Vector Auto Regressions in macroeconomics

Daniele Caratelli

Stanford University

April 3, 2018

Chocolate!

Disclaimer

The views expressed here are those of the author and the author alone. They do not reflect those of Profs. Tom MaCurdy and Frank Wolak or those of Stanford University more broadly.

Macroeconomic data

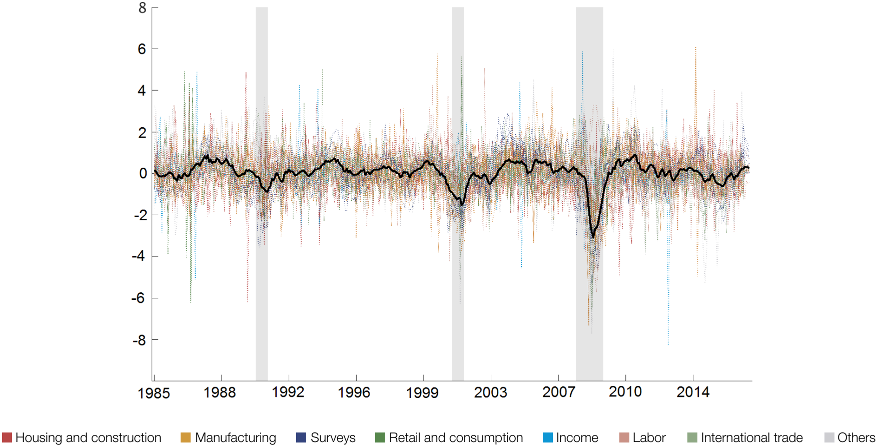

Figure: (Some) macro data by category

Macroeconomic data, cnt'd

- lots of data
- move along the *business cycle*
- idiosyncratic (leads and lags)

Macroeconomic data, cnt'd

- lots of data
- move along the *business cycle*
- idiosyncratic (leads and lags)

Uses:

1. Forecasting
2. Data description
3. Structural inference
4. Policy analysis

Macroeconomic data, cnt'd

- lots of data
- move along the *business cycle*
- idiosyncratic (leads and lags)

Uses:

1. Forecasting
2. Data description
3. Structural inference
4. Policy analysis

$\Rightarrow$ want statistical model to deal with macroeconomic variables.

A brief history

- Collecting macroeconomic data (e.g. NIPA) — '30s-'40s
- Large-scale models — '60s-present
- VARs — '70s-present
- DFMs — '80s-present
- Nowcasting — '00s-present

A brief history

- Large-scale models '60s-'80s
  - many equations informed by macro theory,
  - each with just a few variables, mostly considered exogenous,
  - problem: hard to discover empirical facts.

A brief history

- VARs '70s-present
  - × let the data speak (under identification assumptions),
  - × consistent with rational expectations.

From AR to V-AR

An AR(p) process for y is:

$$y_t = \mu + \rho_1 y_{t-1} + \dots \rho_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

From AR to V-AR

An AR(p) process for y is:

$$y_t = \mu + \rho_1 y_{t-1} + \dots \rho_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

What if we are interested in n-many variables each having an AR(p) process?

From AR to V-AR

An AR(p) process for y is:

$$y_t = \mu + \rho_1 y_{t-1} + \dots \rho_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

What if we are interested in n-many variables each having an AR(p) process? (1) holds for each of these and so:

$$\begin{cases} y_{1t} = \mu_1 + \rho_1^1 y_{1t-1} + \dots \rho_p^1 y_{1t-p} + \varepsilon_{1t}, \\ \vdots \\ y_{nt} = \mu_n + \rho_1^n y_{nt-1} + \dots \rho_p^n y_{1t-p} + \varepsilon_{nt}, \end{cases}$$

with $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2)$.

From AR to V-AR, cnt'd

Stack!!

$$\underbrace{\begin{bmatrix} y_{1t} \\ \vdots \\ y_{nt} \end{bmatrix}}_{\mathbf{Y_t}} = \underbrace{\begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}}_{\boldsymbol{\mu}} + \Phi_1 \times \underbrace{\begin{bmatrix} y_{1t-1} \\ \vdots \\ y_{nt-1} \end{bmatrix}}_{\mathbf{Y_{t-1}}} \cdots + \Phi_p \times \underbrace{\begin{bmatrix} y_{1t-p} \\ \vdots \\ y_{nt-p} \end{bmatrix}}_{\mathbf{Y_{t-p}}} + \underbrace{\begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix}}_{\boldsymbol{\varepsilon}_\mathbf{t}}$$

where

$$\Phi_i = \begin{bmatrix} \rho_i^1 & \dots & \dots & 0 \\ 0 & \rho_i^2 & \dots & 0 \\ & & \ddots & \\ \dots & \dots & & \dots \\ 0 & \dots & \dots & \rho_i^n \end{bmatrix} \qquad \boldsymbol{\varepsilon} \sim \sigma^2 I_{n \times n}$$

VAR(p)

As (macroeconomic) variables are all “connected” we can extend the above so that each variable depends not only on its own lags but also on the lags of the other variables.
Then we have the following system:

$$\begin{cases} y_{1t} = \mu_1 + \rho^1_{11} y_{1t-1} + \dots \rho^1_{1p} y_{1t-p} + \dots + \rho^1_{1n} y_{nt-1} + \dots \rho^1_{pn} y_{nt-p} + \varepsilon_{1t}, \\ \vdots \\ y_{nt} = \mu_n + \rho^n_{11} y_{1t-1} + \dots \rho^n_{1p} y_{1t-p} + \dots + \rho^n_{1n} y_{nt-1} + \dots \rho^n_{pn} y_{nt-p} + \varepsilon_{nt}, \end{cases}$$

Chol

VAR(p)

As (macroeconomic) variables are all "connected" we can extend the above so that each variable depends not only on its own lags but also on the lags of the other variables.
Then we have the following system:

$$\mathbf{Y_t} = \boldsymbol{\mu} + \Phi_1 \mathbf{Y_{t-1}} + \ldots \Phi_p \mathbf{Y_{t-p}} + \varepsilon_\mathbf{t}$$

VAR(p)

As (macroeconomic) variables are all "connected" we can extend the above so that each variable depends not only on its own lags but also on the lags of the other variables.
Then we have the following system:

$$\mathbf{Y_t} = \boldsymbol{\mu} + \Phi_1 \mathbf{Y_{t-1}} + \ldots \Phi_p \mathbf{Y_{t-p}} + \varepsilon_\mathbf{t}$$

$$\Phi_i = \begin{bmatrix} \rho_{i1}^1 & \rho_{i2}^1 & \cdot & \rho_{in}^1 \\ \cdot & \cdot & \cdot & \cdot \\ \rho_{i1}^n & \rho_{i2}^n & \cdot & \rho_{in}^n \end{bmatrix}$$

VAR, OLS formulation

Finally, we can can stack once again and write the VAR(p) as a VAR(1):

$$\underbrace{\begin{bmatrix} \mathbf{Y_t} \\ \mathbf{Y_{t-1}} \\ \vdots \\ \mathbf{Y_{t-(p-1)}} \end{bmatrix}}_{np\times 1} = \underbrace{\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix}}_{np\times 1} + \underbrace{\begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ I_{n\times n} & 0_{n\times n} & \dots & 0_{n\times n} \\ \dots & \ddots & \dots & \dots \\ 0_{n\times n} & \dots & I_{n\times n} & 0_{n\times n} \end{bmatrix}}_{np\times np} \underbrace{\begin{bmatrix} \mathbf{Y_{t-1}} \\ \mathbf{Y_{t-1}} \\ \vdots \\ \mathbf{Y_{t-p}} \end{bmatrix}}_{np\times 1} + \underbrace{\begin{bmatrix} \varepsilon_{\mathbf{t}} \\ 0_{n\times 1} \\ \vdots \\ 0_{n\times 1} \end{bmatrix}}_{np\times 1}$$

[1] Note this is known as the *companion-form*.

VAR, OLS formulation

Finally, we can can stack once again and write the VAR(p) as a VAR(1):

$$\underbrace{\begin{bmatrix} \mathbf{Y_t} \\ \mathbf{Y_{t-1}} \\ \vdots \\ \mathbf{Y_{t-(p-1)}} \end{bmatrix}}_{np\times 1} = \underbrace{\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix}}_{np\times 1} + \underbrace{\begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ I_{n\times n} & 0_{n\times n} & \dots & 0_{n\times n} \\ \dots & \ddots & \dots & \dots \\ 0_{n\times n} & \dots & I_{n\times n} & 0_{n\times n} \end{bmatrix}}_{np\times np} \underbrace{\begin{bmatrix} \mathbf{Y_{t-1}} \\ \mathbf{Y_{t-1}} \\ \vdots \\ \mathbf{Y_{t-p}} \end{bmatrix}}_{np\times 1} + \underbrace{\begin{bmatrix} \varepsilon_{\mathbf{t}} \\ 0_{n\times 1} \\ \vdots \\ 0_{n\times 1} \end{bmatrix}}_{np\times 1}$$

$$\bar{\mathbf{Y}}_{\mathbf{t}} = \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\Phi}}\bar{\mathbf{Y}}_{\mathbf{t-1}} + \bar{\varepsilon}_{\mathbf{t}},^{1} \qquad \bar{\varepsilon} \sim \mathcal{N}(\bar{\mathbf{0}}, \bar{\boldsymbol{\Sigma}}),$$

[1] Note this is known as the *companion-form*.

VAR, OLS formulation

Finally, we can can stack once again and write the VAR(p) as a VAR(1):

$$\underbrace{\begin{bmatrix} \mathbf{Y_t} \\ \mathbf{Y_{t-1}} \\ \vdots \\ \mathbf{Y_{t-(p-1)}} \end{bmatrix}}_{np\times 1} = \underbrace{\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix}}_{np\times 1} + \underbrace{\begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ I_{n\times n} & 0_{n\times n} & \dots & 0_{n\times n} \\ \dots & \ddots & \dots & \dots \\ 0_{n\times n} & \dots & I_{n\times n} & 0_{n\times n} \end{bmatrix}}_{np\times np} \underbrace{\begin{bmatrix} \mathbf{Y_{t-1}} \\ \mathbf{Y_{t-1}} \\ \vdots \\ \mathbf{Y_{t-p}} \end{bmatrix}}_{np\times 1} + \underbrace{\begin{bmatrix} \varepsilon_\mathbf{t} \\ 0_{n\times 1} \\ \vdots \\ 0_{n\times 1} \end{bmatrix}}_{np\times 1}$$

$$\mathbf{\bar{Y}_t} = \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\Phi}}\mathbf{\bar{Y}_{t-1}} + \bar{\varepsilon}_\mathbf{t},^1 \qquad \bar{\varepsilon} \sim \mathcal{N}(\bar{\mathbf{0}}, \bar{\boldsymbol{\Sigma}}),$$

$\longrightarrow$ run OLS and get $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\Phi}}$

[1] Note this is known as the *companion-form*.

VAR, OLS formulation

Finally, we can can stack once again and write the VAR(p) as a VAR(1):

$$\underbrace{\begin{bmatrix} \mathbf{Y_t} \\ \mathbf{Y_{t-1}} \\ \vdots \\ \mathbf{Y_{t-(p-1)}} \end{bmatrix}}_{np\times 1} = \underbrace{\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix}}_{np\times 1} + \underbrace{\begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ I_{n\times n} & 0_{n\times n} & \dots & 0_{n\times n} \\ \dots & \ddots & \dots & \dots \\ 0_{n\times n} & \dots & I_{n\times n} & 0_{n\times n} \end{bmatrix}}_{np\times np} \underbrace{\begin{bmatrix} \mathbf{Y_{t-1}} \\ \mathbf{Y_{t-1}} \\ \vdots \\ \mathbf{Y_{t-p}} \end{bmatrix}}_{np\times 1} + \underbrace{\begin{bmatrix} \varepsilon_\mathbf{t} \\ 0_{n\times 1} \\ \vdots \\ 0_{n\times 1} \end{bmatrix}}_{np\times 1}$$

$$\bar{\mathbf{Y}}_\mathbf{t} = \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\Phi}}\bar{\mathbf{Y}}_{\mathbf{t-1}} + \bar{\varepsilon}_\mathbf{t},^{1} \qquad \bar{\varepsilon} \sim \mathcal{N}(\bar{\mathbf{0}}, \bar{\boldsymbol{\Sigma}}),$$

$\longrightarrow$ run OLS and get $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\Phi}}$ $(n + pn^2$ parameters!)

[1] Note this is known as the *companion-form*.

Forecasting

$$
\begin{aligned}
E_{t+1}[\bar{\mathbf{Y}}_{\mathbf{t+1}}] &= E_{t+1}[\bar{\mu} + \bar{\mathbf{\Phi}}\bar{\mathbf{Y}}_{\mathbf{t}} + \bar{\varepsilon}_{\mathbf{t+1}}] = \bar{\mu} + E[\bar{\mathbf{\Phi}}\bar{\mathbf{Y}}_{\mathbf{t-1}}] \\
&= \bar{\mu} + \bar{\mathbf{\Phi}}E[\bar{\mu} + \bar{\mathbf{\Phi}}\bar{\mathbf{Y}}_{\mathbf{t-1}} + \bar{\varepsilon}_{\mathbf{t}}] \\
&= \bar{\mu} + \bar{\mathbf{\Phi}}\bar{\mu} + E[\bar{\mathbf{\Phi}}\bar{\mathbf{Y}}_{\mathbf{t-1}} + \bar{\varepsilon}_{\mathbf{t}}] = \dots \\
&= \left(\sum_{j=0}^{t} \bar{\mathbf{\Phi}}^{j}\right)\bar{\mu} + \bar{\mathbf{\Phi}}^{t+1}\bar{\mathbf{Y}}_{\mathbf{0}}
\end{aligned}
$$

Forecasting

$$
\begin{aligned}
E_{t+1}[\bar{\mathbf{Y}}_{\mathbf{t+1}}] &= E_{t+1}[\bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\Phi}}\bar{\mathbf{Y}}_{\mathbf{t}} + \bar{\boldsymbol{\varepsilon}}_{\mathbf{t+1}}] = \bar{\boldsymbol{\mu}} + E[\bar{\boldsymbol{\Phi}}\bar{\mathbf{Y}}_{\mathbf{t-1}}] \\
&= \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\Phi}} E[\bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\Phi}}\bar{\mathbf{Y}}_{\mathbf{t-1}} + \bar{\boldsymbol{\varepsilon}}_{\mathbf{t}}] \\
&= \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\Phi}}\bar{\boldsymbol{\mu}} + E[\bar{\boldsymbol{\Phi}}\bar{\mathbf{Y}}_{\mathbf{t-1}} + \bar{\boldsymbol{\varepsilon}}_{\mathbf{t}}] = \dots \\
&= \left(\sum_{j=0}^{t} \bar{\boldsymbol{\Phi}}^{j}\right) \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\Phi}}^{t+1}\bar{\mathbf{Y}}_{\mathbf{0}}
\end{aligned}
$$

Remark: # of parameters grows as the square of the # of variables.

$\Rightarrow$ Bayesian-VARs: discipline parameters penalizing over-fitting.

How well do VARs forecast?, cnt'd

$n = 3$ variables, $p = 4$ lags.

horizon	Inflation		FFR		Unemployment	
	AR	VAR	AR	VAR	AR	VAR
2	.74	.76	.80	.75	.49	.49
4	.82	.84	1.04	.98	.68	.61
8	.99	.97	1.28	1.27	.90	.90

Table: Small model: out-of-sample RMSFE (1985:Q1-2000:Q4)

How well do VARs forecast?, cnt'd

$n = 3$ variables, $p = 4$ lags.

horizon	Inflation			FFR			Unemployment		
	AR	VAR	BVAR	AR	VAR	BVAR	AR	VAR	BVAR
2	.74	.76	.75	.80	.75	.79	.49	.49	.47
4	.82	.84	.82	1.04	.98	1.03	.68	.61	.60
8	.99	.97	.99	1.28	1.27	1.25	.90	.90	.82

Table: Small model: out-of-sample RMSFE (1985:Q1-2000:Q4)

How well do VARs forecast?

$n = 12$ variables, $p = 4$ lags.

horizon	Inflation		FFR		Unemployment	
	AR	VAR	AR	VAR	AR	VAR
2	.74	1.04	.80	1.01	.49	.58
4	.82	1.08	1.04	1.15	.68	.63
8	.99	1.23	1.28	1.37	.90	.92

Table: Large model: out-of-sample RMSFE (1985:Q1-2000:Q4)

How well do VARs forecast?

$n = 12$ variables, $p = 4$ lags.

horizon	Inflation			FFR			Unemployment		
	AR	VAR	BVAR	AR	VAR	BVAR	AR	VAR	BVAR
2	.74	1.04	.77	.80	1.01	.69	.49	.58	.42
4	.82	1.08	.85	1.04	1.15	.93	.68	.63	.54
8	.99	1.23	1.09	1.28	1.37	1.18	.90	.92	.76

Table: Large model: out-of-sample RMSFE (1985:Q1-2000:Q4)

Data description

1. Coefficients don't tell you Jack!

2. Impulse Response Functions (IRFs)
   $\rightarrow$ how do variables respond to shocks?
   $\rightarrow$ need assumptions about the errors (recursive or structural VARs)

Impulse Response Functions

- Aim: determine $\frac{\partial Y_{it+j}}{\partial \varepsilon_{kt}}$.
- Problem: we don't know how ε_{kt} works.

$\Rightarrow$ Need to pin down dynamics and interactions of $\boldsymbol{\varepsilon_t}$

IRFs with Cholesky decomposition

- Error term in each regression equation is uncorrelated with the error in the preceding equations. (system)
- This is the Cholesky decomposition of the reduced form VAR covariance matrix.
- Can interpret **correlation** as **causation**.
- E.g. inflation, interest rate, unemployment.

IRFs: the impulse

Impulse

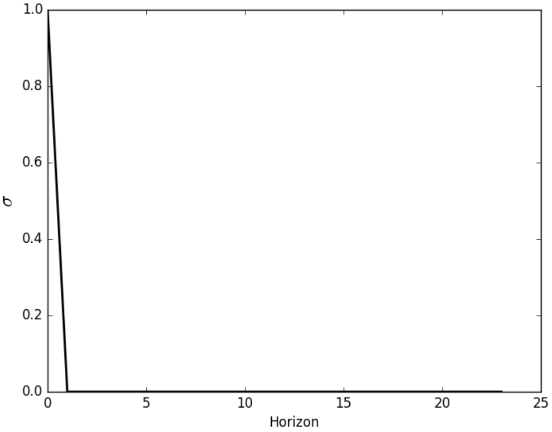

Figure: Standard deviation shock.

IRFs: computation

- Need to orthogonalize shocks
- Recall that $\boldsymbol{\varepsilon_t} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$
- Write $\boldsymbol{\Sigma} = PP'$ where P is lower-triangular $\Rightarrow \boldsymbol{\eta_t} : P^{-1}\boldsymbol{\varepsilon_t} \sim \mathcal{N}(0, I)$
- $\boldsymbol{\eta_t}$ is our orthogonal shock!

$$
\begin{aligned}
\frac{\partial Y_{t+j}}{\partial \eta_{kt}} &= \frac{\partial(\mu + \Phi Y_{t+j-1} + \varepsilon_{t+j})}{\partial \eta_{kt}} = \frac{\partial(\mu + \Phi Y_{t+j-1} + P\eta_{t+j})}{\partial \eta_{kt}} \\
&= \cdots = \frac{\partial\left(\left(\sum_{s=0}^{h} \Phi^s \mu\right) + \Phi^{h+1} Y_t + \Phi^{h+1} P \eta_t\right)}{\partial \eta_{kt}} \\
&= \left[\Phi^{h+1} P\right]_k
\end{aligned}
$$

IRFs: responses to interest rate shock

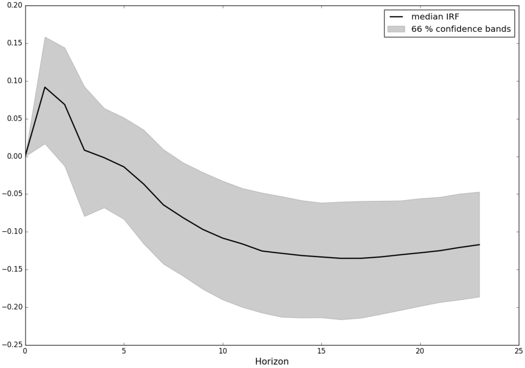

Figure: Response of inflation to interest rate shock.

IRFs: responses to interest rate shock

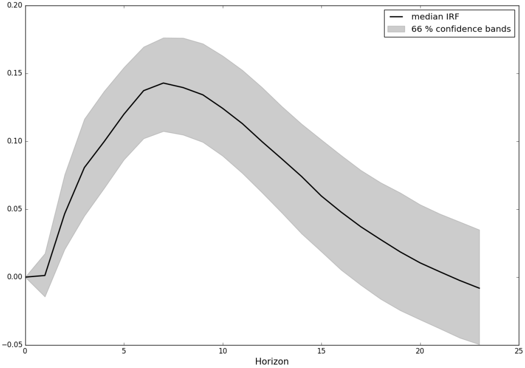

Figure: Response of unemployment to interest rate shock.

IRFs: responses to interest rate shock

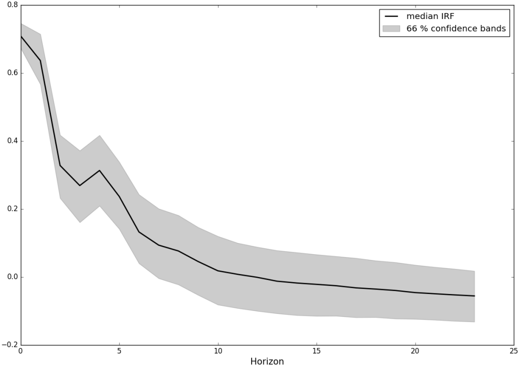

Figure: Response of interest rate to interest rate shock.

IRFs: responses to inflation shock

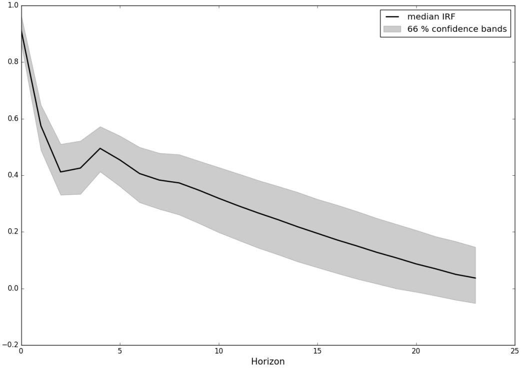

Figure: Response of inflation to inflation shock.

IRFs: responses to inflation shock

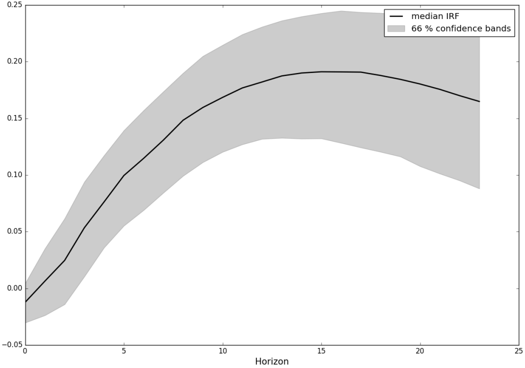

Figure: Response of unemployment to inflation shock.

IRFs: responses to inflation shock

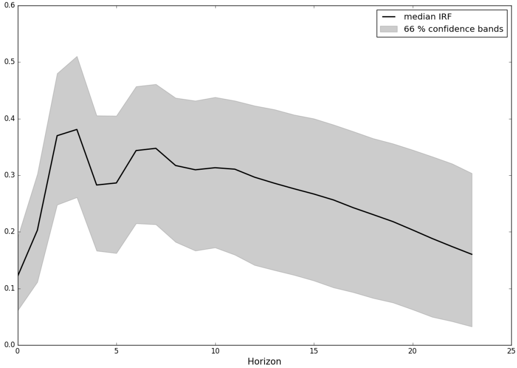

Figure: Response of interest rate to inflation shock.

More on VARs

- More identification techniques (e.g. sign restrictions; partial identification)

More on VARs

- More identification techniques (e.g. sign restrictions; partial identification)
- Including more structure (e.g. Taylor rule)

More on VARs

- More identification techniques (e.g. sign restrictions; partial identification)
- Including more structure (e.g. Taylor rule)
- Time-varying parameters (TVARs)

More on VARs

- More identification techniques (e.g. sign restrictions; partial identification)
- Including more structure (e.g. Taylor rule)
- Time-varying parameters (TVARs)

How to deal with...

More on VARs

- More identification techniques (e.g. sign restrictions; partial identification)
- Including more structure (e.g. Taylor rule)
- Time-varying parameters (TVARs)

How to deal with...

- Mixed-frequency data

More on VARs

- More identification techniques (e.g. sign restrictions; partial identification)
- Including more structure (e.g. Taylor rule)
- Time-varying parameters (TVARs)

How to deal with...

- Mixed-frequency data
- Missing data

More on VARs

- More identification techniques (e.g. sign restrictions; partial identification)
- Including more structure (e.g. Taylor rule)
- Time-varying parameters (TVARs)

How to deal with...

- Mixed-frequency data
- Missing data
- Jagged-edges data

References:

Main:

- Doan, Thomas, et al. "Forecasting and Conditional Projection Using Realistic Prior Distributions." Econometric Reviews, vol. 3, no. 1, 1984, pp. 1–100.
- Hamilton, James D. "Time Series Analysis." Princeton University Press, 1994.
- Sims, Christopher A. "Macroeconomics and Reality." *Econometrica* 48, no. 1 (1980): 1-48.
- Stock, James H., and Mark W. Watson. "Vector Autoregressions." *The Journal of Economic Perspectives* 15, no. 4 (2001): 101-15.
- Uhlig, H. "Economics and Reality." *MFI Working Paper Series* no. 2011-006.