contains all the unknowns $\mathbf{c}, \mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p$ and $\mathbf{\Omega}$.

ML Estimation

- For notational convenience, let

$$\mathbf{x}_t = \begin{pmatrix} 1 \\ \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p} \end{pmatrix}, \quad \mathbf{\Pi} = \begin{pmatrix} \mathbf{c}' \\ \mathbf{\Phi}_1' \\ \mathbf{\Phi}_2' \\ \vdots \\ \mathbf{\Phi}_p' \end{pmatrix} \Longrightarrow \mathbf{y}_t = \mathbf{\Pi}'\mathbf{x}_t + \varepsilon_t$$

- It thus follows that

$$\mathbf{y}_t | \mathbf{y}_{t-1}, \ldots, \mathbf{y}_{-p+1} \sim N(\mathbf{\Pi}'\mathbf{x}_t, \mathbf{\Omega})$$

and

$$f_t = (2\pi)^{-n/2} |\mathbf{\Omega}|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{y}_t - \mathbf{\Pi}'\mathbf{x}_t)' \mathbf{\Omega}^{-1} (\mathbf{y}_t - \mathbf{\Pi}'\mathbf{x}_t) \right\}$$

ML Estimation

- The joint density is the product of the individual conditional densities:

$$
\begin{aligned}
& f_{\mathbf{Y}_T,\mathbf{Y}_{T-1},\ldots,\mathbf{Y}_1|\mathbf{Y}_0,\ldots,\mathbf{Y}_{-p+1}}(\mathbf{y}_T,\mathbf{y}_{T-1},\ldots,\mathbf{y}_1|\mathbf{y}_0,\ldots,\mathbf{y}_{-p+1};\boldsymbol{\theta}) \\
&= \prod_{t=1}^{T} f_{\mathbf{Y}_t|\mathbf{Y}_{t-1},\ldots,\mathbf{Y}_{t-p}}(\mathbf{y}_t|\mathbf{y}_{t-1},\ldots,\mathbf{y}_{t-p};\boldsymbol{\theta}) \\
&= \prod_{t=1}^{T} f_t.
\end{aligned}
$$

- Eventually, we end up with the log likelihood

$$
\mathcal{L}(\theta) = c + \frac{T}{2}\log(|\boldsymbol{\Omega}^{-1}|) - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)
$$

ML Estimation

- To maximize, differentiate with respect to $\mathbf{\Pi}$, set to 0 and solve
- Useful derivative: For a symmetric matrix $\mathbf{W}$,

$$\frac{\partial}{\partial \mathbf{A}}(\mathbf{x} - \mathbf{A}\mathbf{s})'\mathbf{W}(\mathbf{x} - \mathbf{A}\mathbf{s}) = -2\mathbf{W}(\mathbf{x} - \mathbf{A}\mathbf{s})\mathbf{s}'$$

- Hence,

$$\begin{aligned}\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \mathbf{\Pi}'} &= \sum_{t=1}^{T} \mathbf{\Omega}^{-1}(\mathbf{y}_t - \mathbf{\Pi}'\mathbf{x}_t)\mathbf{x}_t' = \mathbf{0} \\ &= \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{\Pi}'\mathbf{x}_t)\mathbf{x}_t' = \mathbf{0}\end{aligned}$$

- The ML estimator is therefore the OLS estimator,

$$\hat{\mathbf{\Pi}} = \left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{y}_t'\right)$$

- Furthermore, maximum (conditional) likelihood estimation of a VAR is equivalent to equation-by-equation OLS
- By straight-forward matrix calculus, the ML estimator of the error covariance matrix can be shown to be

$$\hat{\boldsymbol{\Omega}} = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

where $\hat{\varepsilon}_t = \mathbf{y}_t - \hat{\boldsymbol{\Pi}}' \mathbf{x}_t$, i.e. evaluated at $\hat{\boldsymbol{\Pi}}$.

Lag selection

- Likelihood ratio test for
  H_0: The model has p_0 lags
  H_1: The model has $p_1 > p_0$ lags
- The likelihood under H_0 ($i = 0$) and H_1 ($i = 1$)

$$\mathcal{L}(\hat{\boldsymbol{\Omega}}_i, \hat{\boldsymbol{\Pi}}_i) = c + \frac{T}{2}\log(|\hat{\boldsymbol{\Omega}}_i^{-1}|) - \frac{1}{2}\sum_{t=1}^{T}\hat{\varepsilon}'_{t,i}\hat{\boldsymbol{\Omega}}_i^{-1}\hat{\varepsilon}_{t,i}, \quad i = 0, 1$$

- The last term is for both $i = 0$ and $i = 1$ equal to $-Tn/2$. Put this into a new constant $c^* = c - Tn/2$. Thus, minus two times the log likelihood ratio is

$$\begin{aligned}
\Lambda &= -2\left(\mathcal{L}(\hat{\boldsymbol{\Omega}}_0, \hat{\boldsymbol{\Pi}}_0) - \mathcal{L}(\hat{\boldsymbol{\Omega}}_1, \hat{\boldsymbol{\Pi}}_1)\right) = \\
&= -2\left(c^* + \frac{T}{2}\log(|\hat{\boldsymbol{\Omega}}_0^{-1}|) - c^* - \frac{T}{2}\log(|\hat{\boldsymbol{\Omega}}_1^{-1}|)\right) \\
&= T\left(\log(|\hat{\boldsymbol{\Omega}}_1^{-1}|) - \log(|\hat{\boldsymbol{\Omega}}_0^{-1}|)\right) \\
&= T\left(\log(|\hat{\boldsymbol{\Omega}}_0|) - \log(|\hat{\boldsymbol{\Omega}}_1|)\right)
\end{aligned}$$

Lag selection

- How many restrictions are imposed under H_0?
  - Each equation has $p_1 - p_0$ fewer lags per variable, i.e. $n(p_1 - p_0)$ parameters are restricted to 0
  - n equations, so $n^2(p_1 - p_0)$ restrictions
- Under H_0, $\Lambda \sim \chi^2 \left(n^2(p_1 - p_0)\right)$
- Finding the lag length using this procedure means sequential testing of hypotheses, quite complicated to control significance levels
- Sometimes prediction is the objective - the correct order of the model is uninteresting, a model suitable for forecasting is desired

Lag selection

- It is important to choose an appropriate lag length: too few will make the residuals correlated, too many make estimates imprecise and forecasts worse
- If forecasting is the objective, one can find the order which minimizes some forecast measure
- The usual model selection criteria are often used:

$$AIC(p) = \ln\left|\hat{\boldsymbol{\Omega}}\right| + \frac{2}{T}n(np+1)$$
$$BIC(p) = \ln\left|\hat{\boldsymbol{\Omega}}\right| + \frac{\ln T}{T}n(np+1)$$
$$HQ(p) = \ln\left|\hat{\boldsymbol{\Omega}}\right| + \frac{2\ln(\ln T)}{T}n(np+1)$$

- It is common to use these criteria together with residual tests (e.g. for autocorrelation)

Lag selection

- Having selected the lag length, how do we summarize and present the results?
- There are often a huge number of parameters involved, so looking at the estimated coefficients individually is pointless
- Some main tools:
  - Impulse response analysis: summarizes the dynamics in the model
  - Granger causality: are certain variables important for the prediction of others?
  - Variance decomposition: how much of the unexplained variance in one variables can be traced back to unexplained shocks to other variables?

Impulse responses

- Impulse responses are often of great interest to researchers
- How do shocks transmit in the system?
- Consider a trivariate VAR and a shock in variable 1 at time $t = 0$
  - Because of the lag structure, the shock in variable 1 affects variables 1-3 at $t = 1$
  - Similarly, at $t = 2$, *all* variables are affected by *all* variables
- Example of a trivariate VAR:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{pmatrix} = \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{pmatrix}$$

- So the shock affects i) y_1 directly, ii) y_2 at $t = 1$, and iii) y_3 at $t = 2$ (since the shock in y_1 goes through y_2 as there is no direct connection between y_1 and y_3)

Impulse responses

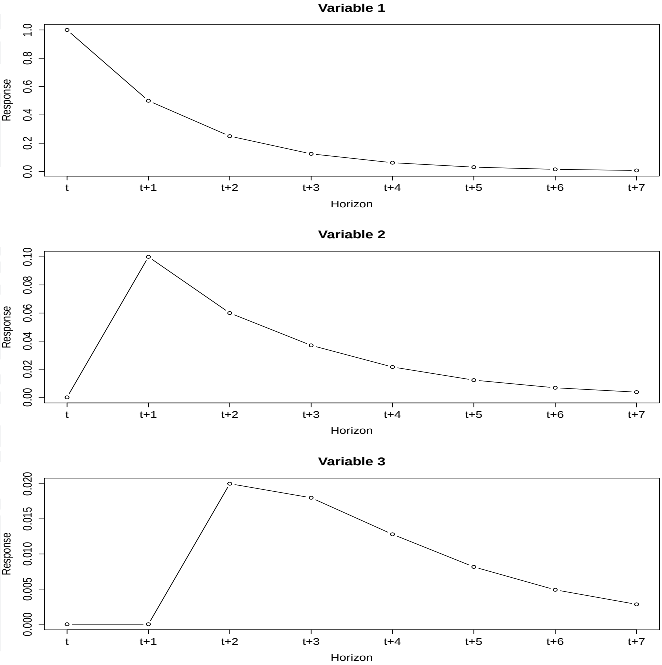
- The impulse response for a univariate process is $\frac{\partial y_{t+s}}{\partial \epsilon_t}$
- For a multivariate model, we might be interested in the effect on variable i at time $t+s$ of a shock to variable j at time t:

$$\frac{\partial y_{i,t+s}}{\partial \epsilon_{j,t}}$$

- Example: what is the effect on the inflation rate of a monetary policy shock?
- From the VMA(∞) form,

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\varepsilon}_t'} = \frac{\partial}{\partial \boldsymbol{\varepsilon}_t'} \left(\boldsymbol{\mu} + \boldsymbol{\varepsilon}_{t+s} + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t+s-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t+s-2} + \cdots \right) = \boldsymbol{\Psi}_s$$

- Element (i,j) is $\frac{\partial y_{i,t+s}}{\partial \epsilon_{j,t}}$, so plotting this for $s = 0, 1, 2, \ldots$ produces a plot of the impulse response function


Variable 1
Response
1.0
0.8
0.6
0.4
0.2
0.0
t
t+1
t+2
t+3
t+4
t+5
t+6
t+7
Horizon
Variable 2
Response
0.10
0.08
0.06
0.04
0.02
0.00
t
t+1
t+2
t+3
t+4
t+5
t+6
t+7
Horizon
Variable 3
Response
0.020
0.015
0.010
0.005
0.000
t
t+1
t+2
t+3
t+4
t+5
t+6
t+7
Horizon

IRF: $\mathbb{E}(y_{i,t+h} \mid F_t, \varepsilon_{it} = \delta)$

$- \mathbb{E}(y_{i,t+h} \mid F_t, \varepsilon_{it} = 0)$

Cholesky: $\Omega = PP'$

$P = \triangle$ matrix

$\varepsilon_t \sim (0, \Omega)$

$u_t = P^{-1}\varepsilon_t \sim (0, 1)$

$$P^{-1}\varepsilon_t = \begin{bmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} = \begin{bmatrix} p_{11}\varepsilon_{1t} \\ \cdots \end{bmatrix}$$

Impulse responses

- One serious problem: what is the meaning of this?
- Recall: $E(\varepsilon_t \varepsilon_t') = \mathbf{\Omega}$, which is (usually) not a diagonal matrix
- Example for the trivariate VAR:

$$\mathbf{\Omega} = \begin{pmatrix} 2.25 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 0.74 \end{pmatrix}$$

- If we again consider a shock in y_1, this means that this shock is likely accompanied by a shock in y_2 as well
- The workaround is to orthogonalize the system

Impulse responses

- Cholesky decomposition: $\boldsymbol{\Omega} = \mathbf{P}\mathbf{P}'$, and we let $\mathbf{v}_t = \mathbf{P}^{-1}\varepsilon_t$:

$$E(\mathbf{v}_t\mathbf{v}_t') = \mathbf{P}^{-1}E(\varepsilon_t\varepsilon_t')\left(\mathbf{P}^{-1}\right)' = \mathbf{P}^{-1}\mathbf{P}\mathbf{P}'\left(\mathbf{P}^{-1}\right)' = \mathbf{I}$$

- VMA(∞) form again:

$$\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\mu} + \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s \varepsilon_{t-s} \\
&= \boldsymbol{\mu} + \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s \mathbf{P}\mathbf{P}^{-1} \varepsilon_{t-s} \\
&= \boldsymbol{\mu} + \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s^* \mathbf{v}_{t-s}
\end{aligned}$$

Impulse responses

- With orthogonal errors, the derivative $\frac{\partial y_{i,t+s}}{\partial v_{j,t}}$ makes sense as an isolated change
- However, new problems arise: the order of the variables matter because of the decomposition

$$\frac{\partial \mathbf{y}_{t+s}}{\partial v_{j,t}} = \mathbf{\Psi}_s \mathbf{p}_j$$

where $\mathbf{p}_j$ is column j of $\mathbf{P}$, a lower triangular matrix.
- Example: three variables

$$\frac{\partial \mathbf{y}_{t+s}}{\partial v_{1,t}} = \mathbf{\Psi}_s \begin{pmatrix} p_{11} \\ p_{21} \\ p_{31} \end{pmatrix}, \quad \frac{\partial \mathbf{y}_{t+s}}{\partial v_{2,t}} = \mathbf{\Psi}_s \begin{pmatrix} 0 \\ p_{22} \\ p_{32} \end{pmatrix}, \quad \frac{\partial \mathbf{y}_{t+s}}{\partial v_{3,t}} = \mathbf{\Psi}_s \begin{pmatrix} 0 \\ 0 \\ p_{33} \end{pmatrix}$$

- Order matters, and it cannot be determined by statistical procedures but must be chosen

Impulse responses

- Consider a simple bivariate VAR(1):

$$
\begin{aligned}
y_t &= 0.5y_{t-1} + 0.2x_{t-1} + \epsilon_{y,t} \\
x_t &= 0.3y_{t-1} - 0.1x_{t-1} + \epsilon_{x,t}
\end{aligned}
$$

which we write as

$$
\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0.5 & 0.2 \\ 0.3 & -0.1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{y,t} \\ \epsilon_{x,t} \end{pmatrix}.
$$

- With covariance

$$
\mathbf{\Omega} = \begin{pmatrix} 2 & 0.2 \\ 0.2 & 3 \end{pmatrix} = \overbrace{\begin{pmatrix} 1.41 & 0 \\ 0.14 & 1.73 \end{pmatrix}}^{\mathbf{P}} \overbrace{\begin{pmatrix} 1.41 & 0.14 \\ 0 & 1.73 \end{pmatrix}}^{\mathbf{P}'}
$$

Impulse responses

- Thus, for an orthogonal shock in y:

$$\frac{\partial}{\partial v_{y,t}} \begin{pmatrix} y_{t+1} \\ x_{t+1} \end{pmatrix} = \mathbf{\Psi}_1 \begin{pmatrix} p_{11} \\ p_{21} \end{pmatrix} = \begin{pmatrix} 0.5 & 0.2 \\ 0.3 & -0.1 \end{pmatrix} \begin{pmatrix} 1.41 \\ 0.14 \end{pmatrix} = \begin{pmatrix} 0.75 \\ 0.27 \end{pmatrix}$$

- What if we instead had ordered x before y?

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} -0.1 & 0.3 \\ 0.2 & 0.5 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{x,t} \\ \epsilon_{y,t} \end{pmatrix}$$

$$\mathbf{\Omega} = \begin{pmatrix} 3 & 0.2 \\ 0.2 & 2 \end{pmatrix} = \overbrace{\begin{pmatrix} 1.73 & 0 \\ 0.12 & 1.41 \end{pmatrix}}^{\mathbf{P}_*} \overbrace{\begin{pmatrix} 1.73 & 0.12 \\ 0 & 1.41 \end{pmatrix}}^{\mathbf{P}_*'}$$

- Another orthogonal shock, still in y:

$$\frac{\partial}{\partial v_{y,t}} \begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \mathbf{\Psi}_1 \begin{pmatrix} 0 \\ p_{22}^* \end{pmatrix} = \begin{pmatrix} -0.1 & 0.3 \\ 0.2 & 0.5 \end{pmatrix} \begin{pmatrix} 0 \\ 1.41 \end{pmatrix} = \begin{pmatrix} 0.28 \\ 0.70 \end{pmatrix}$$

Forecast error variance decomposition

- A closely related concept is forecast error variance decomposition
- How much of the variance of the forecast error of $y_{i,t+s}$ is due to an exogenous shock to $y_{j,t}$?
- Recall two of our previous expressions:

$$\begin{aligned}\mathbf{y}_{t+s} &= \boldsymbol{\mu} + \mathbf{F}_{11}^{(s)}(\mathbf{y}_t - \boldsymbol{\mu}) + \mathbf{F}_{12}^{(s)}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \cdots + \mathbf{F}_{1p}^{(s)}(\mathbf{y}_{t-p+1} - \boldsymbol{\mu}) \\ &\quad + \varepsilon_{t+s} + \boldsymbol{\Psi}_1 \varepsilon_{t+s-1} + \boldsymbol{\Psi}_2 \varepsilon_{t+s-2} + \cdots + \boldsymbol{\Psi}_{s-1} \varepsilon_{t+1} \\ \hat{\mathbf{y}}_{t+s} &= \boldsymbol{\mu} + \mathbf{F}_{11}^{(s)}(\mathbf{y}_t - \boldsymbol{\mu}) + \mathbf{F}_{12}^{(s)}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \cdots + \mathbf{F}_{1p}^{(s)}(\mathbf{y}_{t-p+1} - \boldsymbol{\mu})\end{aligned}$$

- The forecast error is therefore:

$$\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s} = \varepsilon_{t+s} + \boldsymbol{\Psi}_1 \varepsilon_{t+s-1} + \boldsymbol{\Psi}_2 \varepsilon_{t+s-2} + \cdots + \boldsymbol{\Psi}_{s-1} \varepsilon_{t+1}$$

- This means that the forecast error is due to exogenous innovations

Forecast error variance decomposition

- The MSE of the forecast is:

$$
\begin{aligned}
MSE(\hat{\mathbf{y}}_{t+s|t}) &= E[(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})'] \\
&= \mathbf{\Omega} + \mathbf{\Psi}_1 \mathbf{\Omega} \mathbf{\Psi}_1' + \mathbf{\Psi}_2 \mathbf{\Omega} \mathbf{\Psi}_2' + \cdots + \mathbf{\Psi}_{s-1} \mathbf{\Omega} \mathbf{\Psi}_{s-1}' \qquad (1)
\end{aligned}
$$

  since $E(\varepsilon_t \varepsilon_\tau') = \mathbf{0}$ if $t \neq \tau$

- **Key idea**: how much does each of the *orthogonal* disturbances contribute to this MSE?
- To orthogonalize, let $\mathbf{\Omega} = \mathbf{P}\mathbf{P}'$ and

$$
\varepsilon_t = \mathbf{P}\mathbf{v}_t = \mathbf{p}_1 v_{1,t} + \mathbf{p}_2 v_{2,t} + \cdots + \mathbf{p}_n v_{n,t}
$$

- The v_{it} and v_{jt} terms are orthogonal and have unit variance, so

$$
\begin{aligned}
\mathbf{\Omega} &= E(\varepsilon_t \varepsilon_t') \\
&= \mathbf{p}_1 \mathbf{p}_1' V(v_{1,t}) + \mathbf{p}_2 \mathbf{p}_2' V(v_{2,t}) + \cdots + \mathbf{p}_n \mathbf{p}_n' V(v_{n,t}) \\
&= \mathbf{p}_1 \mathbf{p}_1' + \mathbf{p}_2 \mathbf{p}_2' + \cdots + \mathbf{p}_n \mathbf{p}_n' \qquad (2)
\end{aligned}
$$

Forecast error variance decomposition

- Now: take the MSE expression in (1) and replace $\boldsymbol{\Omega}$ with (2)

$$MSE(\hat{\mathbf{y}}_{t+s|t}) =$$
$$\sum_{j=1}^{n} \left(\mathbf{p}_j \mathbf{p}_j' + \boldsymbol{\Psi}_1 \mathbf{p}_j \mathbf{p}_j' \boldsymbol{\Psi}_1' + \boldsymbol{\Psi}_2 \mathbf{p}_j \mathbf{p}_j' \boldsymbol{\Psi}_2' + \cdots + \boldsymbol{\Psi}_{s-1} \mathbf{p}_j \mathbf{p}_j' \boldsymbol{\Psi}_{s-1}' \right)$$

- Each j in the sum is the contribution by each variable to the MSE at horizon s
- Notation: call the term in brackets $\boldsymbol{\Xi}_{j,s}$ and $\sum_{j=1}^{n} \boldsymbol{\Xi}_{j,s} = \boldsymbol{\Xi}_s$
- The proportion of forecast error variance of variable m attributable to variable j at horizon s is then

$$\frac{\boldsymbol{\Xi}_{j,s}(m,m)}{\sum_{j=1}^{n} \boldsymbol{\Xi}_{j,s}(m,m)} = \frac{\boldsymbol{\Xi}_{j,s}(m,m)}{\boldsymbol{\Xi}_s(m,m)}$$

- Numerator: the diagonal of $\boldsymbol{\Xi}_{j,s}$ gives the contribution of variable j to MSE
- Denominator: the diagonal contains the variables' total MSEs

Granger causality

- Granger causality has little to do with causality; it is used to see if lags of one variable are useful in forecasting another
- A variable y Granger-causes x if lags of y improve forecasts of x
- More specifically, if the MSE of a forecast of x_{t+s} based on $(x_t, x_{t-1}, \dots)$ is the same as a forecast based on both $(x_t, x_{t-1}, \dots)$ and $(y_t, y_{t-1}, \dots)$, then y does *not* Granger-cause x
- In a VAR model, this is simply a joint test of certain coefficients being zero

Granger causality

- A bivariate VAR(p):

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \phi_{11}^{(p)} & \phi_{12}^{(p)} \\ \phi_{21}^{(p)} & \phi_{22}^{(p)} \end{pmatrix} \begin{pmatrix} x_{t-p} \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$$

- The optimal forecast for x_{t+1} is:

$$\begin{aligned} &\hat{E}(x_{t+1}|x_t, x_{t-1}, \ldots, y_t, y_{t-1}, \ldots) \\ &= \phi_{11}^{(1)} x_t + \cdots + \phi_{11}^{(p)} x_{t-p+1} + \phi_{12}^{(1)} y_t + \cdots + \phi_{12}^{(p)} y_{t-p+1} \end{aligned}$$

- Thus, if $\phi_{12}^{(1)} = \cdots = \phi_{12}^{(1)} = 0$, the forecast depends only on lagged values of x itself and we get:

$$\hat{E}(x_{t+1}|x_t, x_{t-1}, \ldots, y_t, y_{t-1}, \ldots) = \hat{E}(x_{t+1}|x_t, x_{t-1}, \ldots)$$

Granger causality

- To test for Granger causality, we regress x_t on lags of x and y:

$$x_t = \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{i=1}^{p} \beta_i y_{t-i} + u_t$$

- Conduct an F-test with the null hypothesis:

$$H_0 : \quad \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

Extensions

There are many extensions

- Threshold VARs
- Smooth Transition VARs
- Markov-switching VARs
- Time-varying parameters VARs
- VARs with stochastic volatility
- Factor-augmented VARs
- etcetera...

Stock and Watson (2001)

- Stock and Watson: two of the leading macroeconometricians in the world
- This paper discusses how well VAR models handle what they're frequently used to do
  - Data description
  - Forecasting
  - (Structural inference)
  - (Policy analysis)

Stock and Watson (2001)

- VAR models come in one of three forms: reduced, recursive or structural
- **Reduced** form is what we have discussed so far, where each variable is a (linear) function of past values of itself and the other variables
- **Recursive** form we used when we had orthogonalized impulse responses, since adding contemporaneous lags is equivalent to doing a Cholesky decomposition
- **Structural** VARs are based on economic theory and make use of identifying assumptions therein
- Their data ranges from 1960Q1-2000Q4 and includes π (inflation rate), u (unemployment rate) and R (the federal funds rate, i.e. an interest rate)

Recursive:

$$P^{-1}\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix}, \quad P^{-1}\varepsilon_t = V_t \sim (0, I)$$

$$\begin{pmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{pmatrix}\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} \cdot \end{pmatrix}$$

$$P_{11} Y_{1t} = \ldots$$

$$Y_{1t} = \frac{1}{P_{11}}(\ldots)$$

$$P_{22} Y_{2t} = P_{21} Y_{1t} + \ldots$$

$$= \left(\frac{P_{21}}{P_{11}}\right)(\ldots)$$

Data description: Granger-Causality

- Granger-Causality Test: what variables help predict others?
- H_0 : the regressor does not Granger-cause the dependent variable

A. Granger-Causality Tests

	Dependent Variable in Regression		
Regressor	π	u	R
π	0.00	0.31	0.00
u	0.02	0.00	0.00
R	0.27	0.01	0.00

Figure: *p*-values of Granger-causality tests (Table 1, Panel A)

Data description: Variance decomposition

- The (forecast error) variance decomposition tells us the percentage of the error in forecasting a variable (e.g. inflation) that is due to specific shocks in another variable (such as unemployment) at a specific horizon (like 4 quarters)

B.i. Variance Decomposition of π

Forecast Horizon	*Forecast Standard Error*	*Variance Decomposition (Percentage Points)*		
		π	u	R
1	0.96	100	0	0
4	1.34	88	10	2
8	1.75	82	17	1
12	1.97	82	16	2

Figure: Variance decomposition (Table 1, Panel B.i)

Data description: Variance decomposition

B.i. Variance Decomposition of π

Forecast Horizon		Forecast Standard Error	Variance Decomposition (Percentage Points) π	u	R
1	$\sqrt{\Xi_1(1,1)}$	0.96	100	0	0
4	$\sqrt{\Xi_4(1,1)}$	1.34	88	10	2
8	$\sqrt{\Xi_8(1,1)}$	1.75	82	17	1
12	$\sqrt{\Xi_{12}(1,1)}$	1.97	82	16	2

- The numbers in the u column are given by

$$\frac{\Xi_{2,1}(1,1)}{\Xi_1(1,1)} = 0, \quad \frac{\Xi_{2,4}(1,1)}{\Xi_4(1,1)} = 0.10$$

$$\frac{\Xi_{2,8}(1,1)}{\Xi_8(1,1)} = 0.17, \quad \frac{\Xi_{2,12}(1,1)}{\Xi_{12}(1,1)} = 0.16$$

- Error for variable 1
- at horizon 1, 4, 8, 12
- due to variable 2

Data description: Variance decomposition

B.ii. Variance Decomposition of u

Forecast Horizon		Forecast Standard Error	Variance Decomposition (Percentage Points) π	u	R
1	$\sqrt{\Xi_1(2,2)}$	0.23	1	99	0
4	$\sqrt{\Xi_4(2,2)}$	0.64	0	98	2
8	$\sqrt{\Xi_8(2,2)}$	0.79	7	82	11
12	$\sqrt{\Xi_{12}(2,2)}$	0.92	16	66	18

- The numbers in the u column are given by

$$\frac{\Xi_{2,1}(2,2)}{\Xi_1(2,2)} = 0.99, \quad \frac{\Xi_{2,4}(2,2)}{\Xi_4(2,2)} = 0.98$$
$$\frac{\Xi_{2,8}(2,2)}{\Xi_8(2,2)} = 0.82, \quad \frac{\Xi_{2,12}(2,2)}{\Xi_{12}(2,2)} = 0.66$$

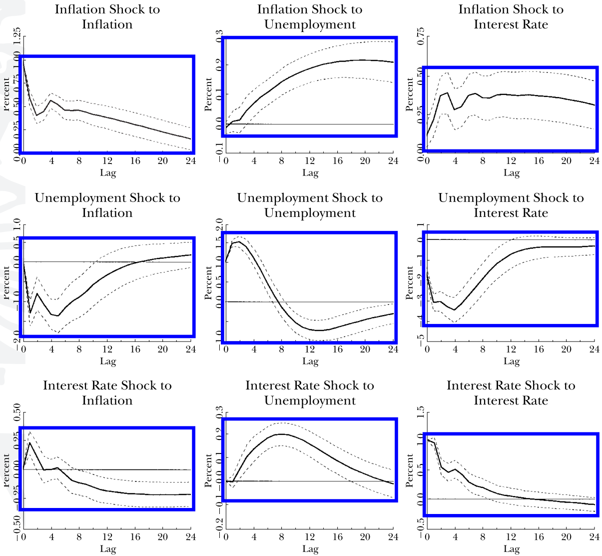
Data description: Variance decomposition

B.iii. Variance Decomposition of R

Forecast Horizon		*Forecast Standard Error*	*Variance Decomposition (Percentage Points)* π	u	R
1	$\sqrt{\Xi_1(3,3)}$	0.85	2	19	79
4	$\sqrt{\Xi_4(3,3)}$	1.84	9	50	41
8	$\sqrt{\Xi_8(3,3)}$	2.44	12	60	28
12	$\sqrt{\Xi_{12}(3,3)}$	2.63	16	59	25

- The numbers in the u column are given by

$$\frac{\Xi_{2,1}(3,3)}{\Xi_1(3,3)} = 0.19, \quad \frac{\Xi_{2,4}(3,3)}{\Xi_4(3,3)} = 0.50$$
$$\frac{\Xi_{2,8}(3,3)}{\Xi_8(3,3)} = 0.60, \quad \frac{\Xi_{2,12}(3,3)}{\Xi_{12}(3,3)} = 0.59$$


Inflation Shock to Inflation
Inflation Shock to Unemployment
Inflation Shock to Interest Rate
Unemployment Shock to Inflation
Unemployment Shock to Unemployment
Unemployment Shock to Interest Rate
Interest Rate Shock to Inflation
Interest Rate Shock to Unemployment
Interest Rate Shock to Interest Rate
Percent
Lag

Data description: Forecasting

- VARs are often used for forecasting
- Pseudo out-of-sample forecasting exercise for the period 1985Q1-2000Q4 using a rolling forecast window:
  - Estimate model on data through 1984Q4 and predict h steps ahead
  - Add one more data point: estimate model on data through 1985Q1 and predict h steps ahead
  - Continue until the sample ends, repeat for $h = 2, 4, 8$
- Call the forecasts $\hat{\pi}_t^{(h)}$, $\hat{u}_t^{(h)}$ and $\hat{R}_t^{(h)}$
- Stock and Watson evaluate the forecasts using the standard measure RMSE:

$$RMSE_h(\pi) = \sqrt{\sum_{t=1984Q4+h}^{2000Q4} \frac{\left(\pi_t - \hat{\pi}_t^{(h)}\right)^2}{\#\text{ of forecasts}}}$$

Data description: Forecasting

- It is common practice to include AR(1) and random walks as benchmark models

Forecast Horizon	*Inflation Rate*			*Unemployment Rate*			*Interest Rate*		
	RW	*AR*	*VAR*	*RW*	*AR*	*VAR*	*RW*	*AR*	*VAR*
2 quarters	0.82	0.70	0.68	0.34	0.28	0.29	0.79	0.77	0.68
4 quarters	0.73	0.65	0.63	0.62	0.52	0.53	1.36	1.25	1.07
8 quarters	0.75	0.75	0.75	1.12	0.95	0.78	2.18	1.92	1.70

Figure: RMSE of pseudo out-of-sample forecasts

- It is often quite difficult to beat simple AR models, but here the VAR is most of the time slightly better

Conclusions

- VAR models have been very useful tools for macroeconometricians for almost four decades
- There are limitations, but competing models are usually much more complicated for little or no gain
- Much of the recent research is focused on fixing its limitations: dealing with overparametrization, allowing for nonlinearities in various ways, using larger data sets

To be continued! Thank you!