A Primer on Vector Autoregressions

Ambrogio Cesa-Bianchi

[DISCLAIMER]

These notes are meant to provide intuition on the basic mechanisms of VARs

As such, most of the material covered here is treated in a very informal way

If you crave a formal treatment of these topics, you should stop here and buy a copy of Hamilton's "Time Series Analysis"

VARs & Macro-econometricians' job

- According to a well-known paper by Stock & Watson (2001, JEP) macroeconometricians (would like to) do four things
  1. Describe and summarize macroeconomic time series
  2. Make forecasts
  3. Recover the true structure of the macroeconomy from the data
  4. Advise macroeconomic policymakers
- Vector autoregressive models are a statistical tool to address these tasks

What can we do with vector autoregressive models?

- 3 variables: real GDP growth (Δy), inflation (π) and the policy rate (i)
- A VAR can help us answering the following questions
  1. What is the dynamic behaviour of these variables? How do these variables interact?
  2. What is the profile of GDP conditional on a specific future path for the policy rate?
  3. What is the effect of a monetary policy shock on GDP and inflation?
  4. What has been the contribution of monetary policy shocks to the behaviour of GDP over time?

What is a Vector Autoregression (VAR)?

- Given, for example, a (3×1) vector of time series $\mathbf{x}_t$ where

$$\mathbf{x}_t = \begin{bmatrix} \Delta y_1 & \Delta y_2 & ... & \Delta y_T \\ \pi_1 & \pi_2 & ... & \pi_T \\ r_1 & r_2 & ... & r_T \end{bmatrix}$$

- A stationary structural VAR of order 1 is

$$\mathbf{A}\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \varepsilon_t \qquad \text{for } t = 1, ..., T$$

  - $\mathbf{A}$ and $\mathbf{B}$ are (3×3) matrices of coefficients
  - ε_t is an (3×1) vector of unobservable zero mean white noise processes

Three different ways of writing the same thing

- There are different ways to represent the VAR(1)

$$\mathbf{A}\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t$$

- For example, we can also write it:
  - In matrix form

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{31} & a_{33} \end{bmatrix} \begin{bmatrix} \Delta y_t \\ \pi_t \\ r_t \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \pi_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{\Delta yt} \\ \varepsilon_{\pi t} \\ \varepsilon_{rt} \end{bmatrix}$$

  - As a system of linear equation

$$\begin{cases} a_{11}\Delta y_t + a_{12}\pi_t + a_{13}r_t = b_{11}\Delta y_{t-1} + b_{12}\pi_{t-1} + b_{13}r_{t-1} + \varepsilon_{\Delta yt} \\ a_{21}\Delta y_t + a_{22}\pi_t + a_{13}r_t = b_{21}\Delta y_{t-1} + b_{22}\pi_{t-1} + b_{23}r_{t-1} + \varepsilon_{\pi t} \\ a_{31}\Delta y_t + a_{32}\pi_t + a_{33}r_t = b_{31}\Delta y_{t-1} + b_{32}\pi_{t-1} + b_{33}r_{t-1} + \varepsilon_{rt} \end{cases}$$

The structural innovations

- We defined ε_t as a "vector of unobservable zero mean white noise processes". What does it mean?
- This simply means that they are serially uncorrelated and independent of each other
- In other words

$$\varepsilon_t = (\varepsilon'_{\Delta y t}, \varepsilon'_{\pi t}, \varepsilon'_{rt})' \sim \mathcal{N}(0, \mathbf{I})$$

or

$$\text{VCV}(\varepsilon_t) = \begin{bmatrix} 1 & 0 & 0 \\ - & 1 & 0 \\ - & - & 1 \end{bmatrix} \quad \text{and} \quad \text{CORR}(\varepsilon_t) = \begin{bmatrix} 1 & 0 & 0 \\ - & 1 & 0 \\ - & - & 1 \end{bmatrix}$$

[Back to basics] What is a variance-covariance matrix?

- The formula for the variance of a univariate time series $x = [x_1, x_2, ..., x_T]$ is

$$\text{VAR} = \sum_{t=0}^{T} \frac{(x_t - \bar{x})^2}{N} = \sum_{t=0}^{T} \frac{(x_t - \bar{x})(x_t - \bar{x})}{N}$$

- If we have a bivariate time series

$$\mathbf{x}_t = \begin{bmatrix} x_1 & x_2 & ... & x_T \\ y_1 & y_2 & ... & y_T \end{bmatrix}$$

the formula becomes

$$\text{VCV} = \begin{bmatrix} \sum_{t=0}^{T} \frac{(x_t-\bar{x})(x_t-\bar{x})}{N} & \sum_{t=0}^{T} \frac{(x_t-\bar{x})(y_t-\bar{y})}{N} \\ \sum_{t=0}^{T} \frac{(y_t-\bar{y})(x_t-\bar{x})}{N} & \sum_{t=0}^{T} \frac{(y_t-\bar{y})(y_t-\bar{y})}{N} \end{bmatrix} = \begin{bmatrix} \text{VAR}(x) & \text{COV}(x,y) \\ \text{COV}(x,y) & \text{VAR}(y) \end{bmatrix}$$

The general form of the stationary structural VAR(p) model

- The basic VAR(1) model may be too poor to sufficiently summarize the main characteristics of the data
  - Deterministic terms (such as time trend or seasonal dummy variables)
  - Exogenous variables (such as the price of oil)
- The general form of the VAR(p) model with deterministic terms ($\mathbf{Z}_t$) and exogenous variables ($\mathbf{W}_t$) is given by

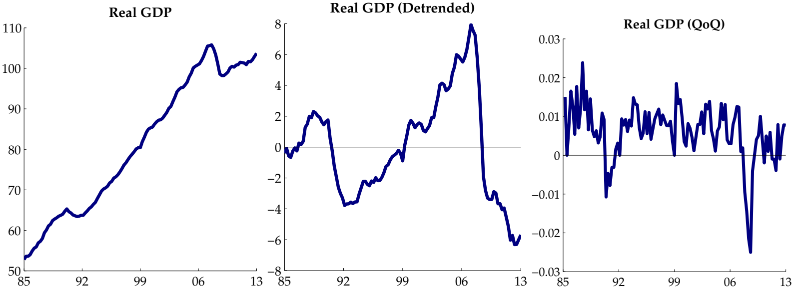
$$\mathbf{A}\mathbf{x}_t = \mathbf{B}_1\mathbf{x}_{t-1} + \mathbf{B}_2\mathbf{x}_{t-2} + ... + \mathbf{B}_p\mathbf{x}_{t-p} + \mathbf{\Lambda}\mathbf{Z}_t + \mathbf{\Psi}\mathbf{W}_t + \varepsilon_t$$

Why is it called structural VAR?

- The equations of a structural VAR define the true structure of the economy
- The fact that $\varepsilon_t = (\varepsilon'_{\Delta yt}, \varepsilon'_{\pi t}, \varepsilon'_{rt})' \sim \mathcal{N}(0, \mathbf{I})$ implies that we can interpret ε_t as structural shocks
- For example we could interpret
  - $\varepsilon_{\Delta yt}$ as an aggregate shock
  - $\varepsilon_{\pi t}$ as a cost-push shock
  - ε_{rt} as a monetary policy shock

Why is it called stationary VAR?

- One of the main assumptions of standard VARs is stationarity of the data
- A stochastic process is said covariance stationary if its first and second moments, $\mathrm{E}(x)$ and $\mathrm{VCV}(x)$ respectively, exist and are constant over time



Structural VARs potentially answers many interesting questions

- For example in our VAR(1)

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \Delta y_t \\ \pi_t \\ r_t \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \pi_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{\Delta y t} \\ \varepsilon_{\pi t} \\ \varepsilon_{rt} \end{bmatrix}$$

  - a_{31} is the impact multiplier of monetary policy shocks on GDP
  - a_{32} is the impact multiplier of monetary policy shocks on inflation
  - If we simulate the model, we can evaluate the time profile of a monetary policy shock on GDP
  - We can add additional variables (and equations) and simulate other shocks: credit supply, oil price, QE,...

However... the estimation of structural VARs is problematic

- The equations of $\mathbf{A}\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \varepsilon_t$ cannot be estimated with OLS because they violate one important assumption $\Longrightarrow$ the regressor cannot be correlated with the error term
  - To see that, take the GDP equation

$$a_{11}\Delta y_t + a_{12}\pi_t + a_{13}r_t = b_{11}\Delta y_{t-1} + b_{12}\pi_{t-1} + b_{13}r_{t-1} + \varepsilon_{\Delta yt}$$

    and compute $\mathrm{COV}[\pi_t, \varepsilon_{yt}]$

- OLS estimation of $\mathbf{A}\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \varepsilon_t$ would produce inconsistent estimates of the parameters, impulse responses, etc

How to solve this problem?

- In the GDP equation

$$a_{11}\Delta y_t + a_{12}\pi_t + a_{13}r_t = b_{11}\Delta y_{t-1} + b_{12}\pi_{t-1} + b_{13}r_{t-1} + \varepsilon_{\Delta yt}$$

  the terms $a_{12}\pi_t$ and $a_{13}r_t$ are the ones generating problems for OLS estimation

- This endogeneity problem disappears if we remove the contemporaneous dependence of Δy_t on the other endogenous variables

- More in general the $\mathbf{A}$ matrix is problematic (since it includes all the contemporaneous relation among the endogenous variables)

- We can solve this problem by simply pre-multiplying the VAR by $\mathbf{A}^{-1}$

$$\mathbf{A}^{-1}\mathbf{A}\mathbf{x}_t = \mathbf{A}^{-1}\mathbf{B}\mathbf{x}_{t-1} + \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t$$
$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t$$
$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{u}_t$$

- That is, we moved the contemporaneous dependence of the endogenous variables (which is given by $\mathbf{A}$) into the "modified" error terms $\mathbf{u}_t$
- This implies that now $\text{CORR}(\mathbf{u}_t) \neq \mathbf{I}$

[Back to basics] **The inverse of a matrix**

- The inverse of a 2×2 matrix

$$\mathbf{X} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \mathbf{X}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

- This implies that $\mathbf{u}_t = \mathbf{A}^{-1} \boldsymbol{\varepsilon}_t$ can be computed as

$$u_{1t} = \frac{a_{22}\varepsilon_{1t} - a_{21}\varepsilon_{2t}}{\Delta}$$
$$u_{2t} = \frac{-a_{21}\varepsilon_{1t} + a_{11}\varepsilon_{2t}}{\Delta}$$

where $\Delta = a_{11}a_{22} - a_{12}a_{21}$

The reduced-form VAR

- This alternative formulation of the VAR

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{u}_t$$

  is called the *reduced-form* representation

- In matrix form

$$\begin{bmatrix} \Delta y_t \\ \pi_t \\ r_t \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \pi_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} u_{\Delta y t} \\ u_{\pi t} \\ u_{it} \end{bmatrix}$$

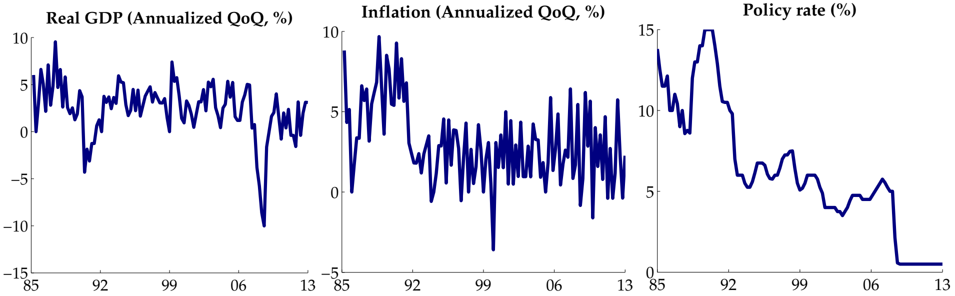
  where

$$\mathbf{u}_t \sim \mathcal{N}(0, \Sigma_{\mathbf{u}}) \quad \text{and} \quad \text{CORR}(\mathbf{u}_t) = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ - & 1 & \rho_{23} \\ - & - & 1 \end{bmatrix}$$

VAR Estimation

We can now estimate the VAR with some real data

- UK quarterly data from 1985.I to 2013.III
- VAR(1) with a constant $\mathbf{x}_t = \mathbf{c} + \mathbf{F}\mathbf{x}_{t-1} + \mathbf{u}_t$



OLS estimation – Typical VAR output

- Matrix of coefficients $\mathbf{F}'$

	Real GDP	GDP Deflator	Policy Rate
c	1.14	1.19	-0.11
Real GDP(-1)	0.61	-0.07	0.06
GDP Deflator(-1)	-0.09	0.02	0.03
Policy Rate(-1)	0.01	0.30	0.96

- Correlation between reduced-form residuals

	Real GDP	GDP Deflator	Policy Rate
Real GDP	1.000	-0.178	0.373
GDP Deflator	–	1.000	0.137
Policy Rate	–	–	1.000

Debunking the typical VAR output

- The constant is not the mean nor the long-run equilibrium of a variable
  - In our example, the mean of the policy rate is not negative

	Real GDP	GDP Deflator	Policy Rate
c	1.14	1.19	-0.11

- The correlation of the residuals reflects the contemporaneous relation between our variables
  - In our example GDP growth and inflation are contemporaneously negatively correlated

	Real GDP	GDP Deflator	Policy Rate
Real GDP	1.000	-0.178	0.373

Model checking & tuning

- We do not cover this in detail but before interpreting the VAR results you should check a number of assumptions
- Loosely speaking we need to check that the reduced-form residuals are
  - Normally distributed
  - Not autocorrelated
  - Not heteroskedastic (i.e., have constant variance)
- ... and that the VAR is stationary (we'll see in a second what it means)

Model checking: why the residuals?

- The VAR believes that

$$\mathbf{x} \sim \mathcal{N}(\mu, \sigma) \implies \begin{cases} \Delta y & \sim & \mathcal{N}(\mu^{\Delta y}, \sigma^{\Delta y}) \\ \pi & \sim & \mathcal{N}(\mu^{\pi}, \sigma^{\pi}) \\ i & \sim & \mathcal{N}(\mu^{i}, \sigma^{i}) \end{cases}$$

- If the data that we feed into the VAR has not these features, the residuals will inherit them
- Note that
  - Mean (μ) and variance (σ) are constant $\implies$ the data have to be stationary!
  - The mean (μ) and variance (σ) are not known a priori
  - At each point in time $\mathbf{x}_t$ does not generally coincide with μ because of (i) shocks hitting at t and (ii) shocks that hit in the past and that are slowly dying out
  - The average of $\mathbf{x}_t$ over a given sample does not necessarily coincide with μ

Can we recover the mean of our endogenous variables?

- Yes. It is given by $\boldsymbol{\mu} = \mathsf{E}[\mathbf{x}_t]$

$$\begin{aligned}
\mathsf{E}[\mathbf{x}_t] &= \mathsf{E}[\mathbf{c} + \mathbf{F}\mathbf{x}_{t-1} + \mathbf{u}_t] = \\
&= \mathsf{E}[\mathbf{c} + \mathbf{F}(\mathbf{c} + \mathbf{F}\mathbf{x}_{t-2} + \mathbf{u}_{t-1}) + \mathbf{u}_t] = \mathbf{E}[\mathbf{c} + \mathbf{F}\mathbf{c} + \mathbf{F}^2\mathbf{x}_{t-2} + \mathbf{F}\mathbf{u}_{t-1} + \mathbf{u}_t] = \\
&= \mathsf{E}[\mathbf{c} + \mathbf{F}\mathbf{c} + \mathbf{F}^2\mathbf{c} + \ldots + \mathbf{F}^{t-2}\mathbf{c} + \mathbf{F}^{t-1}\mathbf{x}_1 + \mathbf{F}^{t-2}\mathbf{u}_2 + \ldots + \mathbf{F}^2\mathbf{u}_{t-2} + \mathbf{F}\mathbf{u}_{t-1} + \mathbf{u}_t] = \\
&= \ldots = \\
&= \mathsf{E}\left[\sum_{j=0}^{t-2} \mathbf{F}^j\mathbf{c} + \mathbf{F}^{t-1}\mathbf{x}_1 + \sum_{j=0}^{t-2} \mathbf{F}^j\mathbf{u}_{t-j}\right]
\end{aligned}$$

- If VAR is stationary, from the last expression we get

$$\mathsf{E}[\mathbf{x}_t] = (\mathbf{I} - \mathbf{F})^{-1}\mathbf{c} = \boldsymbol{\mu}$$

[Back to basics] **Geometric series**

- Consider the following sum $\sum_{j=0}^{T} \mathbf{Y}^j$
- When $T \rightarrow \infty$ we have:

$$(\mathbf{1} + \mathbf{Y} + \mathbf{Y}^2 + ... + \mathbf{Y}^\infty) = (\mathbf{I} - \mathbf{Y})^{-1}$$

  if and only if:
  - $|eig(\mathbf{Y})| < 1$ when $\mathbf{Y}$ is a matrix
  - $\mathbf{Y} < 1$ when $\mathbf{Y}$ is a number
- Therefore, for T large enough we have

$$\mathsf{E}\left[\sum_{j=0}^{t-2} \mathbf{F}^j \mathbf{c} + \mathbf{F}^{t-1}\mathbf{x}_0 + \sum_{j=0}^{t-2} \mathbf{F}^j \mathbf{u}_{t-j}\right] = \mathsf{E}\left[(\mathbf{I} - \mathbf{F})^{-1}\mathbf{c}\right] + \underbrace{\mathbf{F}^{t-1}\mathbf{x}_1}_{\approx 0} + \underbrace{\mathsf{E}\left[\sum_{j=0}^{t-2} \mathbf{F}^j \mathbf{u}_{t-j}\right]}_{\approx 0}$$

Stationary VAR, stationary data

- A VAR is stationary (or stable) when

$$|eig(\mathbf{F})| < 1$$

- In that case, the VAR thinks that $(\mathbf{I} - \mathbf{F})^{-1}\mathbf{c}$ is the unconditional mean of the stochastic processes governing our variables

- VAR models stationary data
  - GDP level clearly does not have a well defined mean
  - GDP growth does!

Unconditional mean in practice

- The unconditional mean is an interesting element of VAR analysis but it is often ignored
- From the estimated VAR, recover both $\mathbf{c}$ and $\mathbf{F}$
- Compute the unconditional mean as

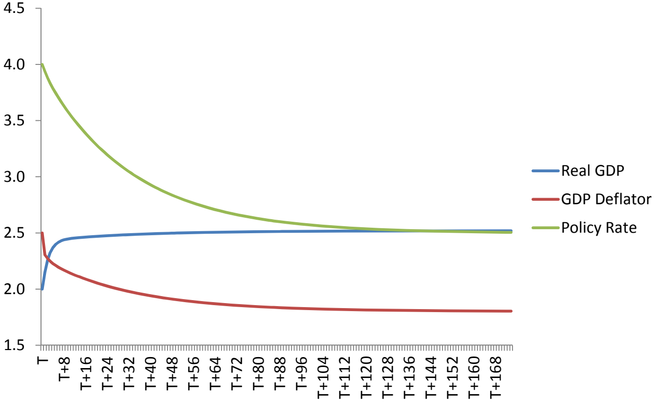
$$(\mathbf{I}-\mathbf{F})^{-1}\mathbf{c} = \begin{bmatrix} 2.32 & -0.28 & -1.89 \\ 1.34 & 1.24 & 10.58 \\ 4.89 & 0.67 & 33.87 \end{bmatrix} \begin{bmatrix} 1.14 \\ 1.19 \\ -0.11 \end{bmatrix} = \begin{bmatrix} 2.51 \\ 1.80 \\ 2.49 \end{bmatrix} = \begin{bmatrix} \mu^{\Delta y} \\ \mu^{\pi} \\ \mu^{i} \end{bmatrix}$$

Why stationarity of the data is important

- In absence of shocks, each variable will converge to its unconditional mean

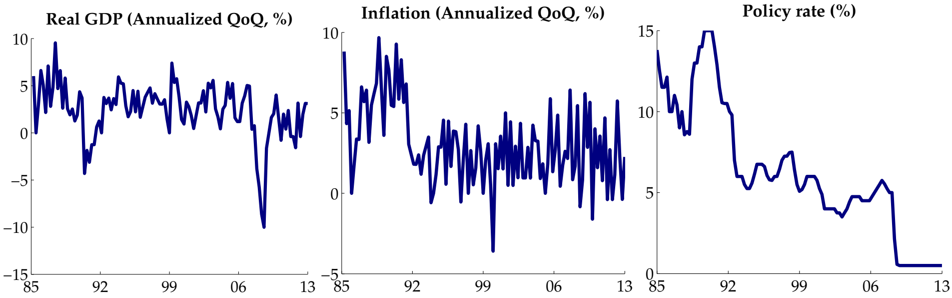
For example, start from a point in time T where:

$$y_T = 2\%$$
$$\pi_T = 2\%$$
$$r_T = 4\%$$



In light of these considerations: is our data sensible?

- UK quarterly data from 1985.I to 2013.III



Some VAR limitations

- VARs are linear models of stationary data
- ... but often macro data
  - is non-linear (crisis periods)
  - is non-stationary (trends, breaks, etc)
  - displays time-varying variance (Great Moderation Vs Great Recession)
- All these elements have to be taken into account when analyzing the output from VAR analysis (especially when estimated over the sample period we use these days which features all the above)

Back to our estimated VAR

	Real GDP	GDP Deflator	Policy Rate
c	1.14	1.19	-0.11
Real GDP(-1)	0.61	-0.07	0.06
GDP Deflator(-1)	-0.09	0.02	0.03
Policy Rate(-1)	0.01	0.30	0.96

- What can we do with it?
  - What are the dynamic properties of these variables? [Look at lagged coefficients]
  - How do these variables interact? [Look at cross-variable coefficients]
  - What will be inflation tomorrow? [???]

Forecasting

- Forecasting is one of the main objectives of multivariate time series analysis
- The best linear predictor (in terms of minimum mean squared error) of $\mathbf{x}_{T+1}$ based on information available at time T (today) is

$$\mathbf{x}^{\mathbf{f}}_{T+1} = \mathbf{F}\mathbf{x}_T$$

- For example, if today (T) we have

$$\mathbf{x}_T \begin{bmatrix} \Delta y_T = 2\% \\ \pi_T = 2\% \\ r_T = 4\% \end{bmatrix} \quad \Longrightarrow \quad \mathbf{x}^{\mathbf{f}}_{T+1} = \mathbf{F}\mathbf{x}_T \begin{bmatrix} \Delta y_{T+1} = 2.15\% \\ \pi_{T+1} = 2.31\% \\ r_{T+1} = 3.94\% \end{bmatrix}$$

- Note that we can also construct conditional forecasts by specifying an exogenous path for a variable (the policy rate for example)

The Identification Problem

Back to our estimated VAR

	Real GDP	GDP Deflator	Policy Rate
c	1.14	1.19	-0.11
Real GDP(-1)	0.61	-0.07	0.06
GDP Deflator(-1)	-0.09	0.02	0.03
Policy Rate(-1)	0.01	0.30	0.96

- What can we do with it?
  - What are the dynamic properties of these variables? [Look at lagged coefficients]
  - How do these variables interact? [Look at cross-variable coefficients]
  - What will be inflation tomorrow [Forecasting]
  - What is the effect of a monetary policy shock on GDP and inflation? [???]

Reduced-form VARs do not tell us anything about the structure of the economy

- We cannot interpret the reduced-form error terms ($\mathbf{u}$) as structural shocks
  - How do we interpret a movement in u_r? Since it is a linear combination of ε_r, $\varepsilon_{\Delta y}$, and ε_π it is hard to know what is the nature of the shock
- Is it a shock to aggregate demand that induces policy to move the interest rate? Or is it a monetary policy shock?
  - This is the very nature of the identification problem
- To answer this question we need to get back to the structural representation (where the error terms are uncorrelated)

From the reduced-form back to the structural form

- We normally estimate

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{u}_t \quad \text{and} \quad \Sigma_{\mathbf{u}}$$

- But our ultimate objective is to recover

$$\mathbf{A}\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \varepsilon_t$$

- Does not sound too difficult... We know that

$$\mathbf{F} = \mathbf{A}^{-1}\mathbf{B}$$
$$\mathbf{u}_t = \mathbf{A}^{-1}\varepsilon_t$$

- We also know that

$$\Sigma_{\mathbf{u}} = \mathsf{E}\left[\mathbf{u}_t\mathbf{u}_t'\right] = \mathsf{E}\left[\mathbf{A}^{-1}\varepsilon\left(\mathbf{A}^{-1}\varepsilon\right)'\right] = \mathbf{A}^{-1}\Sigma_{\varepsilon}\left(\mathbf{A}^{-1}\right)' = \mathbf{A}^{-1}\mathbf{A}^{-1\prime}$$

  since $\Sigma_{\varepsilon} = \mathbf{I}$

- In other words if we pin down $\mathbf{A}^{-1}$ we are done, since we can recover
  - $\varepsilon_t = \mathbf{A}\mathbf{u}_t$
  - $\mathbf{B} = \mathbf{A}\mathbf{F}$

  and we would know the structural representation of the economy

- You can think of the idnetified $\mathbf{A}\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \varepsilon_t$ model as the 3 equation New Keynesian model

The variance–covariance matrix of residuals

- Why $\Sigma_{\mathbf{u}} = \mathsf{E}\left[\mathbf{u}_t \mathbf{u}_t'\right]$?
- The formula for the variance of a univariate time series is $x = \left[x_0, x_1, ..., x_T\right]$

$$\mathsf{VAR} = \sum_{t=0}^{T} \frac{(x_t - \bar{x})^2}{N}$$

- But the residuals (both the $\mathbf{u}_t$ and the ε_t) have zero mean which implies that formula would be

$$\mathsf{VAR} = \sum_{t=0}^{T} \frac{x_t^2}{N}$$

The variance–covariance matrix of residuals

- In a bivariate VAR we would have

$$\mathbf{u}_t\mathbf{u}_t' = \begin{bmatrix} u_1^1 & u_2^1 & \dots & u_T^1 \\ u_1^2 & u_2^2 & \dots & u_T^2 \end{bmatrix} \begin{bmatrix} u_1^1 & u_1^2 \\ u_2^1 & u_2^2 \\ \dots & \dots \\ u_T^1 & u_T^2 \end{bmatrix} = \begin{bmatrix} \sum_{t=0}^{T}\left(u_t^1\right)^2 & \sum_{t=0}^{T}\left(u_t^1 u_t^2\right) \\ - & \sum_{t=0}^{T}\left(u_t^2\right)^2 \end{bmatrix}$$

- And therefore

$$\mathsf{E}\left[\mathbf{u}_t\mathbf{u}_t'\right] = \begin{bmatrix} \mathsf{VAR}\left[u^1\right] & \mathsf{COV}\left[u^1u^2\right] \\ - & \mathsf{VAR}\left[u^1\right] \end{bmatrix}$$

The "identification problem"

- Identification problem boils down to pinning down $\mathbf{A}^{-1}$
- If we write $\Sigma_{\mathbf{u}} = \mathbf{A}^{-1}\mathbf{A}^{-1\prime}$ in matrices

$$\underbrace{\begin{bmatrix} \sigma^2_{u1} & \sigma^2_{u1,u2} & \sigma^2_{u1,u3} \\ - & \sigma^2_{u2} & \sigma^2_{u2,u3} \\ - & - & \sigma^2_{u3} \end{bmatrix}}^{\Sigma_{\mathbf{u}}} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^{-1} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^{-1\prime}}^{\mathbf{A}^{-1}\mathbf{A}^{-1\prime}}$$

  we can derive a system of equations
- However, there are 9 unknowns (the elements of $\mathbf{A}^{-1}\mathbf{A}^{-1\prime}$) but only 6 equations (because the variance-covariance matrix is symmetric)
  - The system is not identified!

Common Identification Schemes

Common identification schemes

- Zero short-run restrictions (also known as recursive, Cholesky, orthogonal)
- Zero long-run restrictions (also known as Blanchard-Quah)
- Theory-based restrictions
- Sign restrictions

Zero short-run restrictions

- Assume that A is lower triangular

$$\begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \Delta y_t \\ \pi_t \\ r_t \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \pi_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{\Delta yt} \\ \varepsilon_{\pi t} \\ \varepsilon_{rt} \end{bmatrix}$$

- Or in other words that
  - $\varepsilon_{\Delta yt}$ affects contemporaneously all variables, namely Δy_t, π_t and r_t
  - $\varepsilon_{\pi t}$ affects contemporaneously only π_t and r_t, but no Δy_t
  - ε_{rt} affects contemporaneously only r_t
- We now have 6 unknowns and 6 equations!

- To see what are the implications of our assumptions on $\mathbf{A}$, first remember that the inverse of a lower triangular matrix is also lower triangular

$$\begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^{-1} = \begin{bmatrix} \tilde{a}_{11} & 0 & 0 \\ \tilde{a}_{21} & \tilde{a}_{22} & 0 \\ \tilde{a}_{31} & \tilde{a}_{32} & \tilde{a}_{33} \end{bmatrix}$$

- Now pre-multiply the VAR by $\mathbf{A}^{-1}$

$$\begin{bmatrix} \Delta y_t \\ \pi_t \\ r_t \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \pi_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} \tilde{a}_{11} & 0 & 0 \\ \tilde{a}_{21} & \tilde{a}_{22} & 0 \\ \tilde{a}_{31} & \tilde{a}_{32} & \tilde{a}_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{\Delta yt} \\ \varepsilon_{\pi t} \\ \varepsilon_{rt} \end{bmatrix}$$

- Which implies that

$$\begin{cases} \Delta y_t & = ... + \tilde{a}_{11}\varepsilon_{\Delta yt} \\ \pi_t & = ... + \tilde{a}_{21}\varepsilon_{\Delta yt} + \tilde{a}_{22}\varepsilon_{\pi t} \\ r_t & = ... + \tilde{a}_{31}\varepsilon_{\Delta yt} + \tilde{a}_{32}\varepsilon_{\pi t} + \tilde{a}_{33}\varepsilon_{rt} \end{cases}$$

- We normally implement this identification scheme *via* a Cholesky decomposition of $\Sigma_{\mathbf{u}}$

$$\Sigma_{\mathbf{u}} = \mathbf{P}'\mathbf{P}$$

where $\mathbf{P}'$ is lower triangular

- Note that

$$\Sigma_{\mathbf{u}} = \mathbf{P}'\mathbf{P} \qquad \text{but also} \qquad \Sigma_{\mathbf{u}} = \mathbf{A}^{-1}\mathbf{A}^{-1\prime}$$

- and that

$$\mathbf{A} \text{ is lower triangular}$$

- Then it musty follow that $\mathbf{P}' = \mathbf{A}^{-1} \implies$ Identification!

Cholesky decomposition of a matrix

- Don't be scared of Cholesky decomposition! It's a kind of square root of a matrix
  - As in Excel you type `sqrt()` in Matlab you type `chol()`
- A symmetric and positive definite matrix $\mathbf{X}$ can be decomposed as:

$$\mathbf{X} = \mathbf{P}'\mathbf{P}$$

  where $\mathbf{P}$ is an upper triangular matrix (and therefore $\mathbf{P}'$ is lower triangular)
- The formula is

$$\mathbf{X} = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \quad \mathbf{P} = \begin{bmatrix} \sqrt{a} & \frac{b}{\sqrt{a}} \\ 0 & \sqrt{c - \frac{b^2}{a}} \end{bmatrix}$$

Zero long-run restrictions

- Re-write the VAR as

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{A}^{-1}\varepsilon_t$$

- If a shock hits in t, its cumulative (long run) impact on $\mathbf{x}_t$ would be

$$\mathbf{x}_{t,t+\infty} = \mathbf{A}^{-1}\varepsilon_t + \mathbf{F}\mathbf{A}^{-1}\varepsilon_t + \mathbf{F}^2\mathbf{A}^{-1}\varepsilon_t + \ldots + \mathbf{F}^{\infty}\mathbf{A}^{-1}\varepsilon_t$$

- We can rewrite

$$\mathbf{x}_{t,t+\infty} = \sum_{j=0}^{\infty} \mathbf{F}^j\mathbf{A}^{-1}\varepsilon_t = (\mathbf{I} - \mathbf{F})^{-1}\mathbf{A}^{-1}\varepsilon_t = \mathbf{D}\varepsilon_t$$

where $\mathbf{D}$ is the cumulative effect of the shock ε_t from time t to ∞

- What is the intuition for $\mathbf{D}$?

$$\begin{bmatrix} \Delta y_{t,t+\infty} \\ \pi_{t,t+\infty} \\ i_{t,t+\infty} \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{\pi t} \\ \varepsilon_{rt} \end{bmatrix}$$

- Take the first equation: $\Delta y_{t,t+\infty} = d_{11}\varepsilon_{yt} + d_{12}\varepsilon_{\pi t} + d_{13}\varepsilon_{rt}$
  - d_{13} represents the cumulative impact of a monetary policy shock (hitting in t) on the level GDP in the long-run
  - If you believe in the neutrality of money you would expect $d_{13} = 0$

- To achieve identification note that

$$\mathbf{D}\mathbf{D}' = (\mathbf{I}-\mathbf{F})^{-1}\mathbf{A}^{-1}\mathbf{A}^{-1\prime}(\mathbf{I}-\mathbf{F})^{-1\prime} = (\mathbf{I}-\mathbf{F})^{-1}\Sigma_{\mathbf{u}}(\mathbf{I}-\mathbf{F})^{-1\prime}$$

- Note also that
    - The right-hand side of the above equation is known
    - Both $\mathbf{D}\mathbf{D}'$ and $(\mathbf{I}-\mathbf{F})^{-1}\Sigma_{\mathbf{u}}(\mathbf{I}-\mathbf{F})^{-1\prime}$ are symmetric matrices
    - There exists an upper triangular matrix $\mathbf{P}$ such that $\mathbf{P}'\mathbf{P} = (\mathbf{I}-\mathbf{F})^{-1}\Sigma_{\mathbf{u}}(\mathbf{I}-\mathbf{F})^{-1\prime}$
- Therefore, if we assume that $\mathbf{D}$ is lower triangular, it must be that $\mathbf{D} = \mathbf{P}'$
- Finally $\mathbf{A}^{-1} = (\mathbf{I}-\mathbf{F})\,\mathbf{D} \Longrightarrow$ Identification!

Sign restrictions

- In the zero short-run restriction identification we used the fact that

$$\boldsymbol{\Sigma}_{\mathbf{u}} = \mathbf{A}^{-1}\mathbf{A}^{-1\prime} \quad \text{and} \quad \boldsymbol{\Sigma}_{\mathbf{u}} = \mathbf{P}'\mathbf{P}$$

  where the lower triangular $\mathbf{P}'$ matrix is the Cholesky decomposition of $\boldsymbol{\Sigma}_{\mathbf{u}}$

- For a given random orthonormal matrix (i.e., such that $\mathbf{S}'\mathbf{S} = \mathbf{I}$) we have that

$$\boldsymbol{\Sigma}_{\mathbf{u}} = \mathbf{A}^{-1}\mathbf{A}^{-1\prime} = \mathbf{P}'\mathbf{S}'\mathbf{S}\mathbf{P} = \mathcal{P}'\mathcal{P}$$

  where $\mathcal{P}'$ is generally not lower triangular anymore

- $\mathbf{A}^{-1} = \mathcal{P}'$ is clearly a valid solution to the identification problem
- But $\mathbf{S}'$ is a random matrix... is the solution $\mathbf{A}^{-1} = \mathcal{P}'$ plausible?
- Identification is achieved by checking whether the impulse responses implied by $\mathbf{S}'$ satisfy a set of a priori (and possibly theory-driven) sign restrictions
- We can draw as many $\mathbf{S}'$ as we want and construct a distribution of the solutions that satisfy the sign restrictions

Sign restriction in steps

1. Draw a random orthonormal matrix $\mathbf{S}'$
2. Compute $\mathbf{A}^{-1} = \mathbf{P}'\mathbf{S}'$ where $\mathbf{P}'$ is the Cholesky decomposition of the reduced form residuals $\Sigma_{\mathbf{u}}$
3. Compute the impulse response associated with $\mathbf{A}^{-1}$
4. Are the sign restrictions satisfied?
   4.1 Yes. Store the impulse response
   4.2 No. Discard the impulse response
5. Perform N replications and report the median impulse response (and its confidence intervals)

Structural Dynamic Analysis

Why do we need identification?

- According to Stock & Watson's list so far we have done
  1. Describe and summarize macroeconomic time series
  2. Make forecasts
  3. Recover the true structure of the macroeconomy from the data
- How about
  1. Advise macroeconomic policymakers
- (A good) identification allows us to address the last point
- This is normally done by means of Impulse responses, Forecast error variance decompositions, and Historical decompositions

Impulse response functions

- Impulse response functions ($\mathcal{IR}$) answer the question
  - What is the response of current and future values of each of the variables to a one-unit increase in the current value of one of the structural errors, assuming that this error returns to zero in subsequent periods and that all other errors are equal to zero
- The implied thought experiment of changing one error while holding the others constant makes sense only when the errors are uncorrelated across equations

How to compute impulse response functions

- As an example, we compute the $\mathcal{IR}$ for a bivariate VAR $\mathbf{x}_t = (x'_{1t}, x'_{2t})$
- Define a vector of exogenous impulses $(\mathbf{s}_\tau)$ that we want to to impose to the structural errors of the system

Time (τ)	1	2	h
Impulse to ε_1 $(s_{1,\tau})$	$s_{1,1} = 1$	$s_{1,2} = 0$	$s_{1,h} = 0$
Impulse to ε_2 $(s_{2,\tau})$	$s_{2,1} = 0$	$s_{2,2} = 0$	$s_{2,h} = 0$

- We can use the following hybrid representation to compute the $\mathcal{IR}$

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{A}^{-1}\mathbf{s}_t,$$

- The impulse response $\mathcal{IR}_\tau$ is given by

$$\begin{cases} \mathcal{IR}_1 = \mathbf{A}^{-1}\mathbf{s}_1, \\ \mathcal{IR}_\tau = \mathbf{F} \cdot \mathcal{IR}_{\tau-1}, \qquad \text{for } \tau = 2, ..., h. \end{cases}$$

Forecast error variance decompositions

- Forecast error variance decompositions ($\mathcal{VD}$) answer the question
  - What portion of the variance of the forecast error in predicting $x_{i,T+h}$ is due to the structural shock ε_i?
- Provide information about the relative importance of each structural shock in affecting the variables in the VAR

How to compute forecast error variance decompositions

- As an example, we compute the $\mathcal{VD}$ for the 1-step ahead forecast error in a bivariate VAR $\mathbf{x}_t = (x'_{1t}, x'_{2t})$
- First, let's define the 1-step ahead forecast error

$$\xi_{T+1} = \mathbf{x}_{T+1} - \mathbf{x}^{\mathbf{f}}_{T+1} = \mathbf{x}_{T+1} - \mathbf{F}\mathbf{x}_T$$

- It follows that

$$\xi_{T+1} = \mathbf{u}_{T+1} = \mathbf{A}^{-1}\varepsilon_{T+1}$$

Forecast error is the yet unobserved realization of the shocks

- In a bivariate VAR we have

$$\begin{bmatrix} \xi_{1,T+1} \\ \xi_{2,T+1} \end{bmatrix} = \begin{bmatrix} \tilde{a}_{11} & \tilde{a}_{12} \\ \tilde{a}_{21} & \tilde{a}_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,T+1} \\ \varepsilon_{2,T+1} \end{bmatrix}$$

where $\tilde{a}$ are the elements of $\mathbf{A}^{-1}$

- Or

$$\begin{cases} \xi_{1,T+1} = \tilde{a}_{11}\varepsilon_{1,T+1} + \tilde{a}_{12}\varepsilon_{2,T+1} \\ \xi_{2,T+1} = \tilde{a}_{21}\varepsilon_{1,T+1} + \tilde{a}_{22}\varepsilon_{2,T+1} \end{cases}$$

- What is the variance of the forecast error?

$$\text{VAR}\left(\xi_{1,T+1}\right) = \tilde{a}_{11}^2 \text{VAR}\left(\varepsilon_{1,T+1}\right) + \tilde{a}_{12}^2 \text{VAR}\left(\varepsilon_{2,T+1}\right) = \tilde{a}_{11}^2 + \tilde{a}_{12}^2$$
$$\text{VAR}\left(\xi_{2,T+1}\right) = \tilde{a}_{21}^2 \text{VAR}\left(\varepsilon_{1,T+1}\right) + \tilde{a}_{22}^2 \text{VAR}\left(\varepsilon_{2,T+1}\right) = \tilde{a}_{21}^2 + \tilde{a}_{22}^2$$

Basic properties of the variance

- If X is a random variable $\mathbf{X}$ and a is a constant
  - $\text{VAR}(\mathbf{X}+a) = \text{VAR}(\mathbf{X})$
  - $\text{VAR}(a\mathbf{X}) = a^2\text{VAR}(\mathbf{X})$
- If $\mathbf{Y}$ is a random variable and b is a constant
  - $\text{VAR}(a\mathbf{X}+b\mathbf{Y}) = a^2\text{VAR}(\mathbf{X}) + b^2\text{VAR}(\mathbf{Y}) + 2ab\text{COV}(\mathbf{X},\mathbf{Y})$
- Since the structural errors are independent, it follows that $\text{COV}(\varepsilon_1,\varepsilon_2) = 0$

- Therefore, we just showed that variance of the 1-step ahead forecast error is

$$\text{VAR}\left(\xi_{1,T+1}\right) = \tilde{a}_{11}^2 + \tilde{a}_{12}^2$$
$$\text{VAR}\left(\xi_{2,T+1}\right) = \tilde{a}_{21}^2 + \tilde{a}_{22}^2$$

- Which portion of the variance is due to each structural error?

$$\underbrace{\begin{cases} \mathcal{VD}_{x_1}^{\varepsilon_1} = \frac{\tilde{a}_{11}^2}{\tilde{a}_{11}^2+\tilde{a}_{12}^2} \\ \mathcal{VD}_{x_1}^{\varepsilon_2} = \frac{\tilde{a}_{12}^2}{\tilde{a}_{11}^2+\tilde{a}_{12}^2} \end{cases}}_{\text{This sums up to 1}} \qquad \underbrace{\begin{cases} \mathcal{VD}_{x_2}^{\varepsilon_1} = \frac{\tilde{a}_{21}^2}{\tilde{a}_{21}^2+\tilde{a}_{22}^2} \\ \mathcal{VD}_{x_2}^{\varepsilon_2} = \frac{\tilde{a}_{22}^2}{\tilde{a}_{21}^2+\tilde{a}_{22}^2} \end{cases}}_{\text{This sums up to 1}}$$

Historical decompositions

- Historical decompositions ($\mathcal{HD}$) answer the question
  - What portion of the deviation of $x_{i,t}$ from its unconditional mean is due to the structural shock ε_i?
- We showed before that each observation of a variable does not generally coincide with its unconditional mean
- This is because, in each period, the structural shocks realize and push all variables away from their equilibrium values

[Back to basics] **Wold representation of a VAR**

- Each observation of our original data can be re-written as the cumulative sum of the structural shocks

$$
\begin{aligned}
\mathbf{x}_2 &= \mathbf{F}\mathbf{x}_1 + \mathbf{A}^{-1}\varepsilon_2 \\
\mathbf{x}_3 &= \mathbf{F}\mathbf{x}_2 + \mathbf{A}^{-1}\varepsilon_3 = \mathbf{F}(\mathbf{F}\mathbf{x}_1 + \mathbf{A}^{-1}\varepsilon_2) + \mathbf{A}^{-1}\varepsilon_3 = \mathbf{F}^2\mathbf{x}_1 + \mathbf{F}\mathbf{A}^{-1}\varepsilon_2 + \mathbf{A}^{-1}\varepsilon_3 \\
&= \ldots = \\
\mathbf{x}_T &= \mathbf{F}^{T-1}\mathbf{x}_1 + (\mathbf{F}^{T-2}\mathbf{A}^{-1}\varepsilon_2 + \ldots + \mathbf{F}\mathbf{A}^{-1}\varepsilon_{T-1} + \mathbf{A}^{-1}\varepsilon_T)
\end{aligned}
$$

- Or, more in general

$$
\mathbf{x}_t = \mathbf{F}^{t-1}\mathbf{x}_1 + \sum_{j=0}^{t-2} \mathbf{F}^j\mathbf{A}^{-1}\varepsilon_{t-j} \qquad \text{for } t > 1
$$

How to compute historical decompositions

- As an example, we compute the $\mathcal{HD}$ of the third observation in a bivariate VAR $\mathbf{x}_t = \left(x'_{1t}, x'_{2t}\right)'$
- First write $\mathbf{x}_3$ as a function of past errors (ε_2 and ε_3) and the initial conditions ($\mathbf{x}_1$)

$$\mathbf{x}_3 = \underbrace{\mathbf{F}^2\mathbf{x}_1}_{init_3} + \underbrace{\mathbf{FA}^{-1}}_{\mathbf{\Theta_1}}\varepsilon_2 + \underbrace{\mathbf{A}^{-1}}_{\mathbf{\Theta_0}}\varepsilon_3$$

- Re-write $\mathbf{x}_3$ in matrices

$$\begin{bmatrix} x_{1,3} \\ x_{2,3} \end{bmatrix} = \begin{bmatrix} init_{1,3} \\ init_{2,3} \end{bmatrix} + \begin{bmatrix} \theta^1_{11} & \theta^1_{12} \\ \theta^1_{21} & \theta^1_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,2} \\ \varepsilon_{2,2} \end{bmatrix} + \begin{bmatrix} \theta^0_{11} & \theta^0_{12} \\ \theta^0_{21} & \theta^0_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,3} \\ \varepsilon_{2,3} \end{bmatrix}$$

- Therefore $\mathbf{x}_3$ can be expressed as

$$\begin{cases} \mathbf{x}_{1,3} = init_{1,3} + \theta_{11}^{1}\varepsilon_{1,2} + \theta_{12}^{1}\varepsilon_{2,2} + \theta_{11}^{0}\varepsilon_{1,3} + \theta_{12}^{0}\varepsilon_{2,3} \\ \mathbf{x}_{2,3} = init_{2,3} + \theta_{21}^{1}\varepsilon_{1,2} + \theta_{22}^{1}\varepsilon_{2,2} + \theta_{21}^{0}\varepsilon_{1,3} + \theta_{22}^{0}\varepsilon_{2,3} \end{cases}$$

- The historical decomposition is given by

$$\underbrace{\begin{cases} \mathcal{HD}_{1,3}^{\varepsilon_1} = \theta_{11}^{1}\varepsilon_{1,2} + \theta_{11}^{0}\varepsilon_{1,3} \\ \mathcal{HD}_{1,3}^{\varepsilon_2} = \theta_{12}^{1}\varepsilon_{2,2} + \theta_{12}^{0}\varepsilon_{2,3} \\ \mathcal{HD}_{1,3}^{init} = init_{1,3} \end{cases}}_{\text{This sums up to } \mathbf{x}_{1,3}} \qquad \underbrace{\begin{cases} \mathcal{HD}_{2,3}^{\varepsilon_1} = \theta_{21}^{1}\varepsilon_{1,2} + \theta_{21}^{0}\varepsilon_{1,3} \\ \mathcal{HD}_{2,3}^{\varepsilon_2} = \theta_{22}^{1}\varepsilon_{2,2} + \theta_{22}^{0}\varepsilon_{2,3} \\ \mathcal{HD}_{2,3}^{init} = init_{2,3} \end{cases}}_{\text{This sums up to } \mathbf{x}_{2,3}}$$

“Famous VAR Examples”

Examples of different identification schemes

- Zero short-run restrictions
  - Stock & Watson (2001). "Vector Autoregressions," Journal of Economic Perspectives
- Zero long-run restrictions
  - Blanchard & Quah (1989). "The Dynamic Effects of Aggregate Demand and Supply Disturbances", American Economic Review
- Sign Restrictions
  - Uhlig (2005). "What are the effects of monetary policy on output? Results from an agnostic identification procedure," Journal of Monetary Economics

Example: zero short-run restrictions

- Stock & Watson (2001). "Vector Autoregressions," Journal of Economic Perspectives
- US quarterly data from 1960.I to 2000.IV

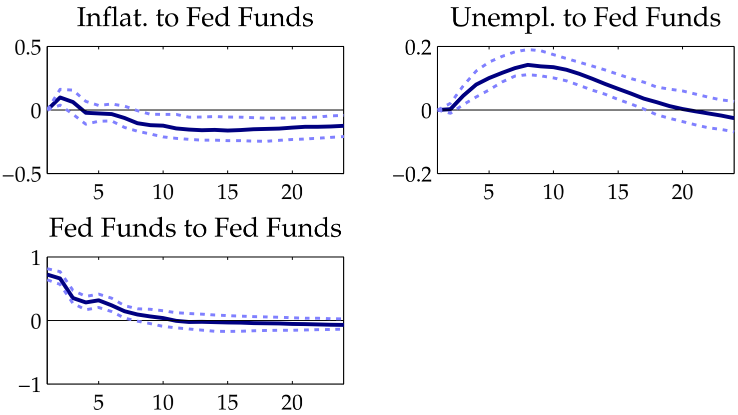
Monetary policy shocks, inflation and unemployment

- Objective: infer the causal influence of monetary policy on unemployment, inflation and interest rates
- Assumptions
  - MP (r_t) reacts contemporaneously to movements in inflation and in unemployment
  - MP shocks (ε_{rt}) do not affect inflation and unemployment within the quarter of the shock

$$\begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \pi_t \\ ur_t \\ r_t \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ ur_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{\pi t} \\ \varepsilon_{urt} \\ \varepsilon_{rt} \end{bmatrix}$$

- Do these assumptions make sense?

The effect of a monetary policy shock



The other two shocks are identified by definition... but how can we interpret them?

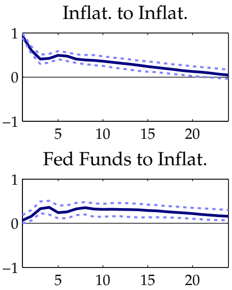
- How about ε_{π} and ε_{ur}? They represent an aggregate supply and a demand equation...
  - The shock to ε_{π} affects all variables contemporaneously
  - The shock to ε_{ur} affects r_t contemporaneously but not π_t

- Do these assumptions make sense?

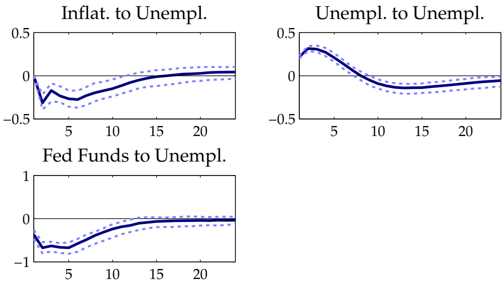
- [Some shocks may be better identified than others]

Aggregate demand and aggregate supply shocks

- Shock to $\varepsilon_{\pi t}$ behaves as a negative aggregate supply shock



- Shock to ε_{ut} behaves as a negative aggregate demand shock

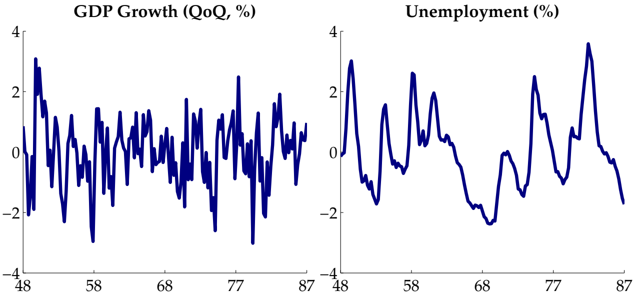


Forecast error variance decomposition

	Inflation			Unemployment			Fed Funds		
	ε_{π}	ε_{ur}	ε_{r}	ε_{π}	ε_{ur}	ε_{r}	ε_{π}	ε_{ur}	ε_{r}
t = 1	1.00	0.00	0.00	0.00	1.00	0.00	0.02	0.20	0.79
t = 4	0.88	0.10	0.01	0.02	0.96	0.02	0.09	0.51	0.41
t = 8	0.83	0.16	0.01	0.10	0.76	0.13	0.11	0.60	0.29
t = 12	0.83	0.15	0.02	0.21	0.60	0.19	0.15	0.59	0.26

Example: zero long-run restrictions

- Blanchard & Quah (1989). "The Dynamic Effects of Aggregate Demand and Supply Disturbances", American Economic Review
- US quarterly data from 1948.I to 1987.IV



Economic theory and the long-run

- Economic theory usually tells us a lot more about what will happen in the long-run, rather than exactly what will happen today
  - Demand-side shocks have no long-run effect on output, while supply-side shocks do
  - Monetary policy shocks have no long-run effect on output
  - ...
- This suggests an alternative approach: to use these theoretically-inspired long-run restrictions to identify shocks and impulse responses

Blanchard & Quah's identification assumptions

- There are two types of disturbances affecting unemployment and output
- The first has no long-run effect on either unemployment or output level
- The second has no long-run effect on unemployment, but may have a long-run effect on output level
- Blanchard & Quah refer to the first as demand disturbances, and to the second as supply disturbances (traditional Keynesian view of fluctuations)

Identification

- We showed that the long-run cumulative impact of a structural shocks is

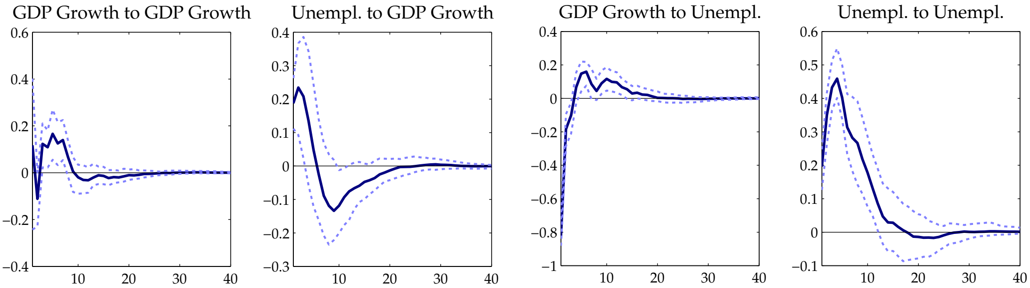
$$\mathcal{IR}_{\tau,\tau+\infty} = \sum_{j=0}^{\infty} \mathbf{F}^j \mathbf{A}^{-1} \varepsilon_t = (\mathbf{I} - \mathbf{F})^{-1} \mathbf{A}^{-1} \varepsilon_t = \mathbf{D} \varepsilon_t$$

- Assume that $\varepsilon_t^{\Delta y}$ is the supply shock and that ε_t^{ur} is the demand shock
- We can rewrite the VAR such that the cumulated effect of ε_t^{ur} on Δy_t is equal to zero by assuming

$$\begin{bmatrix} \Delta y_t \\ ur_t \end{bmatrix} = \begin{bmatrix} d_{11} & 0 \\ d_{21} & d_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_t^{\Delta y} \\ \varepsilon_t^{ur} \end{bmatrix}$$

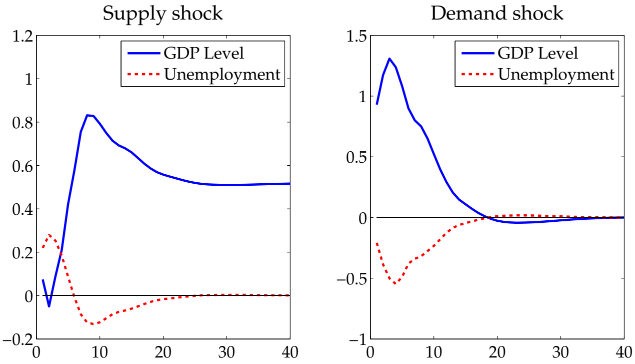
Aggregate demand and supply shocks

- Aggregate supply shock initially increases unemployment (puzzle of hours to producticvity shocks)
- Aggregate demand shocks have a hump-shaped effect on output and unemployment



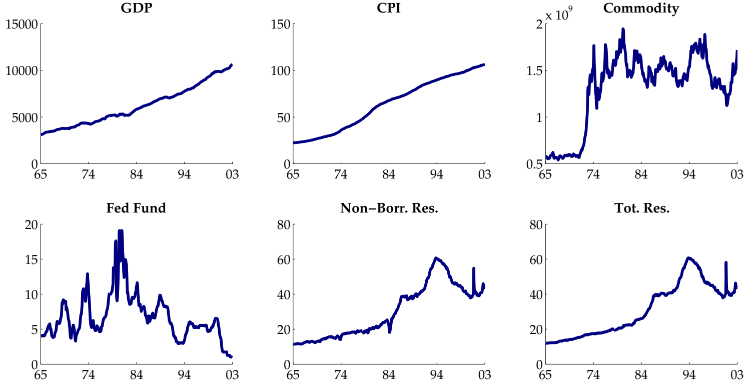
How can we check the long-run "neutrality" of demand shocks on output level?

- Let's simply plot the cumulative sum of the impulse responses of output growth



Example: sign restrictions

- Uhlig (2005). "What are the effects of monetary policy on output? Results from an agnostic identification procedure," Journal of Monetary Economics
- US monthly data from 1965.I to 2003.XII



What are the effects of monetary policy on output?

- Before asking what are the effects of a monetary policy shocks we should be asking, what is a monetary policy shock?
- In the inflation targeting era a monetary policy shock is an increase in the policy rate that
  - is ordered last in a Cholesky decomposition?
  - has no permanent effect on output?
  - ....?

What is a monetary policy shock?

- According to conventional wisdom, monetary contractions should
  1. Raise the federal funds rate
  2. Lower prices
  3. Decrease non-borrowed reserves
  4. Reduce real output

Again... What are the effects of monetary policy on output?

- Standard identification schemes do not fully accomplish the 4 points above
  - Liquidity puzzle: when identifying monetary policy shocks as surprise increases in the stock of money, interest rates tend to go down, not up
  - Price puzzle: after a contractionary monetary policy shock, even with interest rates going up and money supply going down, inflation goes up rather than down
- Successful identification needs to deliver results matching the conventional wisdom

Uhlig's identification assumptions

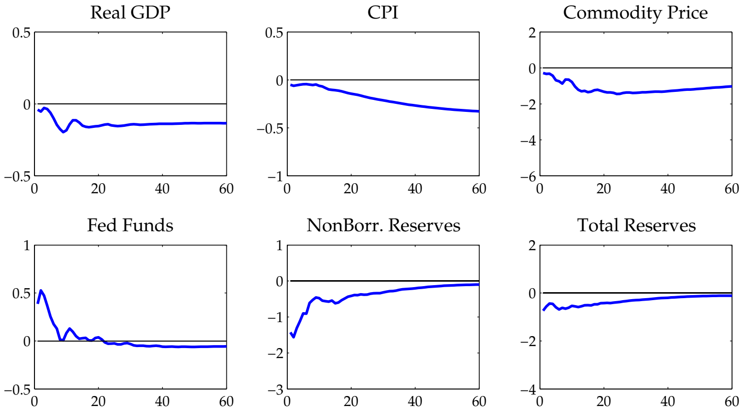
- Uhlig's assumption: a "contractionary" monetary policy shock does not lead to
  - Increases in prices
  - Increase in nonborrowed reserves
  - Decreases in the federal funds rate
- How about output? Since is the response of interest, we leave it un-restricted

[Reminder] How to compute sign restrictions

1. Estimate from the reduced-form VAR $\mathbf{F}$, $\mathbf{u}_t$, and $\Sigma_{\mathbf{u}}$
2. Draw a random orthonormal matrix $\mathbf{S}'$, compute $\mathbf{P}' =$`chol`$(\Sigma_{\mathbf{u}})$ and recover $\mathbf{A}^{-1} = \mathbf{P}'\mathbf{S}'$
3. Compute the impulse response using $\mathcal{IR}_1 = \mathbf{A}^{-1}\varepsilon_t$
4. Are the sign restrictions satisfied? Yes. Store the impulse response // No. Discard the impulse response
5. Perform N replications and report the median impulse response (and its confidence intervals)

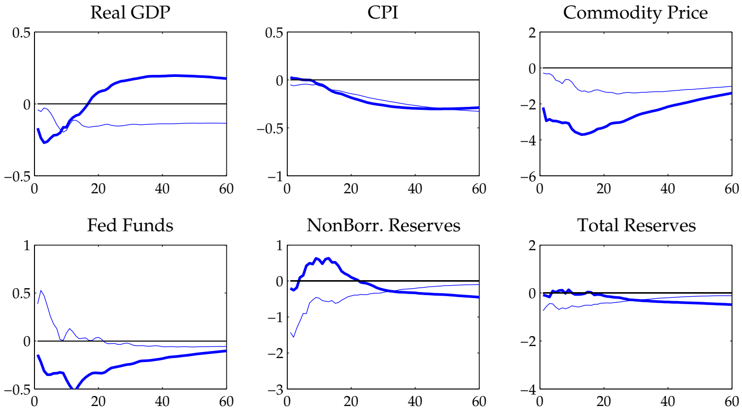
What happens when you do sign restrictions

- First draw: signs are correct, I keep it!



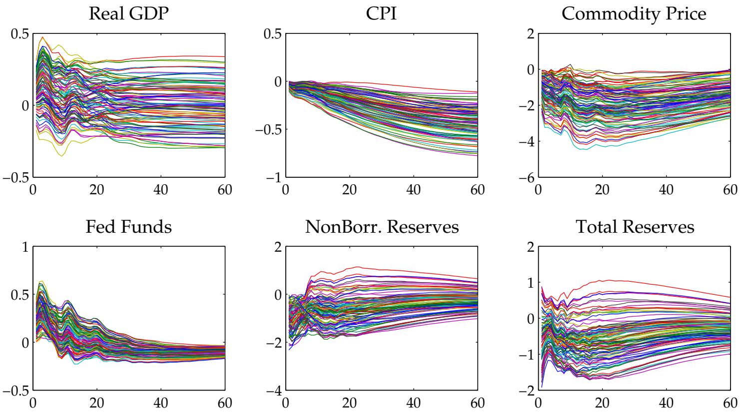
What happens when you do sign restrictions

- Second draw: signs are not correct, I discard it!



What happens when you do sign restrictions

- After a while...



What are the effects of monetary policy on output?

- Ambiguous effect on real GDP $\Longrightarrow$ Long-run monetary neutrality