Direct and Iterated Methods for Impulse Responses: theory and applications

Silvia Miranda-Agrippino
Bank of England and CFM

London Business School - July 1, 2016

[▷]

Outline

i. IRFs from Vector Autoregressions (**VARs**)

ii. Alternative: Local Projection Method (**LPs**)
   - Estimation and Inference
   - Relation with VAR
   - Potential Drawbacks

iii. Direct vs Iterated Methods: Forecasting

iv. Bayesian Local Projections (**BLPs**)
   - Intuition
   - Bayesian VARs: priors and prior tightness
   - Estimation and Inference
   - Application

IRFs from Vector Autoregressions

Multivariate Forecast

Consider the n-dimensional process

$$\underset{[n\times 1]}{y_t} \equiv (y_{1,t}, \dots, y_{n,t})'$$

Conditional on information (i.e. realizations of y_t) up to time $t-1$

$$\mathcal{O}_{t-1} \equiv \text{span}\{y_{t-1}, y_{t-2}, \dots\}$$

the optimal linear forecast for y_t is

$$\hat{y}_{t|t-1} = \text{Proj}(y_t|\mathcal{O}_{t-1}) = \mathbb{E}(y_t|\mathcal{O}_{t-1})$$

INNOVATIONS

Forecast errors / innovations:

$$u_t \equiv y_t - \hat{y}_{t|t-1} = y_t - \text{Proj}(y_t|\mathcal{O}_{t-1})$$

$u_t \sim \text{WN}(0, \Sigma_u)$:

- $\mathbb{E}(u_{i,t}) = 0 \quad i = 1, \dots, n \quad \forall t$
- $\mathbb{E}(u_{i,t}u'_{j,t}) = \sigma_{ij}$
- $\mathbb{E}(u_{i,t}u'_{j,t-s}) = 0 \quad \forall i, j \quad \forall s > 0$

Remark: Optimality $\Rightarrow u_t \perp \mathcal{O}_{t-1}$

Wold (MA) Representation

Any covariance-stationary process y_t can be written as

$$
\begin{aligned}
y_t &= \mu + \psi(L)u_t \\
&= \mu + \sum_{j=0}^{\infty} \psi_j u_{t-j} \\
&= \mu + \psi_0 u_t + \psi_1 u_{t-1} + \ldots
\end{aligned}
$$

where (i) $\psi_0 = 1$ and (ii) $\sum_{j=0}^{\infty} \psi_j < \infty$

IMPULSE RESPONSE FUNCTIONS

$$\psi_j = \frac{\partial y_{t+j}}{\partial u_t'}$$

n x 1

1 x n

n x n

In practice

$$u_t = y_t - \left[c + \sum_{j=1}^{\overset{p}{\cancel{\infty}}} \phi_j y_{t-j} \right]$$

$$\downarrow$$

Finite-order Vector Autoregression: **VAR(p)**

$$\downarrow$$

$$\underset{[n\times 1]}{y_t} = c + \underset{[n\times n]}{\phi_1} y_{t-1} + \ldots + \phi_p y_{t-p} + u_t$$

VAR(P) → VAR(1)

- VAR(p):

$$\underset{[n\times 1]}{y_t} = \underset{[n\times n]}{\phi_1} \; y_{t-1} + \ldots + \phi_p y_{t-p} + u_t$$

- VAR(1) (companion form):

p

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \ldots & \phi_p \\ \mathbb{I}_n & 0 & \ldots & 0 \\ \vdots & \mathbb{I}_n & \ddots & 0 \\ 0 & 0 & \mathbb{I}_n & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\underset{[n(p\cancel{-1})\times 1]}{Y_t} = \underset{[np\times np]}{\Phi} \; Y_{t-1} + U_t$$

IRF Coefficients from VAR(1)

$$
\begin{aligned}
Y_t &= \Phi Y_{t-1} + U_t \\
&= \Phi^2 Y_{t-2} + U_t + \Phi U_{t-1} \\
&\vdots \\
&= \underbrace{\Phi^j Y_{t-j}}_{=0\, j\to\infty} + U_t + \Phi U_{t-1} + \ldots + \Phi^j U_{t-j}
\end{aligned}
$$

$B^j \equiv$ upper left $[n \times n]$ block of Φ^j:

$$y_t = u_t + B u_{t-1} + \ldots + B^j u_{t-j}$$

$$\downarrow$$

$$\psi_j = \frac{\partial y_{t+j}}{\partial u_t'}$$

Equivalently:

$$
\begin{aligned}
Y_{t+1} &= \Phi Y_t + U_t \\
&\vdots \\
Y_{t+h} &= \underbrace{\Phi^h Y_t}_{Y_{t+h|t}} + U_{t+h} + \Phi U_{t+h-1} + \ldots + \Phi^{h-1} U_{t+1}
\end{aligned}
$$

$B^h \equiv$ upper left $[n \times n]$ block of Φ^h:

VAR-IRF

$$IRF_h^{VAR} = B^h \qquad h = 0, \ldots, H$$

Forecasting Model:

- $y_t = \mu + u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \ldots \qquad u_t \sim NW(0, \Sigma)$
- $\mathcal{O}_{t-1} = \text{span}\{y_t, y_{t-1}, \ldots\}$ ➔ Information Set Econometrician
- $u_t = y_t - \text{Proj}(y_t | \mathcal{O}_{t-1})$

Structural Model:

- $y_t = \mu + A_0 e_t + A_1 e_{t-1} + A_2 e_{t-2} + \ldots \qquad e_t \sim NW(0, \mathbb{I}_n)$
- $\mathcal{I}_{t-1} = \text{span}\{e_t, e_{t-1}, \ldots\}$ ➔ Information Set Agents
- $A_0 e_t = y_t - \text{Proj}(y_t | \mathcal{I}_{t-1})$

IDENTIFICATION

$$\mathcal{O}_{t-1} = \mathcal{I}_{t-1}$$

$$\Downarrow$$

$$y_t - \text{Proj}(y_t | \mathcal{O}_{t-1}) = y_t - \text{Proj}(y_t | \mathcal{I}_{t-1})$$

$$\Downarrow$$

$$u_t = A_0 e_t$$

$$\Downarrow$$

$$\Sigma_u = A_0 A_0'$$

Assumption: number of shocks $\leq$ number of variables

i. Specify VAR(p) for y_t:

$$y_t = c + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + u_t$$

ii. Iterate / MA Representation:

$$y_t = \kappa + u_t + B u_{t-1} + \ldots + B^j u_{t-j} + \ldots$$

iii. Shock Identification:

$$u_t = A_0 e_t$$

$$\Downarrow$$

VAR-IRF

$$IRF_h^{VAR} = \underline{B^h A_0} \qquad h = 0, \ldots, H$$

VAR as misspecified representations of the true DGP

There is no reason to believe that the VAR be generally equal to the true DGP ➔ VAR as an approximation of the true model (e.g. VARMA)

VARs are designed for one-step ahead forecast optimality.

If misspecified:

i. can still deliver reasonable $t+1|t$ forecasts [Stock and Watson (1999)]

ii. but misspecification errors will be compounded at further-ahead horizons ➔ potentially important distortions for the IRFs

VAR misspecification: IRFs - Braun and Mittnik (1993)

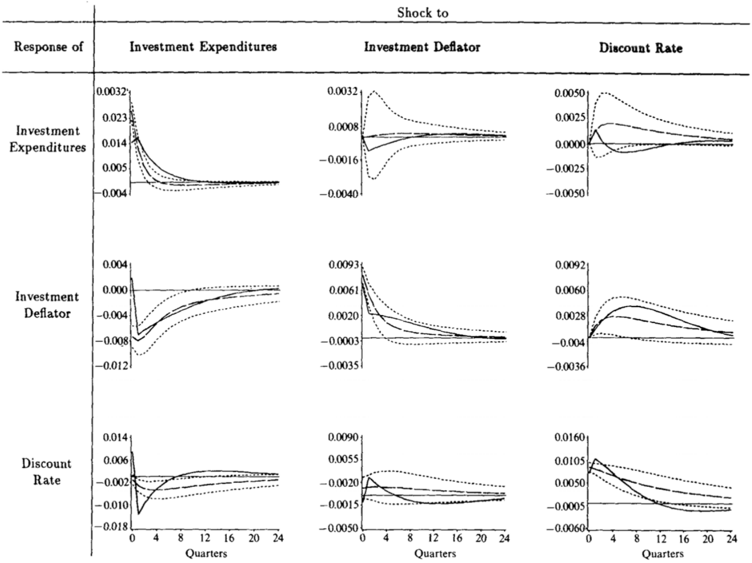


Fig. 1. Impulse response functions – Trivariate VAR(1). Key: trivariate ARMA (1,1) ——, trivariate VAR(1) – – –, ± 2 std. deviation - - - -.

The Local Projection Method

IRFs coefficients are obtained by **projecting the variable of interest at the desired horizon onto the relevant information set**

Alternative to standard methods, LP does not require the specification and estimation of an underlying multivariate model

IRFs based on Local Projection methods were first introduced by Òscar Jordà (2005)

General Idea

Suppose

$n \times p$

$$\underset{[n\times 1]}{y_t} = f(\mathcal{Y}_t)$$

where $\mathcal{Y}_t \equiv [y_{t-1}, y_{t-2}, \ldots, y_{t-p}]'$. We want to compute the response of y_{t+h} to a given shock of size $\mathbf{s}$ *without specifying a model* for y_t

Define the IRFs as:

$$IRF(h, \mathbf{s}) = \mathbb{E}[y_{t+h}|v_t = \mathbf{s}, \mathcal{Y}_t] - \mathbb{E}[y_{t+h}|v_t = 0, \mathcal{Y}_t] \qquad h = 0, \ldots, H$$

where $\mathbb{E}[\,\cdot\,]$ is the best predictor (in mean square sense)

Local Projection IRFs

Consider the linear local projection

$$y_{t+h} = \tilde{c} + \underline{\tilde{B}^{(h)}} y_t + \ldots + \tilde{B}_p^{(h)} y_{t-p} + v_{t+h}$$

then

$$\mathbb{E}[y_{t+h} | v_t = \mathbf{s}, \mathcal{Y}_t] = \underline{\tilde{c} + \tilde{B}^{(h)}(v_t = \mathbf{s}}) + \ldots + \tilde{B}_p^{(h)} y_{t-p} \qquad (1)$$

$$\mathbb{E}[y_{t+h} | v_t = 0, \mathcal{Y}_t] = \tilde{c} \qquad\qquad + \ldots + \tilde{B}_p^{(h)} y_{t-p} \qquad (2)$$

The IRFs from the local linear projection are

LP-IRF

$$\underline{IRF_h^{LP} = \tilde{B}^{(h)} \mathbf{s}} \qquad h = 0, \ldots, H$$

LP-based IRFs

i. Estimate Local Linear Projection for y_{t+h}:

$$y_{t+h} = \tilde{c} + \tilde{B}^{(h)} y_t + \ldots + \tilde{B}_p^{(h)} y_{t-p} + v_{t+h}$$

ii. Shock Identification:

$$u_t = A_0 e_t$$

$$\Downarrow$$

LP-IRF

$$IRF_h^{LP} = \tilde{B}^{(h)} A_0 \qquad h = 0, \ldots, H$$

Relation with VAR

- Iterated VAR(p):

$$y_{t+h} = \kappa + \underline{B^h y_t} + \ldots + \underline{u_{t+h} + \ldots + B^{h-1} u_{t+1}}$$

- LP(p):

$$y_{t+h} = \tilde{c} + \tilde{B}^{(h)} y_t + \ldots + \tilde{B}_p^{(h)} y_{t-p} + \underline{v_{t+h}}$$

If the DGP is the VAR:

i. $\tilde{B}^{(h)} = B^h$

ii. $v_{t+h} = u_{t+h} + \ldots + B^{h-1} u_{t+1}$

Projection residuals are a moving average of forecast errors dated $t+1, \ldots, t+h$

$$y_{t+h} = \tilde{c} + \tilde{B}^{(h)} y_t + \ldots + \tilde{B}_p^{(h)} y_{t-p} + v_{t+h}$$

$$v_{t+h} = f(u_{t+h}, \ldots, u_{t+1})$$

Hence orthogonal to the projection set ➔ $v_{t+h} \perp \text{span}\{y_t, \ldots, y_{t-p}\}$

Result: Local Projection coefficients (IRFs) are consistently estimated using standard Least Squares estimators

$$v_{t+h} \sim MA(h-1)$$

$$\Sigma_v^{(h)} \equiv \mathbb{E}(v_{t+h} v'_{t+h})$$

HAC-consistent (Newey-West) estimator for $\Sigma_v^{(h)}$:

$$\Sigma_{v,HAC}^{(h)} = \hat{\Gamma}_v(0) + \sum_{j=1}^{h-1} \omega_j \left[\hat{\Gamma}_v(j) + \hat{\Gamma}_v(j)'\right]$$

- $\hat{\Gamma}_v(j) = T^{-1} \sum_{t=j+1}^{T} (v_{t+h} v'_{t+h-j})$
- $\omega_j = 1 - j/h$ Bartlett weights

i. Estimate Local Linear Projection for y_{t+h}:

$$y_{t+h} = \tilde{c} + \tilde{B}^{(h)} y_t + \ldots + \tilde{B}_p^{(h)} y_{t-p} + v_{t+h}$$

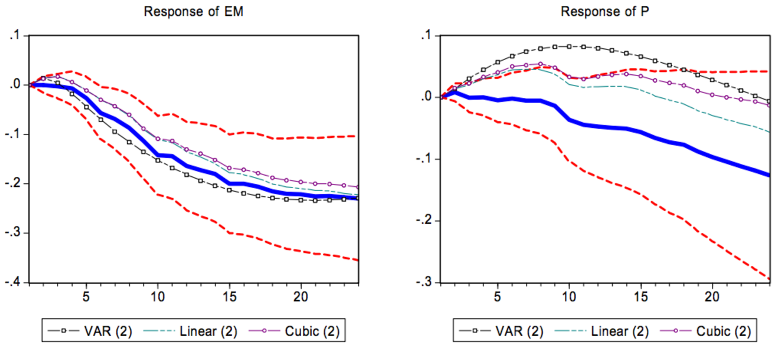
Note: Consistency does not require estimating the full system jointly ➔ the response of each $y_{i,t+h}$ can be estimated by linear regression

ii. Estimate HAC-consistent coefficients variance for each $\tilde{B}^{(h)} \quad h = 1, \ldots, H$

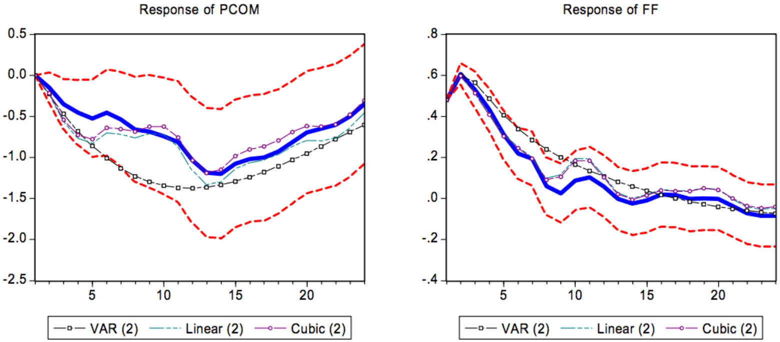
Note: LP residuals up to $h-1$ could be in principle used in regression at horizon h to improve efficiency

iii. Construct robust standard errors around projection coefficients at each horizon h

- True Model: VAR(12) on monthly data 1960:2001
- Misspecification: $p = 2$



- True Model: VAR(12) on monthly data 1960:2001
- Misspecification: $p = 2$



LP: Monte Carlo - Jordà (2005)

- True Model: VAR(12) on monthly data 1960:2001
- Univariate LP(12)

	EM			P			PCOM		
s	True-MC	Newey-West (Linear)	Newey-West (Cubic)	True-MC	Newey-West (Linear)	Newey-West (Cubic)	True-MC	Newey-West (Linear)	Newey-West (Cubic)
1	0.000	0.007	0.008	0.000	0.007	0.007	0.000	0.089	0.096
2	0.008	0.011	0.012	0.007	0.010	0.011	0.094	0.146	0.161
3	0.013	0.015	0.016	0.012	0.014	0.015	0.155	0.191	0.212
4	0.018	0.019	0.021	0.015	0.017	0.018	0.202	0.224	0.250
5	0.022	0.023	0.025	0.018	0.020	0.022	0.240	0.255	0.284
6	0.027	0.026	0.030	0.021	0.023	0.025	0.267	0.279	0.311
7	0.031	0.030	0.033	0.025	0.026	0.029	0.296	0.301	0.335
8	0.035	0.033	0.037	0.028	0.029	0.032	0.325	0.322	0.357
9	0.038	0.036	0.040	0.031	0.032	0.035	0.350	0.340	0.376
10	0.041	0.039	0.043	0.035	0.035	0.039	0.361	0.356	0.392
11	0.044	0.042	0.046	0.038	0.038	0.042	0.377	0.371	0.407
12	0.046	0.044	0.048	0.042	0.042	0.045	0.390	0.380	0.416

Because LP leave the specification of the underlying model largely unrestricted, they can easily accommodate nonlinearities

With quadratic and cubic terms:

$$
\begin{aligned}
y_{t+h} =& \tilde{c} + \tilde{B}^{(h)} y_t + \tilde{Q}^{(h)} y_t^2 + \tilde{C}^{(h)} y_t^3 \\
& + \tilde{B}_2^{(h)} y_{t-2} + \ldots + \tilde{B}_p^{(h)} y_{t-p} + v_{t+h}
\end{aligned}
$$

CUBIC LP-IRF

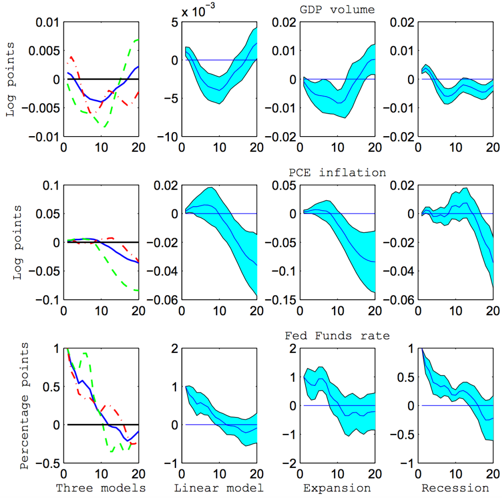
$$
IRF_h^{CLP} = \tilde{B}^{(h)}\mathbf{s} + \tilde{Q}^{(h)}(2y_t\mathbf{s} + \mathbf{s}^2) + \tilde{C}^{(h)}(3y_t^2\mathbf{s} + 3y_t\mathbf{s}^2 + \mathbf{s}^3)
$$

$$IRF_h^{CLP} = \tilde{B}^{(h)}\mathbf{s} + \tilde{Q}^{(h)}(2y_t\mathbf{s} + \mathbf{s}^2) + \tilde{C}^{(h)}(3y_t^2\mathbf{s} + 3y_t\mathbf{s}^2 + \mathbf{s}^3)$$

Inclusion of higher order terms makes responses of y_{t+h} to shock of size $\mathbf{s}$ depend on:

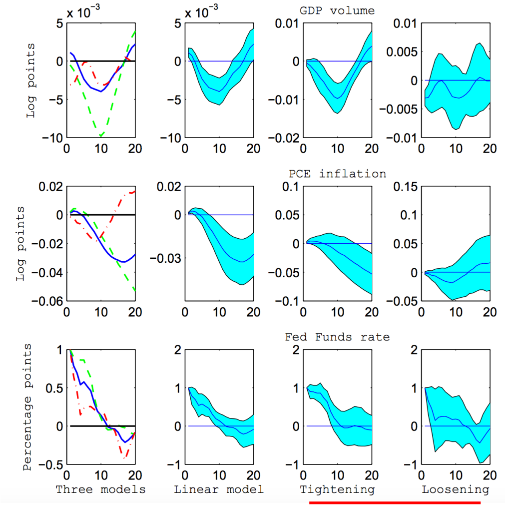
- local history of y_t
- the size of the shock
- the sign of the shock

- Asymmetric effects of monetary policy shocks [Tenreyro and Thwaites (forthcoming)]



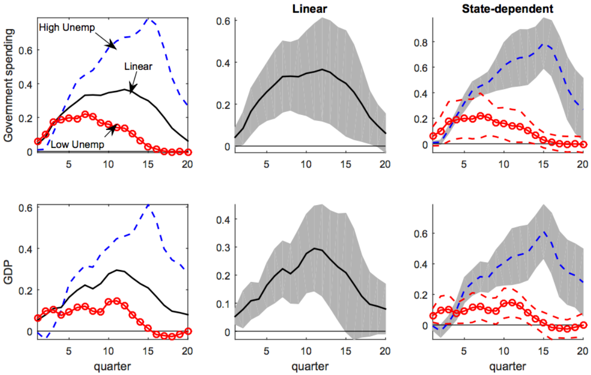
LP: Applications

- Asymmetric effects of monetary policy shocks [Tenreyro and Thwaites (forthcoming)]



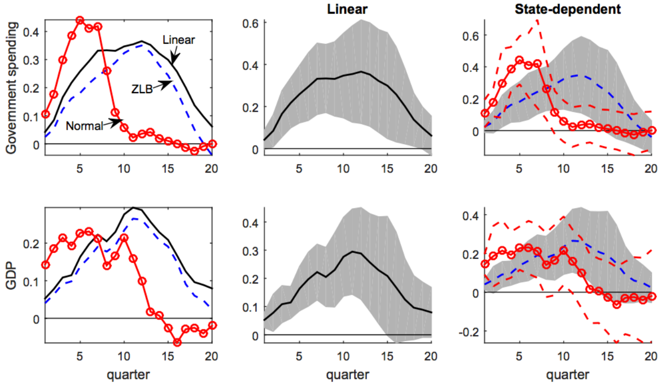
- Government spending multipliers in good and bad times [Ramey and Zubairy (2016)]

Figure 5. Government spending and GDP responses to a news shock: Considering slack states

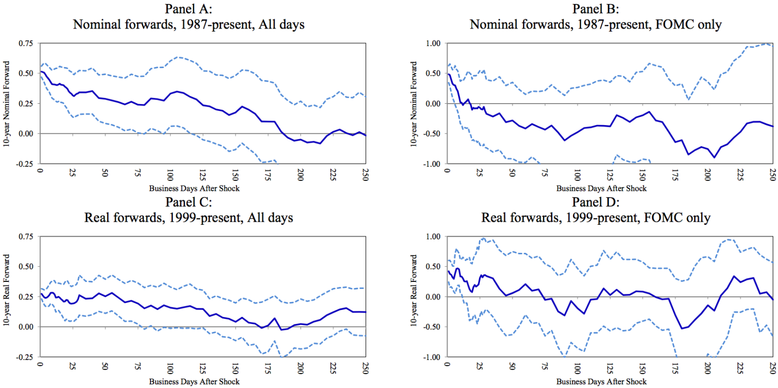


- Government spending multipliers in good and bad times [Ramey and Zubairy (2014)]

Figure 11. Government spending and GDP responses to a news shock: Considering zero lower bound



- Monetary policy on (daily) long-term forward rates
  [Hanson and Stein (2012)]



BANK OF ENGLAND

Direct vs Iterated Methods: Forecasting

Iterated vs Direct Forecasts: Theory

Which one is best?

Tradeoff between bias and estimation variance

- Iterated methods (VAR) give more efficient parameters estimates
- Direct methods (LP) are more robust to misspecifications

- Ignoring estimation uncertainty:

  i. VAR(p) and LP(p) but $p^* > p \Rightarrow MSFE(LP) \leq MSFE(VAR)$

  ii. VAR(p) and LP(p) but $p^* \leq p \Rightarrow MSFE(LP) = MSFE(VAR)$

Related Literature: Marcellino, Stock and Watson (2006) and references therein, Schorfheide (2005)

With estimation uncertainty:

VAR(p) and LP(p) and $p^* = p \Rightarrow MSFE(VAR) \leq MSFE(LP)$

Implausible that low-order VARs are well specified (i.e. best linear predictor)

$\Downarrow$

LP are theoretically preferable due to robustness to misspecification

- **Are gains also realized in practice?**

DGP:

$$y_t = \underbrace{M'FY_{t-1}}_{AR:\quad p \neq p^*} + u_t + \underbrace{\alpha T^{-1/2} \sum_{j=1}^{\infty} M' A_j M u_{t-j}}_{MA:\, \alpha = \text{severity of misspecification}}$$

Design:

- misspecification shrinks as $\underline{T \to \infty}$
- $\underline{p^* = 1,\ VMA(10)}$
- MLE : iterated, LFE : direct
- % change in Prediction Risk: $\mathcal{R}(\hat{y}_{t+h}) = \mathbb{E}[\|\hat{y}_{t+h} - \mathbb{E}_t(y_{t+h})\|^2] \geq 0$

Iterated vs Direct Forecasts: Monte Carlo - Schorfheide (2005)

T		$p=1$		$p=2$		$p=4$	
		MLE	LFE	MLE	LFE	MLE	LFE
	Misspecification $\alpha = 0$						
100	Risk	-67.8	-61.8	-52.9	-41.8	-11.9	0.0
	$\mathbb{E}[PC]$	-48.6	-44.1	-37.2	-27.7	-8.4	0.0
	$\sigma[PC]$	*38.2*	*36.8*	*32.6*	*28.5*	*9.0*	*0.0*
500	Risk	-69.1	-63.5	-54.8	-42.7	-12.5	0.0
	$\mathbb{E}[PC]$	-64.9	-59.5	-51.1	-39.2	-11.6	0.0
	$\sigma[PC]$	*39.2*	*37.8*	*34.1*	*29.6*	*10.4*	*0.0*
5000	Risk	-69.5	-63.9	-54.9	-42.2	-12.1	0.0
	$\mathbb{E}[PC]$	-69.3	-63.8	-55.0	-42.5	-12.5	0.0
	$\sigma[PC]$	*39.7*	*38.1*	*34.7*	*30.1*	*10.8*	*0.0*
∞	Risk	-69.8	-64.3	-55.4	-42.9	-12.5	0.0
	Misspecification $\alpha = 2$						
100	Risk	-51.3	-56.6	-41.5	-36.4	-9.9	0.0
	$\mathbb{E}[PC]$	-32.5	-35.8	-23.0	-23.0	-6.0	0.0
	$\sigma[PC]$	*38.5*	*36.5*	*33.3*	*28.3*	*8.7*	*0.0*
500	Risk	-48.2	-54.1	-36.5	-34.6	-10.6	0.0
	$\mathbb{E}[PC]$	-44.6	-49.1	-31.2	-31.2	-9.5	0.0
	$\sigma[PC]$	*40.5*	*38.5*	*36.8*	*30.5*	*10.5*	*0.0*
5000	Risk	-48.3	-52.9	-35.2	-33.4	-10.3	0.0
	$\mathbb{E}[PC]$	-48.3	-52.4	-34.5	-33.4	-10.5	0.0
	$\sigma[PC]$	*41.5*	*39.3*	*37.9*	*31.6*	*11.0*	*0.0*
∞	Risk	-48.8	-52.6	-35.8	-33.8	-10.7	0.0

Iterated vs Direct Forecasts: Monte Carlo - Schorfheide (2005)

T		$p=1$		$p=2$		$p=4$	
		MLE	LFE	MLE	LFE	MLE	LFE
	Misspecification $\alpha = 5$						
100	Risk	−8.0	−34.0	−19.2	−20.9	0.3	0.0
	$\mathbb{E}[PC]$	6.4	−16.2	−5.2	−12.9	2.8	0.0
	$\sigma[PC]$	*36.0*	*31.1*	*29.2*	*23.9*	*9.7*	*0.0*
500	Risk	25.7	−20.0	19.1	−9.0	−2.2	0.0
	$\mathbb{E}[PC]$	27.7	−14.6	24.2	−7.6	−1.1	0.0
	$\sigma[PC]$	*40.8*	*36.7*	*39.5*	*29.9*	*10.8*	*0.0*
5000	Risk	32.3	−11.7	41.1	−1.0	−3.8	0.0
	$\mathbb{E}[PC]$	32.2	−10.8	42.6	−1.0	−3.9	0.0
	$\sigma[PC]$	*43.9*	*40.3*	*44.6*	*33.7*	*11.2*	*0.0*
∞	Risk	30.6	−8.7	37.8	0.4	−3.7	0.0
	Misspecification $\alpha = 10$						
100	Risk	21.0	−9.3	−7.9	−6.6	14.7	0.0
	$\mathbb{E}[PC]$	30.5	1.7	−0.1	−1.9	13.1	0.0
	$\sigma[PC]$	*30.9*	*23.4*	*20.0*	*17.0*	*11.0*	*0.0*
500	Risk	91.1	15.1	42.4	11.4	15.1	0.0
	$\mathbb{E}[PC]$	91.9	18.7	45.1	11.2	15.5	0.0
	$\sigma[PC]$	*31.9*	*25.6*	*27.8*	*21.0*	*10.1*	*0.0*
5000	Risk	147.3	43.1	139.7	40.6	6.7	0.0
	$\mathbb{E}[PC]$	147.1	44.1	141.1	40.4	6.7	0.0
	$\sigma[PC]$	*37.6*	*33.4*	*40.5*	*28.4*	*9.8*	*0.0*
∞	Risk	147.2	55.9	146.1	50.7	6.6	0.0

Iterated vs Direct Forecasts: Practice - MSW (2006)

- Large-scale exercise on 170 US variables from 1959:1 to 2002:2
- Recursive "pseudo" out-of-sample forecasts at $h = 3, 6, 12, 24$
- $MSFE(LP)/MSFE(VAR)$, $H_0 : VAR$ is efficient

Lag Selection	Mean/Percentile	Forecast Horizon			
		3	6	12	24
AR(4)	mean	0.99 (<.005)	0.99 (<.005)	1.00 (<.005)	1.05 (0.83)
	0.10	0.97 (<.005)	0.92 (<.005)	0.90 (<.005)	0.85 (<.005)
	0.25	0.99 (<.005)	0.98 (<.005)	0.98 (<.005)	0.97 (0.04)
	0.50	1.00 (0.01)	1.00 (0.03)	1.01 (0.25)	1.05 (>.995)
	0.75	1.01 (0.85)	1.02 (0.83)	1.04 (0.55)	1.12 (>.995)
	0.90	1.02 (0.83)	1.04 (0.86)	1.08 (0.82)	1.23 (0.99)
AR(12)	mean	1.01 (>.995)	1.01 (>.995)	1.03 (>.995)	1.10 (>.995)
	0.10	0.98 (>.995)	0.97 (>.995)	0.95 (>.995)	0.93 (>.995)
	0.25	1.00 (>.995)	0.99 (>.995)	1.00 (>.995)	1.02 (>.995)
	0.50	1.00 (>.995)	1.01 (>.995)	1.03 (>.995)	1.09 (>.995)
	0.75	1.01 (>.995)	1.02 (>.995)	1.06 (>.995)	1.17 (>.995)
	0.90	1.02 (0.99)	1.05 (>.995)	1.11 (>.995)	1.29 (>.995)
AR(BIC)	mean	0.98 (<.005)	0.97 (<.005)	0.99 (0.21)	1.05 (0.99)
	0.10	0.92 (<.005)	0.86 (<.005)	0.86 (0.01)	0.88 (0.06)
	0.25	0.97 (<.005)	0.96 (<.005)	0.97 (0.02)	0.98 (0.50)
	0.50	1.00 (<.005)	1.00 (0.01)	1.01 (0.56)	1.04 (>.995)
	0.75	1.01 (0.99)	1.02 (0.91)	1.03 (0.76)	1.12 (>.995)
	0.90	1.03 (>.995)	1.05 (>.995)	1.10 (>.995)	1.20 (0.98)
AR(AIC)	mean	1.00 (>.995)	1.01 (>.995)	1.02 (>.995)	1.09 (>.995)
	0.10	0.97 (0.51)	0.95 (0.99)	0.94 (>.995)	0.91 (0.97)
	0.25	0.98 (0.08)	0.98 (0.90)	0.98 (0.97)	1.00 (>.995)
	0.50	1.00 (0.22)	1.00 (>.995)	1.02 (>.995)	1.07 (>.995)
	0.75	1.01 (>.995)	1.03 (>.995)	1.06 (>.995)	1.18 (>.995)
	0.90	1.04 (>.995)	1.06 (>.995)	1.11 (>.995)	1.29 (>.995)

Which method is best is largely an empirical matter:

i. $p = p^*_{AIC} \Rightarrow MSFE(VAR) \leq MSFE(LP)$
Note: Gains are modest: largely due to reduction in estimation uncertainty

ii. $MSFE(LP) \geq MSFE(VAR)$ the larger h
Note: Again efficiency of VAR outweighs robustness of LP

iii. $MSFE(LP) \leq MSFE(VAR)$ for series with large MA roots
Note: Vanishes for large p, as expected

iv. $MSFE(LP) \geq MSFE(VAR)$ for all other macro variables
Note: Also for low p

Iterated vs Direct Forecasts

VAR likely to be misspecified ($\neq DGP$) ➔ errors compounded at large horizons

LP are theoretically preferable because more robust to misspecification

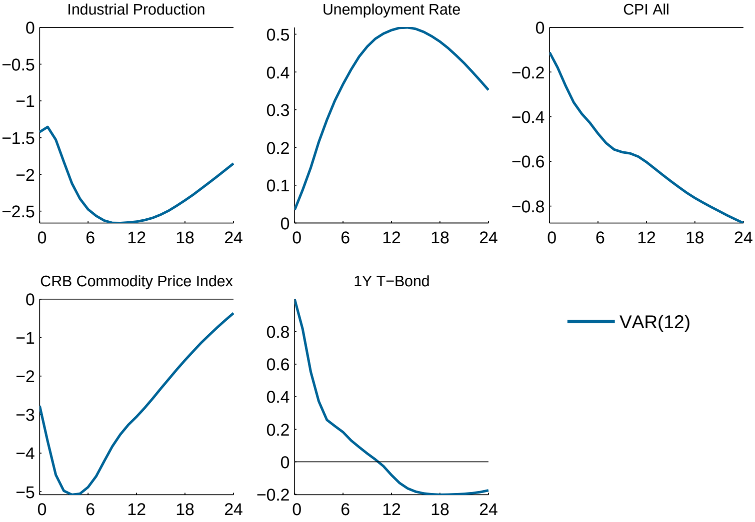
Theoretical gains are hardly realized in practice due to high estimation uncertainty

All else equal, LP deteriorate with small T and large h

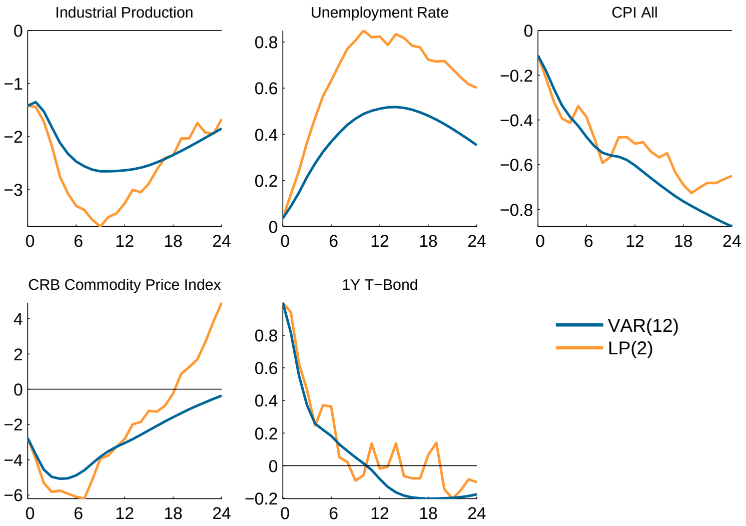
Bayesian Local Projection

Miranda-Agrippino and Ricco (2016)

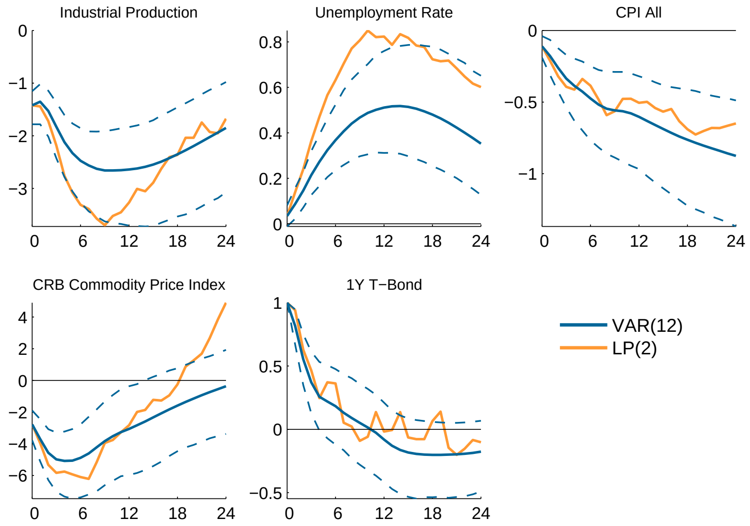
Iterated vs Direct IRFs



Iterated vs Direct IRFs

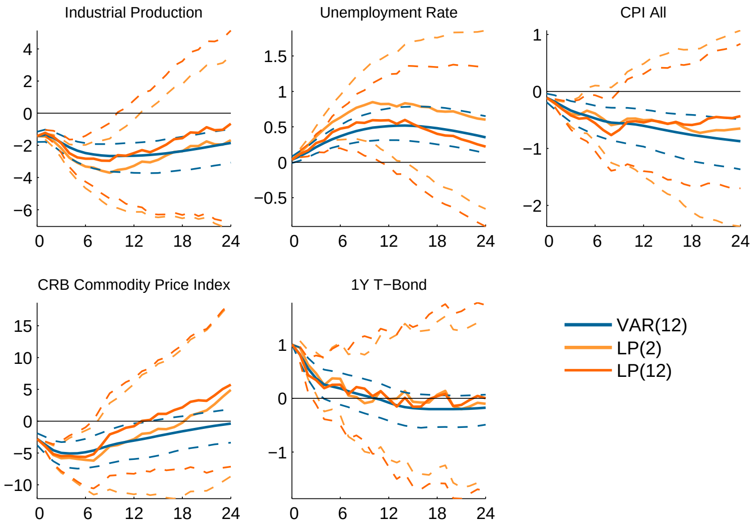


Iterated vs Direct IRFs



Iterated vs Direct IRFs

Iterated vs Direct IRFs



i. If the VAR adequately captures the DGP, IRFs are optimal and consistent. Implausible ➔ misspecification errors compound at higher h

ii. LP are more robust to model misspecification and preferable in theory. Not in practice ➔ high estimation uncertainty

Bias-Variance tradeoff is standard in Bayesian estimation

↓

BLP

↓

optimally bridge between VAR and LP

References: Miranda-Agrippino and Ricco (2016), The Transmission of Monetary Policy Shocks, mimeo, University of Warwick

Miranda-Agrippino and Ricco (2016), Bayesian Direct Methods, Technical Report

BLP: Intuition

- BLP priors give weight to the assumption that macro-variables behaviour is approximately linear and described by a VAR(p)
- Conjugate priors centred around the posterior mean of the VAR-based IRFs (pre-sample)
- The optimal level of informativeness of the priors is chosen to obtain optimal forecasts at all horizons [Giannone, Lenza, and Primiceri (2015)]

Alternative Priors: Additional hyperparameters for autoregressive coefficients

BLP: Intuition

BLP-IRF: Posterior Mean

$$B_{BLP}^{(h)} \quad \propto \left(X'X + \frac{1}{\lambda^{(h)}} \right)^{-1} \left(\frac{1}{\lambda^{(h)}} B_{VAR}^{h} + X'Y^{(h)} \right)$$

- $\lambda^{(h)}$ optimally balances between VAR and LP

  i. $\lambda^{(h)} \to 0 \quad \Rightarrow \quad B_{BLP}^{(h)} \to B_{VAR}^{h}$

  ii. $\lambda^{(h)} \to \infty \quad \Rightarrow \quad B_{BLP}^{(h)} = B_{LP}^{(h)}$

- $\lambda^{(h)}$ is looser the larger h ➔ LP more reliable if VAR is misspecified at large h

Digression: Bayesian VAR

VAR(p) in stacked and "vec" form:

$$\underset{[n\times 1]}{y_t} = \underset{[n\times n]}{\phi_1} y_{t-1} + \ldots + \phi_p y_{t-p} + u_t$$

$$\underset{[(T-p)\times n]}{Y} = \underset{[(T-p)\times k]}{X} \underset{[k\times n]}{B} + U$$

$$\underset{[n(T-p)\times 1]}{y} = \underset{[n(T-p)\times nk]}{(\mathbb{I}_n \otimes X)} \underset{[nk\times n]}{\beta} + u$$

- $k = n(p+1)$
- $y = \text{vec}(Y)$, $\beta = \text{vec}(B)$, $u = \text{vec}(U)$
- $u \sim N(0, \Sigma_u \otimes \mathbb{I}_n)$

Digression: Bayesian Procedure to Inference

i. Formulate a parametric model as a collection of probability distributions of all possible realization of the data Z conditional on different values of the model parameters $\theta \in \Theta$

$$\text{Model} : p(Z|\theta)$$

ii. Organize the beliefs about θ into a (prior) probability distribution over Θ

$$\text{Prior} : p(\theta)$$

iii. Collect the data z, treat them as realizations of Z, and insert them into the family of distributions

$$\text{Likelihood} : \mathcal{L}(z|\theta) = p(z|\theta)$$

iv. Use the Bayes theorem to update the new belief about θ

$$\text{Posterior} : p(\theta|y) \propto \mathcal{L}(z|\theta)p(\theta)$$

Natural Conjugate Priors: prior distributions that, once combined with the likelihood, result in posterior distributions of the same family

NORMAL - INVERSE WISHART PRIOR

$$\Sigma_u \sim IW\left(\Psi_0, d_0\right)$$

$$\beta|\Sigma_u \sim N\left(\beta_0, \Sigma_u \otimes \Omega_0(\lambda)\right)$$

- $\beta_0 \equiv \text{vec}(B)$ ➔ B : equations in columns
- $\lambda =$ prior tightness ➔ controls coefficients variance
- Ψ_0 and d_0: prior scale and degrees of freedom of the IW

VAR POSTERIOR

$$\Sigma_u | y \sim IW\left(\Psi, d\right)$$

$$\beta | \Sigma_u, y \sim N\left(\tilde{\beta}, \Sigma_u \otimes \Omega\right)$$

- $\tilde{\beta} \equiv \text{vec(b)}$

$$\underset{[k \times n]}{\text{b}} = \left(X'X + \Omega_0(\lambda)^{-1}\right)^{-1} \left(X'Y + \Omega_0(\lambda)^{-1}\beta_0\right)$$

- λ = prior tightness ➔ controls coefficients variance
  - i. $\lambda \to 0 \quad \Rightarrow \quad$ prior dominates $\Rightarrow \quad \text{b} \to \beta_0$
  - ii. $\lambda \to \infty \quad \Rightarrow \quad$ OLS dominates $\Rightarrow \quad \text{b} \to (X'X)^{-1}(X'Y)$
- Ψ_0 and d_0: prior scale and degrees of freedom of the IW

Hierarchical Modelling: Treat hyperparameters (λ) as additional unknown coefficients [Giannone, Lenza and Primiceri (2014)]

- $\theta \equiv$ model parameters: B, Σ_u
- $\gamma \equiv$ hyperparameters: λ (there might be more...)

i. specify prior distribution for γ

$$\text{Hyperprior : } p(\gamma)$$

ii. compute:

$$p(\gamma|y) \propto \underbrace{\int p(y|\theta,\gamma)p(\theta|\gamma)d\theta}_{p(y|\gamma)} \times p(\gamma)$$

$$\lambda^* = \text{argmax} p(y|\gamma)$$

under flat hyperprior

BLP PRIOR

$$\Sigma_v^{(h)} | \gamma^{(h)} \sim IW\left(\Psi_0^{(h)}, d_0\right)$$

$$\beta^{(h)} | \Sigma_v^{(h)}, \gamma^{(h)} \sim N\left(\underline{\beta_0^{(h)}}, \Sigma_v^{(h)} \otimes \Omega_0^{(h)}(\lambda^{(h)})\right)$$

Prior mean:

- $\beta^{(h)} \equiv \text{vec}(b^{(h)}) = \text{vec}\left(\left[\tilde{c}, \tilde{B}^{(h)}, \ldots, \tilde{B}_p^{(h)}\right]'\right)$
- $\beta_0^{(h)} = \beta_{T_0}^{(0,h)} = \text{vec}\left(b_{T_0}^{(0,h)}\right)$ ➔ posterior mean of VAR(p) coefficients iterated at h-horizon (pre-sample)

BLP PRIOR

$$\Sigma_v^{(h)}|\gamma^{(h)} \sim IW\left(\Psi_0^{(h)}, d_0\right)$$

$$\beta^{(h)}|\Sigma_v^{(h)}, \gamma^{(h)} \sim N\left(\beta_0^{(h)}, \Sigma_v^{(h)} \otimes \Omega_0^{(h)}(\lambda^{(h)})\right)$$

Prior variance:

- $\Psi_0^{(h)} = \mathrm{diag}\left(\left[(\sigma_1^{(h)})^2, \ldots, (\sigma_n^{(h)})^2\right]\right); \quad d_0 = n+2$
- $\underset{[(np+1)\times(np+1)]}{\Omega_0^{(h)}} = \begin{pmatrix} \epsilon^{-1} & 0 \\ 0 & \mathbb{I}_p \otimes \mathrm{diag}\left(\left[\lambda^{(h)}/\sigma_i^{(h)}\right]^2\right) \end{pmatrix}$
- $\mathbb{V}ar[(\tilde{B}^{(h)})_{ij}|\Sigma_v^{(h)}] = \left(\lambda^{(h)} \frac{\sigma_i^{(h)}}{\sigma_j^{(h)}}\right)^2$

$$y_{t+h} = \tilde{c} + \tilde{B}^{(h)} y_t + ... + \tilde{B}_p^{(h)} y_{t-p} + v_{t+h}$$

$$v_{t+h} \sim N\left(0, \Sigma_v^{(h)}\right) \qquad \forall \; h = 1, \ldots, H$$

$$v_{t+h} \sim MA(h-1)$$

- Frequentist solution: LS + HAC standard errors
- Our solution:
  i. misspecified likelihood ➔ $v_{t+h} \perp \mathrm{span}\{y_t, \ldots, y_{t-p}\}$
  ii. correction to posterior variance ➔ $\mathbb{E}\left[\Sigma_v^{(h)}\right] = \Sigma_{v,HAC}^{(h)}$

Alternative: fully specified VARMA likelihood

BLP POSTERIOR

$$\Sigma_v^{(h)} | \gamma^{(h)}, \mathrm{y} \sim IW\left(\Psi^{(h)}, d\right)$$

$$\beta^{(h)} | \Sigma_v^{(h)}, \gamma^{(h)}, \mathrm{y} \sim N\left(\tilde{\beta}^{(h)}, \Sigma_v^{(h)} \otimes \Omega^{(h)}\right)$$

Misspecified parametric model:

i. Posterior variance-covariance is underestimated ➔ ignores correlation in projection residuals

ii. Posterior beliefs constructed from a misspecified likelihood lead to inadmissible decisions [Müller (2013)] ➔ suboptimal prediction risk $\mathcal{R}(\hat{y}_{t+h})$

BLP HAC-POSTERIOR

$$\Sigma_{v,HAC}^{(h)}|\gamma^{(h)}, \mathrm{y} \sim IW\left(\Psi_{HAC}^{(h)}, d\right)$$

$$\beta^{(h)}|\Sigma_{v,HAC}^{(h)}, \gamma^{(h)}, \mathrm{y} \sim N\left(\tilde{\beta}^{(h)}, \Sigma_{v,HAC}^{(h)} \otimes \Omega^{(h)}\right)$$

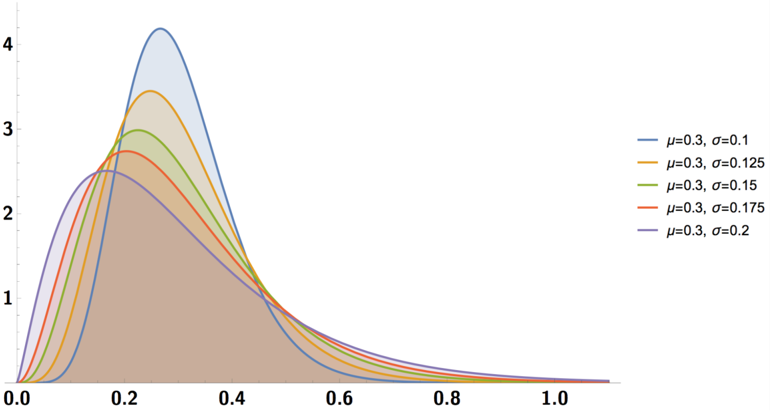
Inference based on an “artificial” Gaussian posterior centred at the MLE with the HAC covariance matrix [Müller (2013)]

Alternative: VARMA ➔ GLS estimator

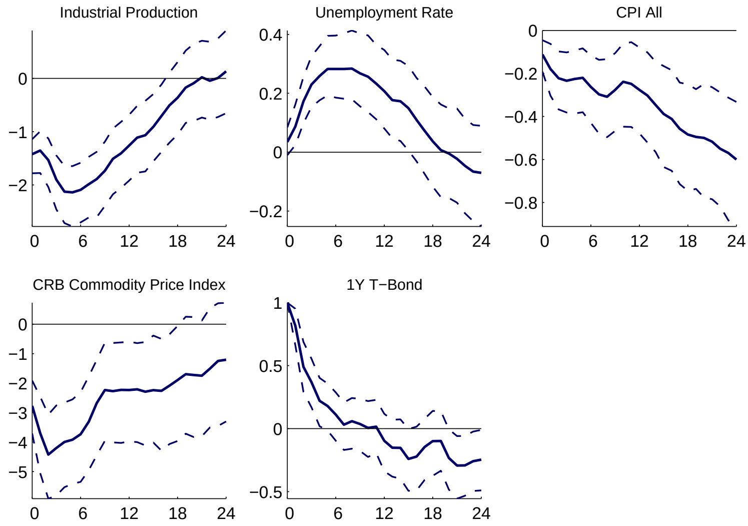
BLP: Hyperprior

$$\lambda^{(h)} \sim \Gamma\left(k^{(h)}, \theta^{(h)}\right)$$

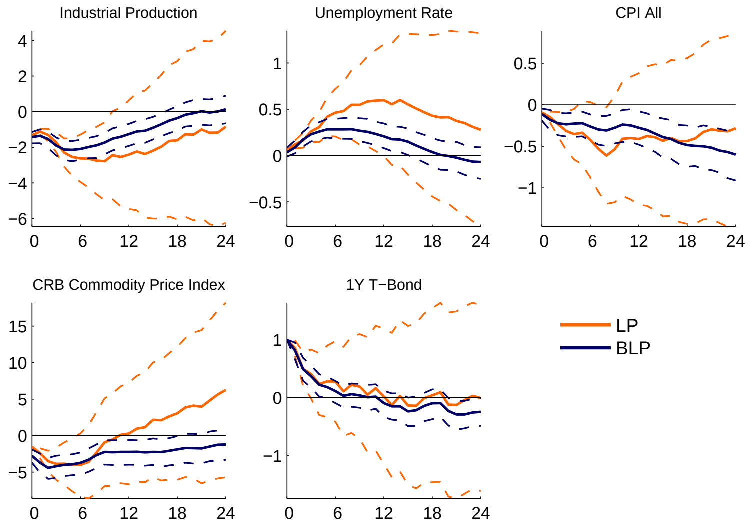
- mode $= 0.3$
- standard deviation $=$ logistic function over h



BLP: IRFs

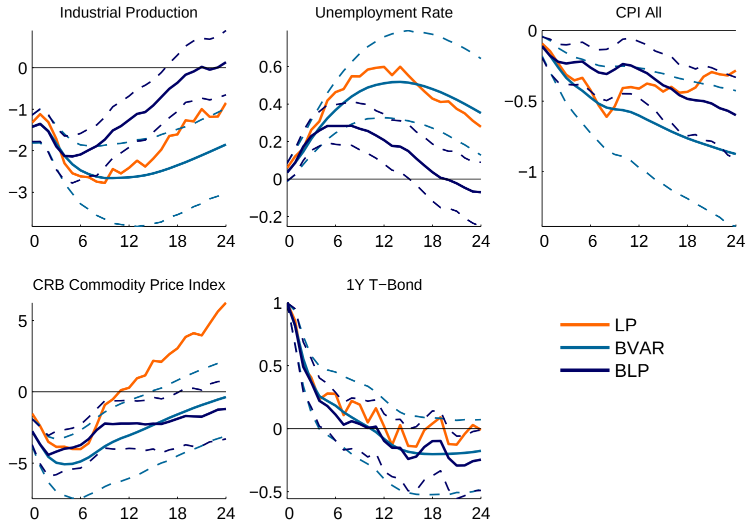

Industrial Production
Unemployment Rate
CPI All
CRB Commodity Price Index
1Y T−Bond
0
6
12
18
24
0
−1
−2
0.4
0.2
−0.2
−0.4
−0.6
−0.8
−3
−4
−5
1
0.5
−0.5

BLP: IRFs



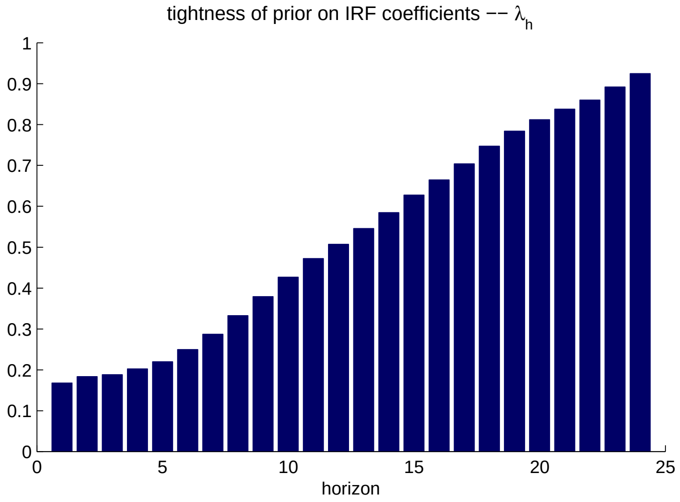
Industrial Production
Unemployment Rate
CPI All
CRB Commodity Price Index
1Y T−Bond
LP
BLP

BLP: IRFs



Industrial Production
Unemployment Rate
CPI All
CRB Commodity Price Index
1Y T−Bond
LP
BVAR
BLP

BLP: Prior Tightness

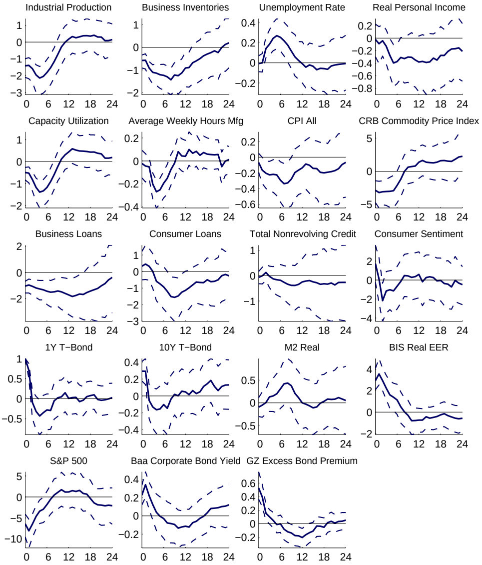


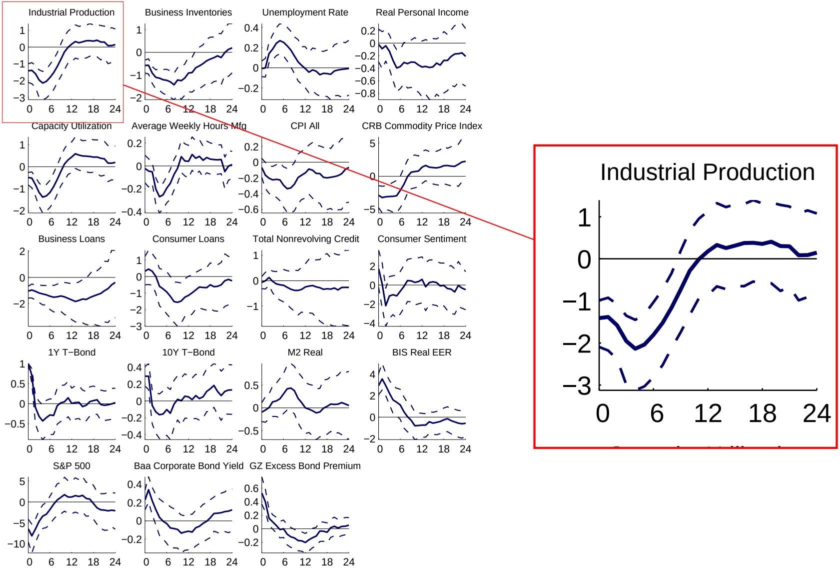
The Transmission of Monetary Policy Shocks

- Heterogeneous information set: $n = 19$
- Estimation: 1979 to 2014
- Identification consistent with information asymmetries/frictions

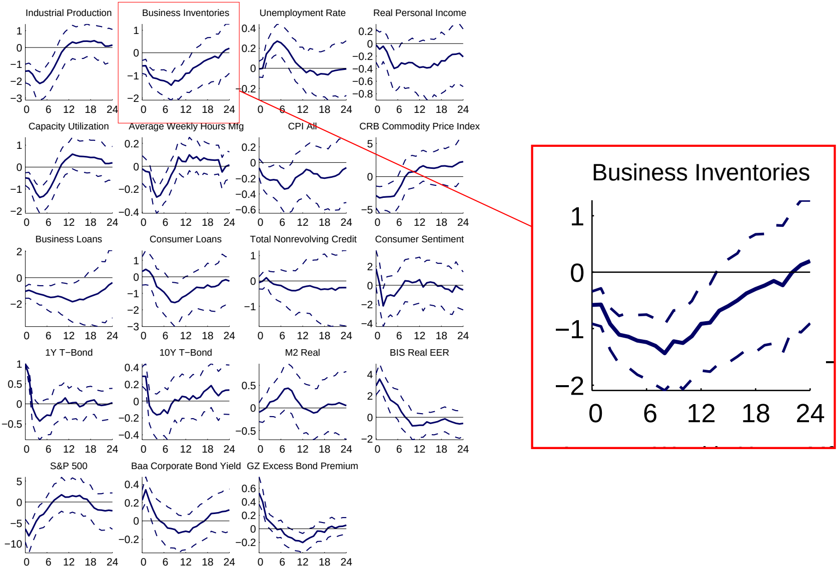
- 12 lags
- 90% posterior coverage bands

The Transmission of Monetary Policy Shocks

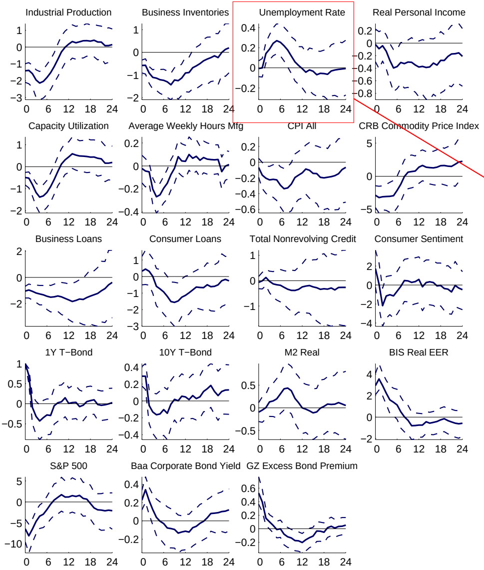


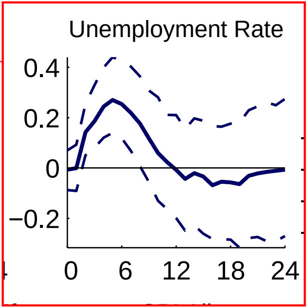

Industrial Production
Business Inventories
Unemployment Rate
Real Personal Income
Capacity Utilization
Average Weekly Hours Mfg
CPI All
CRB Commodity Price Index
Business Loans
Consumer Loans
Total Nonrevolving Credit
Consumer Sentiment
1Y T−Bond
10Y T−Bond
M2 Real
BIS Real EER
S&P 500
Baa Corporate Bond Yield
GZ Excess Bond Premium
Industrial Production

The Transmission of Monetary Policy Shocks

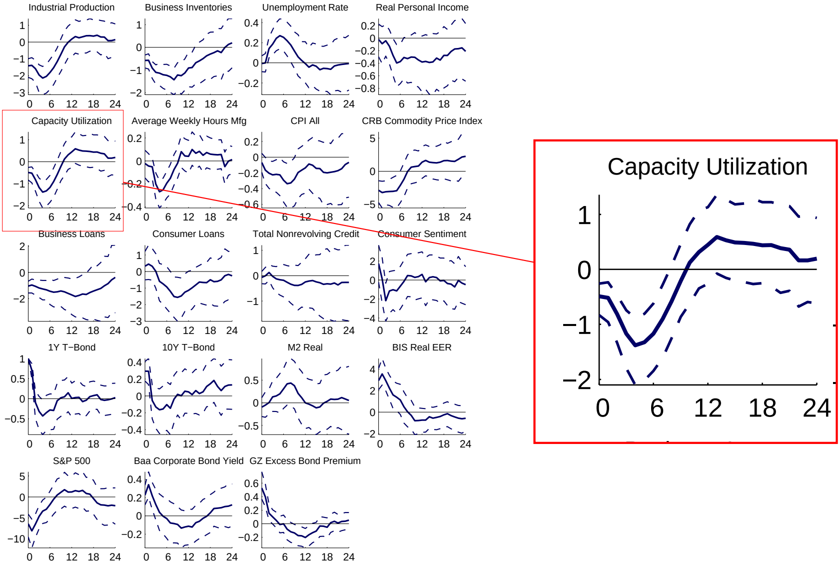


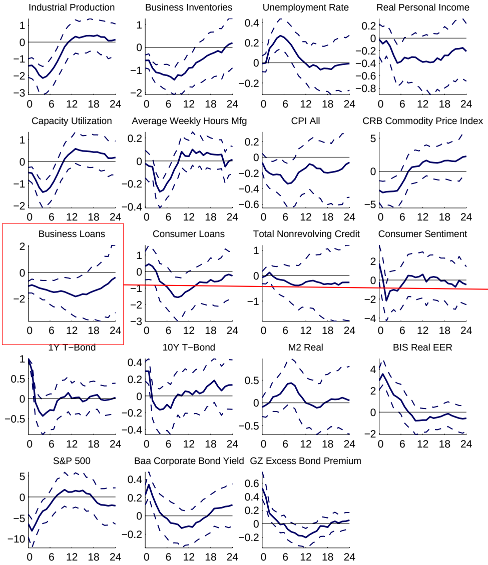
The Transmission of Monetary Policy Shocks

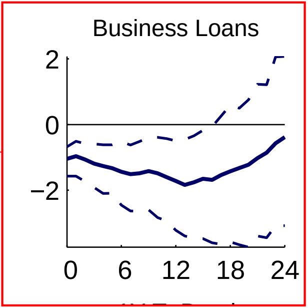




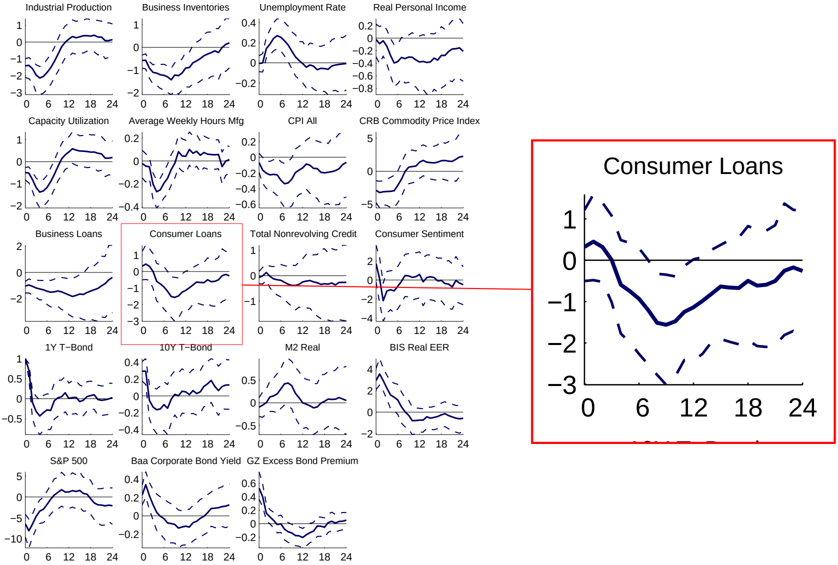
The Transmission of Monetary Policy Shocks



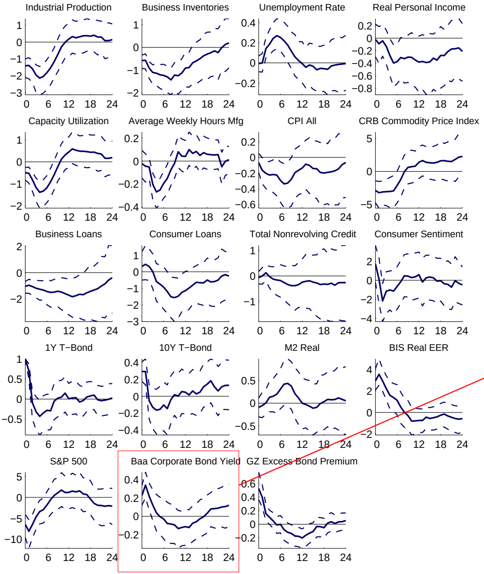

Industrial Production
Business Inventories
Unemployment Rate
Real Personal Income
Capacity Utilization
Average Weekly Hours Mfg
CPI All
CRB Commodity Price Index
Business Loans
Consumer Loans
Total Nonrevolving Credit
Consumer Sentiment
1Y T−Bond
10Y T−Bond
M2 Real
BIS Real EER
S&P 500
Baa Corporate Bond Yield
GZ Excess Bond Premium
0 6 12 18 24

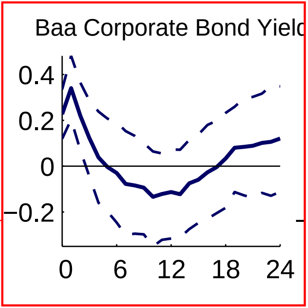

Business Loans
2
0
−2
0 6 12 18 24

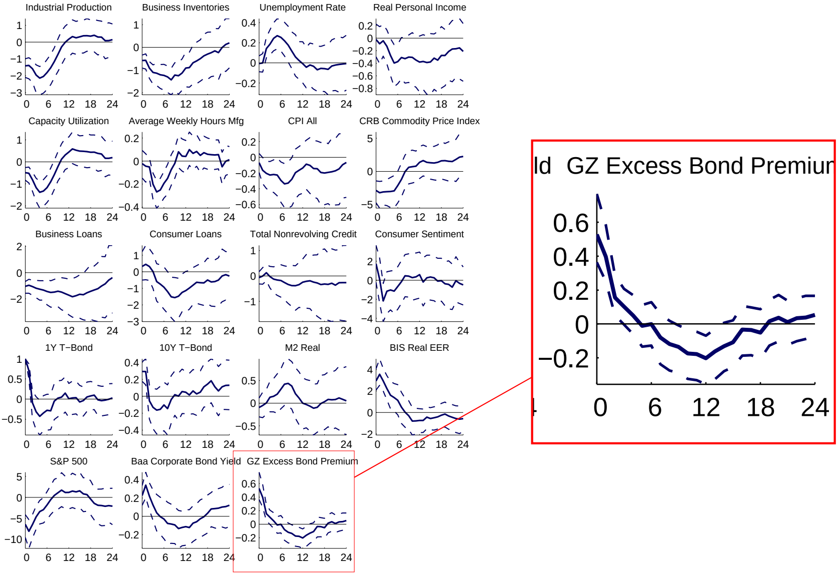
The Transmission of Monetary Policy Shocks



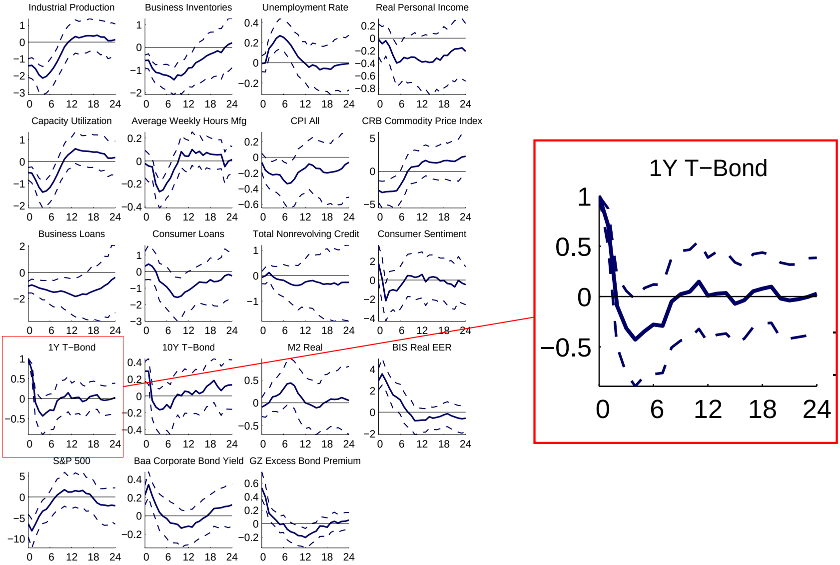
The Transmission of Monetary Policy Shocks



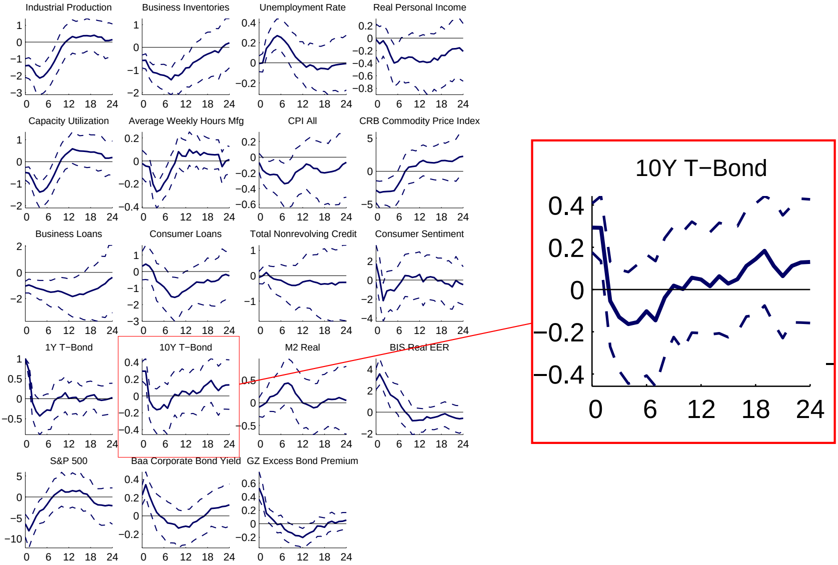



Industrial Production
Business Inventories
Unemployment Rate
Real Personal Income
Capacity Utilization
Average Weekly Hours Mfg
CPI All
CRB Commodity Price Index
Business Loans
Consumer Loans
Total Nonrevolving Credit
Consumer Sentiment
1Y T−Bond
10Y T−Bond
M2 Real
BIS Real EER
S&P 500
Baa Corporate Bond Yield
GZ Excess Bond Premium
GZ Excess Bond Premium

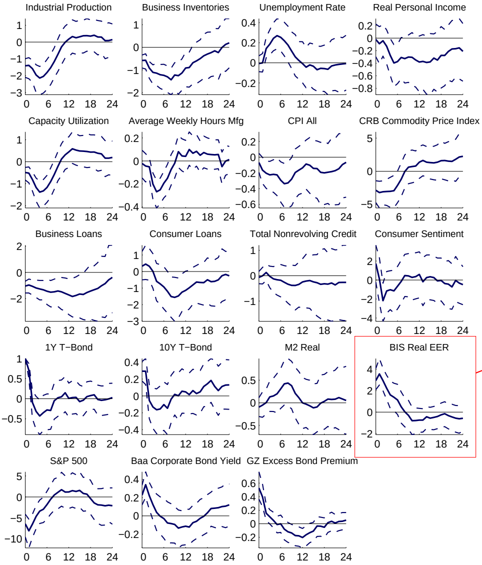
The Transmission of Monetary Policy Shocks

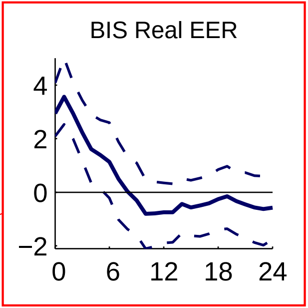


The Transmission of Monetary Policy Shocks

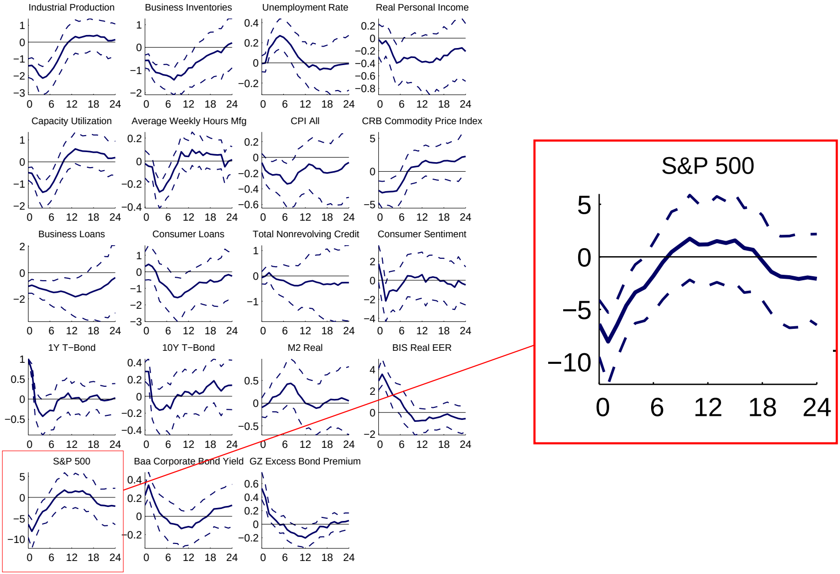


The Transmission of Monetary Policy Shocks





The Transmission of Monetary Policy Shocks



Selected References

Banbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010) "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, Vol. 25, No. 1, pp. 71–92.

Braun, Phillip A. and Stefan Mittnik (1993) "Misspecifications in vector autoregressions and their effects on impulse responses and variance decompositions," *Journal of Econometrics*, Vol. 59, No. 3, pp. 319–341, October.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri (2015) "Prior Selection for Vector Autoregressions," *The Review of Economics and Statistics*, Vol. 2, No. 97, pp. 436–451, May.

Jordà, Òscar (2005) "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, Vol. 95, No. 1, pp. 161–182, March.

Marcellino, Massimiliano, James H. Stock, and Mark W. Watson (2006) "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," *Journal of Econometrics*, Vol. 135, No. 1-2, pp. 499–526.