1 Introduction to Multivariate Models

Beyond univariate models...

- Consider the following AR(2) process for inflation (y_t)

$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \varepsilon_t \quad \varepsilon_t \sim WN$$

In this course we study multivariate generalizations of the above model. Let X_t be the GDP growth we will study model of the form

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} a_{11}^1 & a_{12}^1 \\ a_{21}^1 & a_{22}^1 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} a_{11}^2 & a_{12}^2 \\ a_{21}^2 & a_{22}^2 \end{pmatrix} \begin{pmatrix} Y_{t-2} \\ X_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Both Y_t and X_t are endogenous and depend on lagged values.

- Multivariate time series models are useful for many kinds of analysis and applications.

1. Forecasting macroeconomic variables of interest.

2. Studying the dynamic interrelationships between a number of variables.

3. Studying the effects of some economic shock of interest. What are the effects of monetary policy shocks? What kind of shocks drive the business cycle?

In general the empirical evidence obtained using these models can provide useful information to policymakers and macroeconomic theorists.

Some Preliminary Definitions and Concepts

Random Vector: A vector $X = (X_1, ..., X_n)'$ whose components are scalar-valued random variables on the same probability space.

Vector Random Process: A family of random vectors $\{X_t, t \in T\}$ indexed by t where T is a set of time points. Typically T is the set of natural or integers numbers.

Time Series Vector: A particular realization of random process.

White noise: A n-dimensional vector white noise $\epsilon_t' = [\epsilon_{1t}, ..., \epsilon_{nt}] \sim WN(0, \Omega)$ is such if $E(\epsilon_t) = 0$ and $E(\epsilon_t \epsilon_\tau') = \Omega$ (Ω a symmetric positive definite matrix) if $t = \tau$ and 0 otherwise. If also $\epsilon_t \sim N$ the process is a Gaussian WN.

Important: A vector whose components are white noise is not necessarily a white noise. Example: let u_t be a scalar white noise and define $\epsilon_t = (u_t, u_{t-1})'$. Then $E(\epsilon_t \epsilon_t') = \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}$ and $E(\epsilon_t \epsilon_{t-1}') = \begin{pmatrix} 0 & 0 \\ \sigma_u^2 & 0 \end{pmatrix}$.

Matrix of polynomials in the lag operator: $\Phi(L)$ if its elements are polynomial in the lag operator, i.e.

$$\Phi(L) = \Phi_0 L^0 + \Phi_1 L^1 + \Phi_2 L^2 + ... + \Phi_p L^p$$

and by definition of L when applied to vector the X_t

$$\Phi(L)X_t = \Phi_0 X_t + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + ... + \Phi_p X_{t-p}$$

Note that $\Phi(0) = \Phi_0$ and $\Phi(1) = \Phi_0 + \Phi_1 + \Phi_2 + ... + \Phi_p$

Example

$$\Phi(L) = \begin{pmatrix} 1 & -0.5L \\ L & 1+L \end{pmatrix} = \Phi_0 + \Phi_1 L$$

where

$$\Phi_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \ \Phi_1 = \begin{pmatrix} 0 & -0.5 \\ 1 & 1 \end{pmatrix}, \ \ \Phi_{j>1} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

When applied to a vector X_t we obtain

$$\Phi(L)X_t \quad = \quad \begin{pmatrix} 1 & -0.5L \\ L & 1+L \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} X_{1t} - 0.5X_{2t-1} \\ X_{1t-1} + X_{2t} + X_{2t-1} \end{pmatrix}$$

Covariance Stationarity Let Y_t be a n-dimensional random vector, $Y_t' = [Y_{1t}, ..., Y_{nt}]$. Then Y_t is covariance (weakly) stationary if $E(Y_t) = \mu$, and the *autocovariance* matrix $E(Y_t - \mu)(Y_{t-j} - \mu)' = \Gamma_j$ for all t, j, that is are independent of t and both finite.

- Stationarity of each of the components of Y_t does not imply stationarity of the vector Y_t. Stationarity in the vector case requires that the components of the vector are stationary and costationary.

- Although $\gamma_j = \gamma_{-j}$ for a scalar process, the same is not true for a vector process. The correct relation is

$$\Gamma_j' = \Gamma_{-j}$$

VMA processes

Given the n-dimensional vector White Noise ϵ_t a vector moving average of order q is defined as

$$Y_t = \mu + \epsilon_t + C_1\epsilon_{t-1} + ... + C_q\epsilon_{t-q}$$

where C_j are $n \times n$ matrices of coefficients and μ is the mean of Y_t.

The VMA(1) Let us consider the VMA(1)

$$Y_t = \mu + \epsilon_t + C_1\epsilon_{t-1}$$

with $\epsilon_t \sim WN(0, \Omega)$. The variance of the process is given by

$$\begin{aligned} \Gamma_0 &= E[(Y_t - \mu)(Y_t - \mu)'] \\ &= \Omega + C_1\Omega C_1' \end{aligned}$$

with autocovariances

$$\Gamma_1 = C_1\Omega, \;\; \Gamma_{-1} = \Omega C_1', \;\; \Gamma_j = 0 \;\text{ for }\; |j| > 1$$

The VMA(q) Let us consider the VMA(q)

$$Y_t = \mu + \epsilon_t + C_1\epsilon_{t-1} + ... + C_q\epsilon_{t-q}$$

with $\epsilon_t \sim WN(0, \Omega)$, μ is the mean of Y_t. The variance of the process is given by

$$\begin{aligned} \Gamma_0 &= E[(Y_t - \mu)(Y_t - \mu)'] \\ &= \Omega + C_1\Omega C_1' + C_2\Omega C_2' + ... + C_q\Omega C_q' \end{aligned}$$

with autocovariances

$$\begin{aligned} \Gamma_j &= C_j\Omega + C_{j+1}\Omega C_1' + C_{j+2}\Omega C_2' + ... + C_q\Omega C_{q-j}' \quad for \quad j = 1, 2, ..., q \\ \Gamma_j &= 0 \quad for \quad |j| > q \end{aligned}$$

The VMA(∞) A useful process, as we will see, is the VMA(∞)

$$Y_t = \mu + \sum_{j=0}^{\infty} C_j\varepsilon_{t-j} \tag{1}$$

If the sequence $\{C_j\}$ is absolutely summable, i.e. $\sum_{j=0}^{\infty} |C_{m,n,j}| < \infty$ (for all m, n elements of C_j), then the infinite sequence above generates a well defined (mean square convergent) process. The autocovariances are

$$\Gamma_s = \sum_{v=0}^{\infty} C_{s+v}\Omega C_v'$$

for $s = 0, 1, 2, ...$ That is they can be found by taking the limit of the autocovariance of an $MA(q)$.

Invertibility A MA(q) process defined by the equation $Y_t = C(L)\varepsilon_t$ is said to be invertible is there exists a sequence of absolutely summable matrices of constants $\{A_j\}_{j=0}^{\infty}$ such that $\sum_{j=0}^{\infty} A_j Y_{t-j} = \varepsilon_t$.

Proposition A MA process defined by the equation $Y_t = C(L)\varepsilon_t$ is invertible if and only if the determinant of $C(L)$ vanishes only outside the unit circle, i.e. if $det(C(z)) \neq 0$ for all $|z| \leq 1$.

If the process is invertible it possesses a unique VAR representation (clear later on).

Example Consider the process

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} 1 & L \\ 0 & \theta - L \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

$det(C(z)) = \theta - z$ which is zero for $z = \theta$. Obviously the process is invertible if and only if $|\theta| > 1$.

Funamentalness The VMA is *fundamental* if and only if the $det(C(z)) \neq 0$ for all $|z| < 1$. In the previous example the process is fundamental if and only if $|\theta| \geq 1$. In the case $|\theta| = 1$ the process is fundamental but noninvertible.

Provided that $|\theta| > 1$ the MA process can be inverted and the shock can be obtained as a combination of present and past values of Y_t. In fact

$$\begin{pmatrix} 1 & -\frac{L}{\theta - L} \\ 0 & \frac{1}{\theta - L} \end{pmatrix} \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Notice that for any noninvertible process with determinant that does not vanish on the unit circle there is an invertible process with identical autocovariance structure.

Wold Decomposition Any zero-mean stationary vector process Y_t admits the following representation

$$Y_t = C(L)\varepsilon_t + \kappa_t \tag{2}$$

where $C(L)\epsilon_t$ is the stochastic component with $C(L) = \sum_{i=0}^{\infty} C_i L^i$ and κ_t the purely deterministic component, the one perfectly predictable using linear combinations of past Y_t.

If $\kappa_t = 0$ the process is said *regular*. Here we only consider regular processes.

(2) represents the *Wold representation* of Y_t which is unique and for which the following properties hold:

(a) ϵ_t is innovation for Y_t, i.e. $\epsilon_t = Y_t - \text{Proj}(Y_t|Y_{t-1}, Y_{t-1}, ...)$.

(b) ϵ_t is White noise, $E\epsilon_t = 0$, $E\epsilon_t\epsilon_\tau' = 0$, for $t \neq \tau$, $E\epsilon_t\epsilon_t' = \Omega$

(c) The coefficients are square summable $\sum_{j=0}^{\infty} C_{mn}^2 < \infty$ for all m, n.

(d) $C_0 = I$

The result is very powerful since holds for any covariance stationary process.

However the theorem does not implies that (2) is the *true* representation of the process. For instance the process could be stationary but non-linear or non-invertible.

Other fundamental MA(∞) Representations It is easy to extend the Wold representation to the general class of fundamental MA(∞) representations. For any non singular matrix R of constant we have

$$\begin{aligned} Y_t &= C(L)Ru_t \\ &= D(L)u_t \end{aligned}$$

where $u_t \sim WN(0, R^{-1}\Omega R^{-1\prime})$ and $u_t = R^{-1}\epsilon_t$.

Fundamentalness is ensured since u_t is a linear combination of the Wold shocks. The roots of the determinant of $D(L)$ will coincide with those of $C(L)$. In fact, $det(C(L)R) = det(C(L))det(R)$. Therefore if $det(C(L)) \neq 0 \ \forall |z| < 1$ so will $det(C(L)R)$.

Impulse Response Functions

Impulse response functions represent the mechanisms through which shock spread over time. Let us consider the Wold representation of a covariance stationary VAR(p),

$$\begin{aligned} Y_t &= C(L)\epsilon_t \\ &= \sum_{i=0}^{\infty} C_i \epsilon_{t-i} \end{aligned} \tag{3}$$

The matrix C_j has the interpretation

$$\frac{\partial Y_t}{\partial \epsilon'_{t-j}} = C_j \tag{4}$$

or

$$\frac{\partial Y_{t+j}}{\partial \epsilon'_t} = C_j \tag{5}$$

That is, the row i, column k element of C_j identifies the consequences of a unit increase in the kth variable's innovation at date t for the value of the ith variable at time $t+j$ holding all other innovation at all dates constant.

Example 1 Let us assume that the estimated matrix of VAR coefficients is

$$A = \begin{pmatrix} 0.8 & 0.1 \\ -0.2 & 0.5 \end{pmatrix} \tag{6}$$

with eigenvalues 0.8562 and 0.4438. We generate impulse response functions of the Wold representation

$$C_j = A^j$$

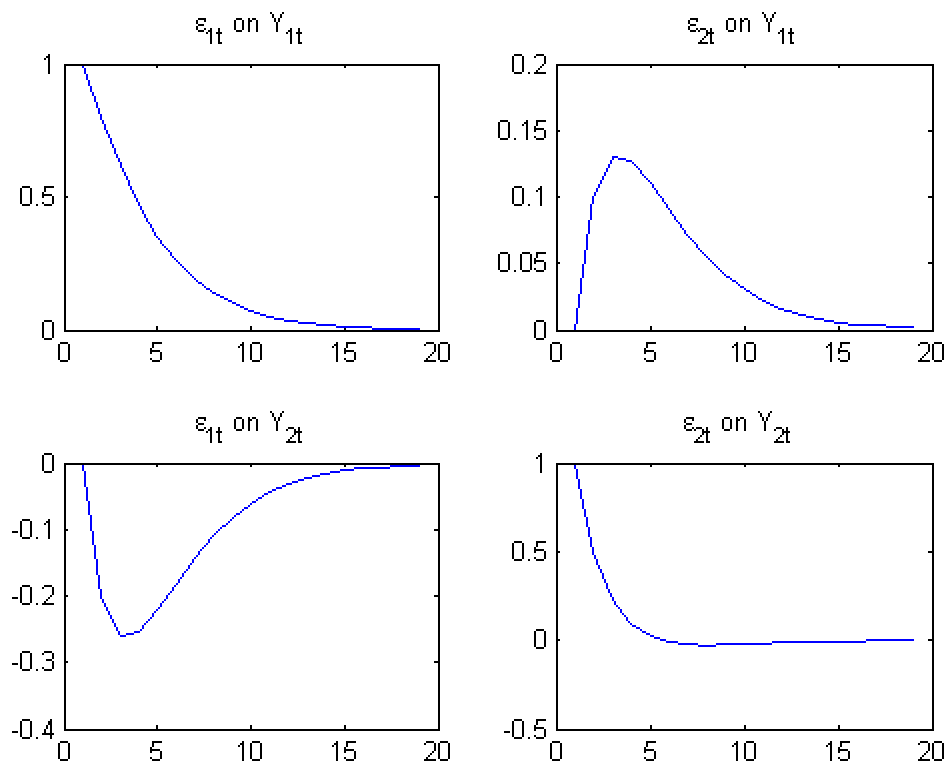


Figure 2: Impulse response functions

Example 2 Let us now assume that

$$A = \begin{pmatrix} 0.8 & 0 \\ -0.2 & 0.5 \end{pmatrix} \tag{7}$$

with eigenvalues 0.8 and 0.5. Impulse response functions are plotted in the next figure

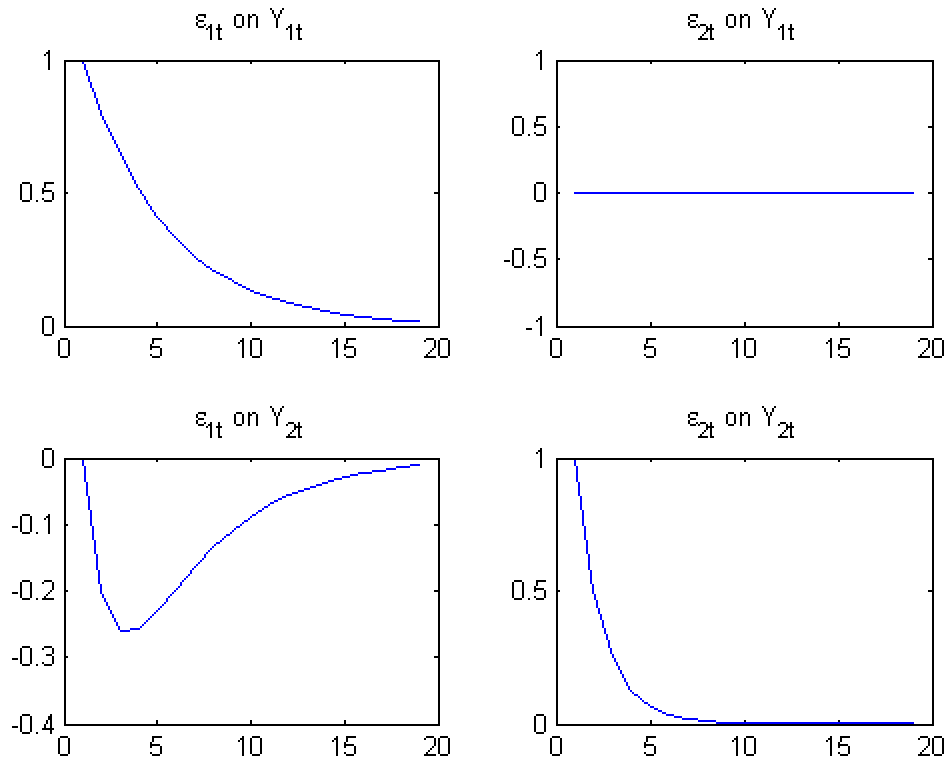


Figure 3.

Cumulated impulse response functions Suppose Y_t is a vector of trending variables (i.e. log prices and output) so we consider the first difference to reach stationarity. So the model is

$$\Delta Y_t = (1-L)Y_t = C(L)\varepsilon_t$$

We know how to estimate, interpret, and conduct inference on $C(L)$. But suppose we are interested in the response of the levels of Y_t rather than their first differences (the level of and prices rather than their growth rates). How can we find these responses? We transform the model

$$Y_t = Y_{t-1} + C(L)\varepsilon_t$$

The effect of ε_t on Y_t is C_0. Now substituting forward we obtain

$$\begin{aligned} Y_{t+1} &= Y_{t-1} + C(L)\varepsilon_t + C(L)\varepsilon_{t+1} \\ &= Y_{t-1} + C_0\varepsilon_{t+1} + (C_0 + C_1)\varepsilon_t + ... \end{aligned} \tag{8}$$

and for two periods ahead

$$\begin{aligned} Y_{t+2} &= Y_{t-1} + C(L)\varepsilon_t + C(L)\varepsilon_{t+1} + C(L)\varepsilon_{t+2} \\ &= Y_{t-1} + C_0\varepsilon_{t+2} + (C_0 + C_1)\varepsilon_{t+1} + (C_0 + C_1 + C_2)\varepsilon_t ... \end{aligned} \tag{9}$$

so the effect of ε_t on Y_{t+1} are $(C_0 + C_1)$ and on Y_{t+2} is $(C_0 + C_1 + C_2)$. In general the effects of ε_t on Y_{t+j} are

$$\tilde{C}_j = C_0 + C_1 + ... + C_j$$

defined as cumulated impulse response functions.

2 Reduced Forms VARs

2.1 Representations

The Vector Autoregressive (VAR) Representation

If the MA matrix of lag polynomials is invertible, then a unique VAR exists.

We define $C(L)^{-1}$ as an $(n \times n)$ lag polynomial matrix such that $C(L)^{-1}C(L) = I$. This operation in effect converts lags of the errors into lags of the vector of dependent variables.

Thus we move from MA coefficient to VAR coefficients.

Define $A(L) = C(L)^{-1.}$ Then given the (invertible) matrix of MA coefficients, it is easy to map these into the VAR coefficients:

$$\begin{aligned} Y_t &= C(L)\epsilon_t \\ A(L)Y_t &= \epsilon_t \end{aligned} \tag{10}$$

where $A(L) = A_0L^0 - A_1L^1 - A_2L^2 - ...$ and A_j for all j are $(n \times n)$ matrices of coefficients.

To show that this matrix lag polynomial exists and how it maps into the coefficients in $C(L)$, note that by definition

$$(A_0 - A_1L^1 - A_2L^2 - ...)(I + C_1L^1 + C_2L^2 + ...) = I$$

After distributing, the identity implies that coefficients on the lag operators must be zero, which implies the following recursive solution for the VAR coefficients:

$$\begin{aligned} A_0 &= I \\ A_1 &= A_0 C_1 \\ &\vdots \end{aligned}$$

As noted, the VAR is possibly of infinite order (i.e. infinite number of lags required to fully represent joint density). In practice, the VAR is usually restricted for estimation by truncating the lag-length.

The *pth-order vector autoregression*, denoted VAR(p) is given by

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \epsilon_t \tag{11}$$

Note: Here we are considering zero mean processes. In case the mean of Y_t is not zero we should add a constant in the VAR equations.

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \epsilon_t \tag{12}$$

Alternative representations: Any VAR(p)

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \epsilon_t \tag{13}$$

can be rewritten as a VAR(1). To form a VAR(1) from the general model we define: $\mathbf{e}'_t = [\epsilon', 0, ..., 0]$, $\mathbf{Y}'_t = [Y'_t, Y'_{t-1}, ..., Y'_{t-p+1}]$

$$\mathbf{A} = \begin{pmatrix} A_1 & A_2 & ... & A_{p-1} & A_p \\ I_n & 0 & ... & 0 & 0 \\ 0 & I_n & ... & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & ... & ... & I_n & 0 \end{pmatrix}$$

$$\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Therefore we can rewrite the VAR(p) as a VAR(1)

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{e}_t$$

This is also known as the companion form of the VAR(p)

Example: Suppose $n = 2$ and $p = 2$. The VAR will be

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} a_{11}^1 & a_{12}^1 \\ a_{21}^1 & a_{22}^1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} a_{11}^2 & a_{12}^2 \\ a_{21}^2 & a_{22}^2 \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ x_{2t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

We can rewrite the above model as

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{1t-1} \\ y_{2t-1} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} a_{11}^1 & a_{12}^1 & a_{11}^2 & a_{12}^2 \\ a_{21}^1 & a_{22}^1 & a_{21}^2 & a_{22}^2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{1t-2} \\ x_{2t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ 0 \\ 0 \end{pmatrix}$$

Setting

$$\mathbf{Y}_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{1t-1} \\ y_{2t-1} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_{11}^1 & a_{12}^1 & a_{11}^2 & a_{12}^2 \\ a_{21}^1 & a_{22}^1 & a_{21}^2 & a_{22}^2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{e}_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ 0 \\ 0 \end{pmatrix}, \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ 0 \\ 0 \end{pmatrix}$$

we obtain the previous VAR(1) representation.

The VAR(p) can be stacked as

$$\mathbf{Y} = \mathbf{X}\Pi + \mathbf{u}$$

where

- $\mathbf{X} = [X_1, ..., X_T]'$ is a $T \times (np + 1)$ matrix
- $X_t = [1, Y'_{t-1}, Y'_{t-2} ..., Y'_{t-p}]'$ is a $(np + 1)$ vector,
- $\mathbf{Y} = [Y_1, ..., Y_T]'$ a $T \times n$ matrix,
- $\mathbf{u} = [\epsilon_1, ..., \epsilon_T]'$ a $T \times n$ vector
- $\Pi = [cA_1, ..., A_p]'$ a $(np + 1) \times n$ matrix of coefficients

Example Again for $n = 2$ $p = 2$ $\mathsf{X}_t = \begin{pmatrix} 1 \\ Y_{1,t-1} \\ Y_{2,t-1} \\ Y_{1,t-2} \\ Y_{2,t-2} \end{pmatrix}$

Stationarity of a VAR

Stability and stationarity Consider the VAR(1)

$$Y_t = c + AY_{t-1} + \varepsilon_t$$

Substituting backward we obtain

$$\begin{aligned}
Y_t &= c + AY_{t-1} + \varepsilon_t \\
&= c + A(c + AY_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= (I + A)c + A^2 Y_{t-2} + A\varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
Y_t &= (I + A + ... + A^{j-1})c + A^j Y_{t-j} + \sum_{i=0}^{j-1} A^i \varepsilon_{t-i}
\end{aligned}$$

If all the eigenvalues of A (the elements of the diagonal matrix Λ) are smaller than one in modulus then

1. $A^j = P\Lambda^j P^{-1} \to 0$.

2. the sequence A^i, $i = 0, 1, ...$ is absolutely summable.

3. the infinite sum $\sum_{i=0}^{j-1} A^i \varepsilon_{t-i}$ exists in mean square (by proposition C.10L);

4. $(I + A + ... + A^j)c \to (I - A)^{-1}c$ and $A^j \to 0$ as j goes to infinity.

Therefore if the eigenvalues are smaller than one in modulus then Y_t has the following representation

$$Y_t = (I - A)^{-1}c + \sum_{i=0}^{\infty} A^i \varepsilon_{t-i}$$

Note that the the eigenvalues correspond to the reciprocal of the roots of the determinant of $A(z) = I - Az$. A VAR(1) is called *stable* if

$$\det(I - Az) \neq 0 \text{ for } |z| \leq 1.$$

For a VAR(p) the stability condition also requires that all the eigenvalues of $\mathbf{A}$ (the AR matrix of the companion form of Y_t) are smaller than one in modulus. Therefore we have that a VAR(p) is called *stable* if

$$\det(I - A_1 z - A_2 z^2 - ... - A_p z^p) \neq 0 \text{ for } |z| \leq 1.$$

A condition for stationarity: A stable VAR process is stationary.

Notice that the converse is not true. An unstable process can be stationary.

Example A stationary VAR(1)

$$Y_t = AY_{t-1} + \epsilon_t,\ A = \begin{pmatrix} 0.5 & 0.3 \\ 0.02 & 0.8 \end{pmatrix},\ \Omega = E(\epsilon_t \epsilon_t') = \begin{pmatrix} 1 & 0.3 \\ 0.3 & .1 \end{pmatrix},\ \lambda = \begin{pmatrix} 0.81 \\ 0.48 \end{pmatrix}$$

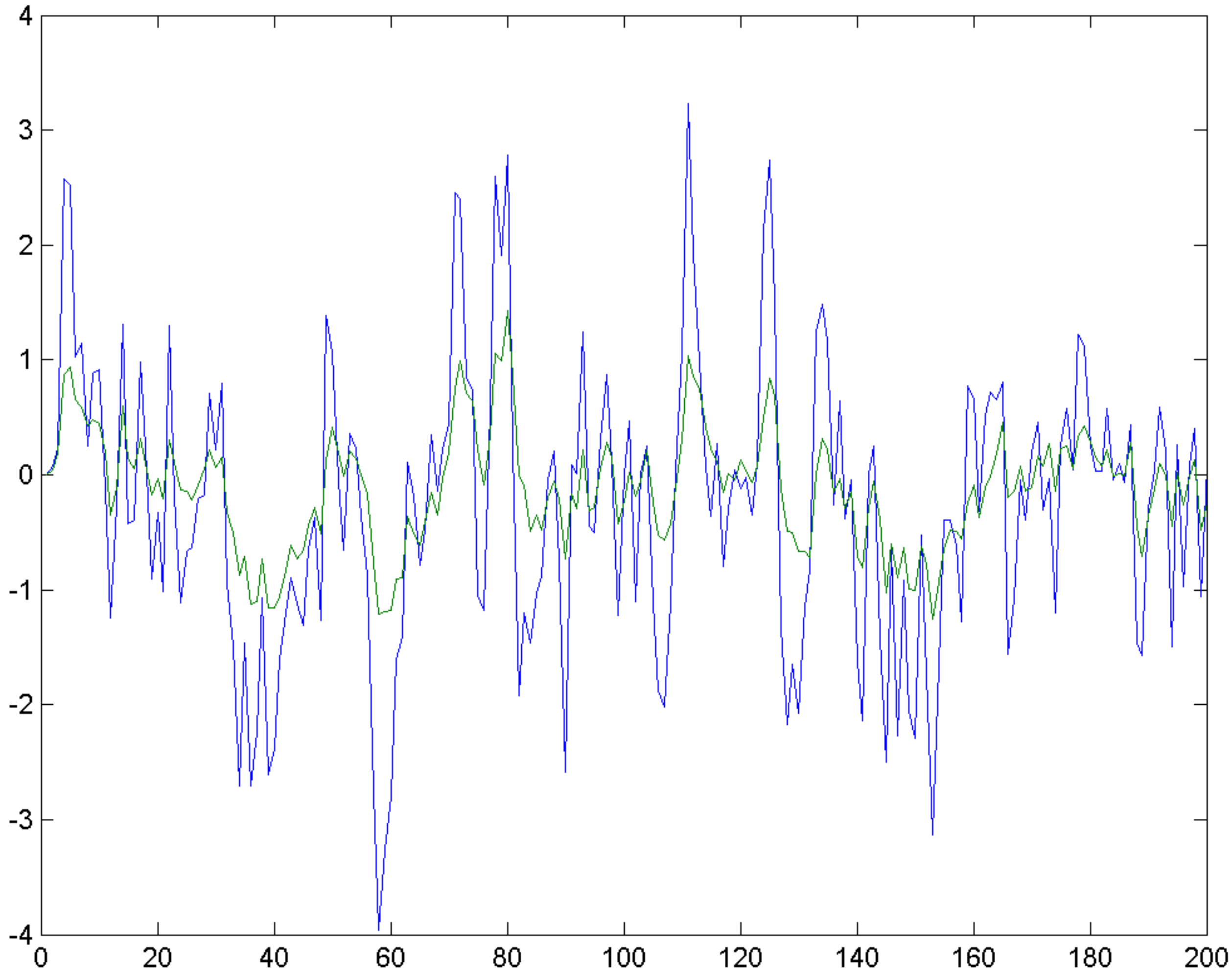


Figure 1: Blu: Y_1, green Y_2.

From a VAR to the Wold Representation

Rewriting the VAR(p) as a VAR(1) it is particularly useful in order to find the Wold representation of Y_t.

We know how to find the MA(∞) representation of a stationary AR(1).

We can proceed similarly for the VAR(1). Substituting backward in the companion form we have

$$\mathbf{Y}_t = \mathbf{A}^j \mathbf{Y}_{t-j} + \mathbf{A}^{j-1}\mathbf{e}_{t-j+1} + ... + \mathbf{A}^1 \mathbf{e}_{t-1} + \mathbf{e}_t$$

If conditions for stationarity are satisfied, the series $\sum_{i=1}^{\infty} \mathbf{A}^j$ converges and Y_t has an VMA(∞) representation in terms of the Wold shock $\mathbf{e}_t$ given by

$$\begin{aligned} \mathbf{Y}_t &= (I - \mathbf{A}L)^{-1}\mathbf{e}_t \\ &= \sum_{i=1}^{\infty} \mathbf{A}^j \mathbf{e}_{t-j} \\ &= \mathbf{C}(L)\mathbf{e}_t \end{aligned}$$

where $\mathbf{C}_0 = \mathbf{A}_0 = I$, $\mathbf{C}_1 = \mathbf{A}_1, \mathbf{C}_2 = \mathbf{A}^2, ..., \mathbf{C}_k = \mathbf{A}^k$. C_j will be the $n \times n$ uuper left matrix of $\mathbf{C}_j$.

2.2 Estimation

Conditional Likelihood

Let us condsider the VAR(p)

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \epsilon_t \tag{14}$$

with $\epsilon_t \sim i.i.dN(0, \Omega)$. Suppose we have a sample of $T + p$ observations for such variables. Conditioning on the first p observations we can form the conditional likelihood

$$f(Y_T, Y_{T-1}, ..., Y_1 | Y_0, Y_{-1}, ..., Y_{-p+1}, \theta) \tag{15}$$

where θ is a vector containing all the parameters of the model. We refer to (2) as "conditional likelihood function".

The joint density of observations 1 through t conditioned on $Y_0, ..., Y_{-p+1}$ satisfies

$$\begin{aligned} f(Y_t, Y_{t-1}, ..., Y_1 | Y_0, Y_{-1}, ..., Y_{-p+1}, \theta) \quad &= \quad f(Y_{t-1}, ..., Y_1 | Y_0, Y_{-1}, ..., Y_{-p+1}, \theta) \\ &\quad\ \times f(Y_t | Y_{t-1}, ..., Y_1, Y_0, Y_{-1}, ..., Y_{-p+1}, \theta) \end{aligned}$$

Applying the formula recursively, the likelihood for the full sample is the product of the individual conditional densities

$$f(Y_t, Y_{t-1}, ..., Y_1 | Y_0, Y_{-1}, ..., Y_{-p+1}, \theta) \quad = \quad \prod_{t=1}^{T} f(Y_t | Y_{t-1}, Y_{t-2}, ..., Y_{-p+1}, \theta) \tag{16}$$

At each t, conditional on the values of Y through date $t-1$

$$Y_t|Y_{t-1}, Y_{t-2}, ..., Y_{-p+1} \sim N(c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p}, \Omega)$$

Recall

$$X_t = \begin{pmatrix} 1 \\ Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{pmatrix}$$

is an $(np+1 \times 1)$ vector and let $\Pi' = [c, A_1, A_2, ..., A_p]$ be an $(n \times np+1)$ matrix of coefficients. Using this notation we have that

$$Y_t|Y_{t-1}, Y_{t-2}, ..., Y_{-p+1} \sim N(\Pi' X_t, \Omega)$$

Thus the conditional density of the tth observation is

$$\begin{aligned} f(Y_t|Y_{t-1}, Y_{t-2}, ..., Y_{-p+1}, \theta) = (2\pi)^{-n/2} \left|\Omega^{-1}\right|^{1/2} \\ \exp\left[(-1/2)(Y_t - \Pi' X_t)' \Omega^{-1} (Y_t - \Pi' X_t)\right] \end{aligned} \tag{17}$$

The sample log-likelihood is found by substituting (4) into (3) and taking logs

$$\begin{aligned} \mathcal{L}(\theta) &= \sum_{t=1}^{T} \log f(Y_t|Y_{t-1}, Y_{t-2}, ..., Y_{-p+1}, \theta) \\ &= -(Tn/2)\log(2\pi) + (T/2)\log|\Omega^{-1}| \\ &\quad (-1/2)\sum_{t=1}^{T}\left[(Y_t - \Pi' X_t)' \Omega^{-1} (Y_t - \Pi' X_t)\right] \end{aligned} \tag{18}$$

Maximum Likelihood Estimate (MLE) of Π

The MLE estimate of Π are given by

$$\widehat{\Pi}' = \left[\sum_{t=1}^{T} Y_t X_t' \right] \left[\sum_{t=1}^{T} X_t X_t' \right]^{-1}$$

The jth row of $\widehat{\Pi}'$ is

$$\widehat{\pi}_j' = \left[\sum_{t=1}^{T} Y_{jt} X_t' \right] \left[\sum_{t=1}^{T} X_t X_t' \right]^{-1}$$

which is the estimated coefficient vector from an OLS regression of Y_{jt} on X_t. Thus the MLE estimates for equation j are found by an OLS regression of Y_{jt} on p lags of all the variables in the system. Recall the SUR representation

$$\mathbf{Y} = \mathbf{X}\Pi + \mathbf{u}$$

The MLE estimator is given by

$$\widehat{\Pi} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

MLE of Ω

The MLE, or the value of Ω that maximizes the likelihood among the class of all positive definite matrices is given by

$$\widehat{\Omega} = (1/T)\sum_{t=1}^{T} \widehat{\epsilon}_t \widehat{\epsilon}_t' \tag{19}$$

Number of Lags

As in the univariate case, care must be taken to account for all systematic dynamics in multivariate models. In VAR models, this is usually done by choosing a sufficient number of lags to ensure that the residuals in each of the equations are white noise.

AIC: Akaike information criterion Choosing the p that minimizes the following

$$AIC(p) = \ln|\hat{\Omega}| + \frac{2n^2p}{T}$$

*BIC: Bayesian information criterion*Choosing the p that minimizes the following

$$BIC(p) = \ln|\hat{\Omega}| + \frac{n^2p\ln T}{T}$$

HQ: Hannan- Quinn information criterion Choosing the p that minimizes the following

$$HQ(p) = \ln|\hat{\Omega}| + \frac{2n^2p\ln\ln T}{T}$$

$\hat{p}$ obtained using BIC and HQ are consistent while with AIC it is not.

AIC overestimate the true order with positive probability and underestimate the true order with zero probability.

Suppose a VAR(p) is fitted to $Y_1, ..., Y_T$ (Y_t not necessarily stationary). In small sample the following relations hold:

$$\hat{p}_{BIC} \leq \hat{p}_{AIC} \text{ if } T \geq 8$$

$$\hat{p}_{BIC} \leq \hat{p}_{HQ} \text{ for all } T$$

$$\hat{p}_{HQ} \leq \hat{p}_{AIC} \text{ if } T \geq 16$$

Error Bands for Impulse Response Functions

Method I: Asymptotic Hamilton (1994).

Method II: Bootstrap The idea behind bootstrapping (Runkle, 1987) is to obtain estimates of the small sample distribution for the impulse response functions without assuming that the shocks are Gaussian. Steps:

1. Estimate the VAR and save the $\widehat{\pi}$ and the fitted residuals $\{\widehat{u}_1, \widehat{u}_2, ..., \widehat{u}_T\}$.

2. Draw uniformly from $\{\widehat{u}_1, \widehat{u}_2, ..., \widehat{u}_T\}$ and set $\tilde{u}_1^{(1)}$ equal to the selected realization and use this to construct
$$Y_1^{(1)} = \widehat{A}_1 Y_0 + \widehat{A}_2 Y_{-1} + ... + \widehat{A}_p Y_{-p+1} + \tilde{u}_1^{(1)} \tag{20}$$

3. Taking a second draw (with replacement) $\tilde{u}_2^{(1)}$ generate
$$Y_2^{(1)} = \widehat{A}_1 Y_1^{(1)} + \widehat{A}_2 Y_0 + ... + \widehat{A}_p Y_{-p+2} + \tilde{u}_2^{(1)} \tag{21}$$

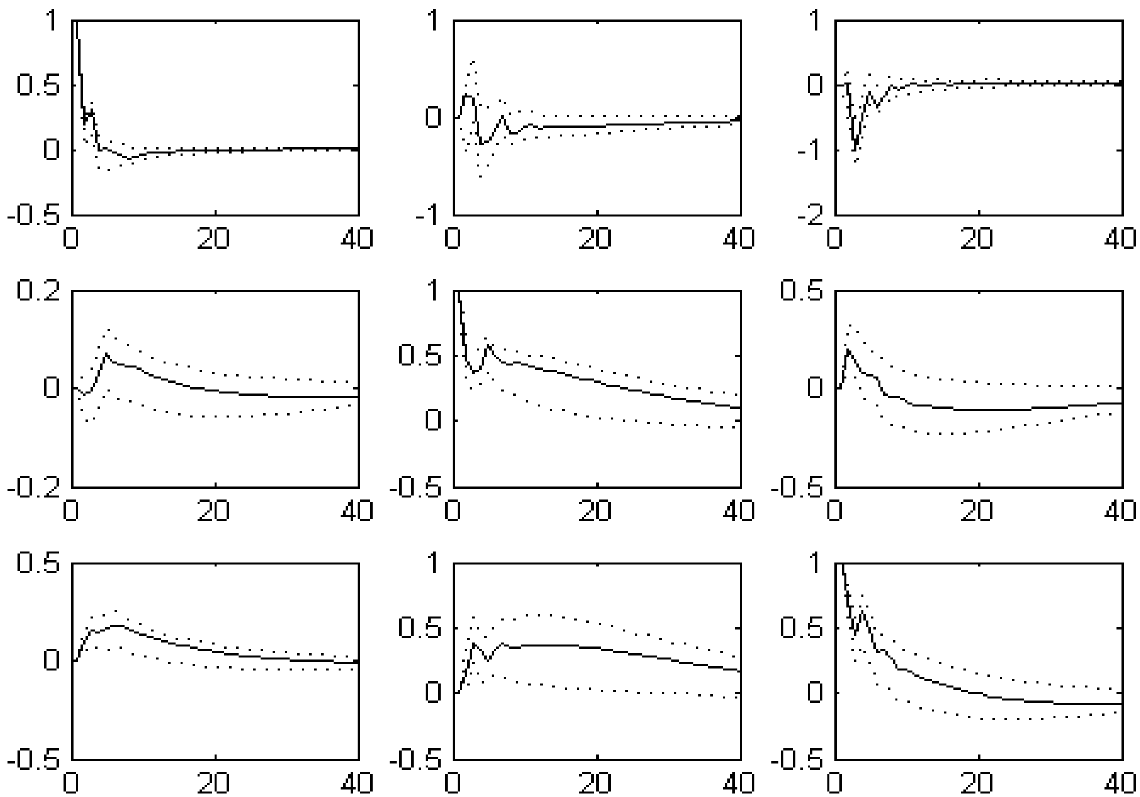
4. Proceeding in this fashion generate a sample of length T $\{Y_1^{(1)}, Y_2^{(1)}, ..., Y_T^{(1)}\}$ and use the sample to compute $\widehat{\pi}^{(1)}$ and the implied impulse response functions $C^{(1)}(L)$.

5. Repeat steps $(3)-(4)$ M times and collect M realizations of $C^{(l)}(L)$, $l = 1, ...M$ and takes for all the elements of the impulse response functions and for all the horizons the αth and $1-\alpha$th percentile to construct confidence bands.

Method III: Montecarlo As bootstrap but drawing the residuals from a normal distribution.

Example: A Monetary VAR with bootstrap bands

We estimate the standard monetary VAR which includes real output growth, the inflation rate and the federal funds rate. These three variables are the core variables for monetary policy analysis in VAR models. Data are taken from the St.Louis Fed FREDII database.



2.3 Granger Causality

Granger Causality

Granger causality If a scalar X cannot help in forecasting Y we say that X does not Granger cause Y. X *fails to Granger cause* Y if for all $s > 0$ the mean squared error of a forecast of $Y_{t+s|t}$ based on $(Y_t, Y_{t-1}, ...,)$ is the same as the MSE of a forecast of Y_{t+s} based on $(Y_t, Y_{t-1}, ...,)$ and $(X_t, X_{t-1}, ...,)$.

Linear projection: multivariate case. Let Y be a $n \times 1$ vector α an $m \times n$ matrix and X an $m \times 1$ vector. The linear projection is defined to be the vector $\alpha' X_t$ satisfying

$$E(Y - \alpha' X)X' = 0'$$

Restricting the attention to linear functions, y fails to Granger-cause x if

$$MSE\left[P(Y_{t+s}|Y_t, Y_{t-1}, ...,)\right] = MSE\left[P(Y_{t+s}|Y_t, Y_{t-1}, ..., X_t, X_{t-1}, ...,)\right] \tag{22}$$

Also we say that Y *is exogenous in the time series sense with respect to* X if (22) holds.

Granger Causality in Bivariate VAR

Let us consider a bivariate VAR

$$\begin{aligned}\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} &= \begin{pmatrix} A_{11}^{(1)} & A_{12}^{(1)} \\ A_{21}^{(1)} & A_{22}^{(1)} \end{pmatrix} \begin{pmatrix} Y_{1t-1} \\ Y_{2t-1} \end{pmatrix} + \begin{pmatrix} A_{11}^{(2)} & A_{12}^{(2)} \\ A_{21}^{(2)} & A_{22}^{(2)} \end{pmatrix} \begin{pmatrix} Y_{1t-2} \\ Y_{2t-2} \end{pmatrix} + \\ &\quad + ... + \begin{pmatrix} A_{11}^{(p)} & A_{12}^{(p)} \\ A_{21}^{(p)} & A_{22}^{(p)} \end{pmatrix} \begin{pmatrix} Y_{1t-p} \\ Y_{2t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \end{aligned} \tag{23}$$

We say that Y_2 *fails to Granger cause* Y_1 if the elements $A_{12}^{(j)} = 0$ for $j = 1, ..., p$. We can check that if $A_{12}^{(j)} = 0$ the two MSE coincide. For $s = 1$ we have

$$\begin{aligned} P(Y_{1t+1}|Y_{1t}, Y_{1t-1}, ..., Y_{2t}, Y_{2t-1}) &= A_{11}^{(1)}Y_{1t} + A_{12}^{(1)}Y_{2t} + A_{11}^{(2)}Y_{1t-1} + A_{12}^{(2)}Y_{2t-1} + \\ &\quad + ... + A_{11}^{(p)}Y_{1t-p+1} + A_{12}^{(p)}Y_{2t-p+1} \end{aligned}$$

clearly if if $A_{12}^{(j)} = 0$

$$\begin{aligned} P(Y_{1t+1}|Y_{1t}, Y_{1t-1}, ..., Y_{2t}, Y_{2t-1}) &= A_{11}^{(1)}Y_{1t} + A_{11}^{(2)}Y_{1t-1} + ... + A_{11}^{(p)}Y_{1t-p+1} \\ &= \hat{P}(Y_{1t+1}|Y_{1t}, Y_{1t-1}, ...) \end{aligned} \tag{24}$$

Implication: if Y_2 *fails to Granger cause* Y_1 then the Wold representation of Y_t is

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} C_{11}(L) & 0 \\ C_{21}(L) & C_{22}(L) \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \tag{25}$$

that is the second Wold shock has no effects on the first variable. This it is easy to show by deriving the Wold representation by invertinG the VAR polynomial matrix.

Econometric test for Granger Causality

The simplest approach to test Granger causality in an autoregressive framework is to estimate the bivariate VAR with p lags by OLS and test the null hypothesis $H_0: \ A_{12}^{(1)} = A_{12}^{(2)} = ...A_{12}^{(p)} = 0$ using

$$S_1 = \frac{(RSS_0 - RSS_1)/p}{RSS_1/(T - 2p - 1)}$$

where RSS_0 are the OLS residuals of the restricted model and RSS_1 are those of the restricted model and reject if $S_1 > F_{(\alpha,p,T-2p-1)}$.

Application 1: Money, Income and Causality, Sims (1972)

"It has long been known that money stock and current dollar measures of economic activity are positively correlated.[...] A body of macro-economic theory, the "Quantity Theory," explains these empirical observations as reflecting a causal relation running from money to income. However, it is widely recognized that no degree of positive association between money and income can by itself prove that variation in money causes variation in income. Money might equally well react passively and very reliably to fluctuations in income." Sims (1972) "Money, Income, and Causality" AER.

Application 1: Money, Income and Causality, Sims (1972)

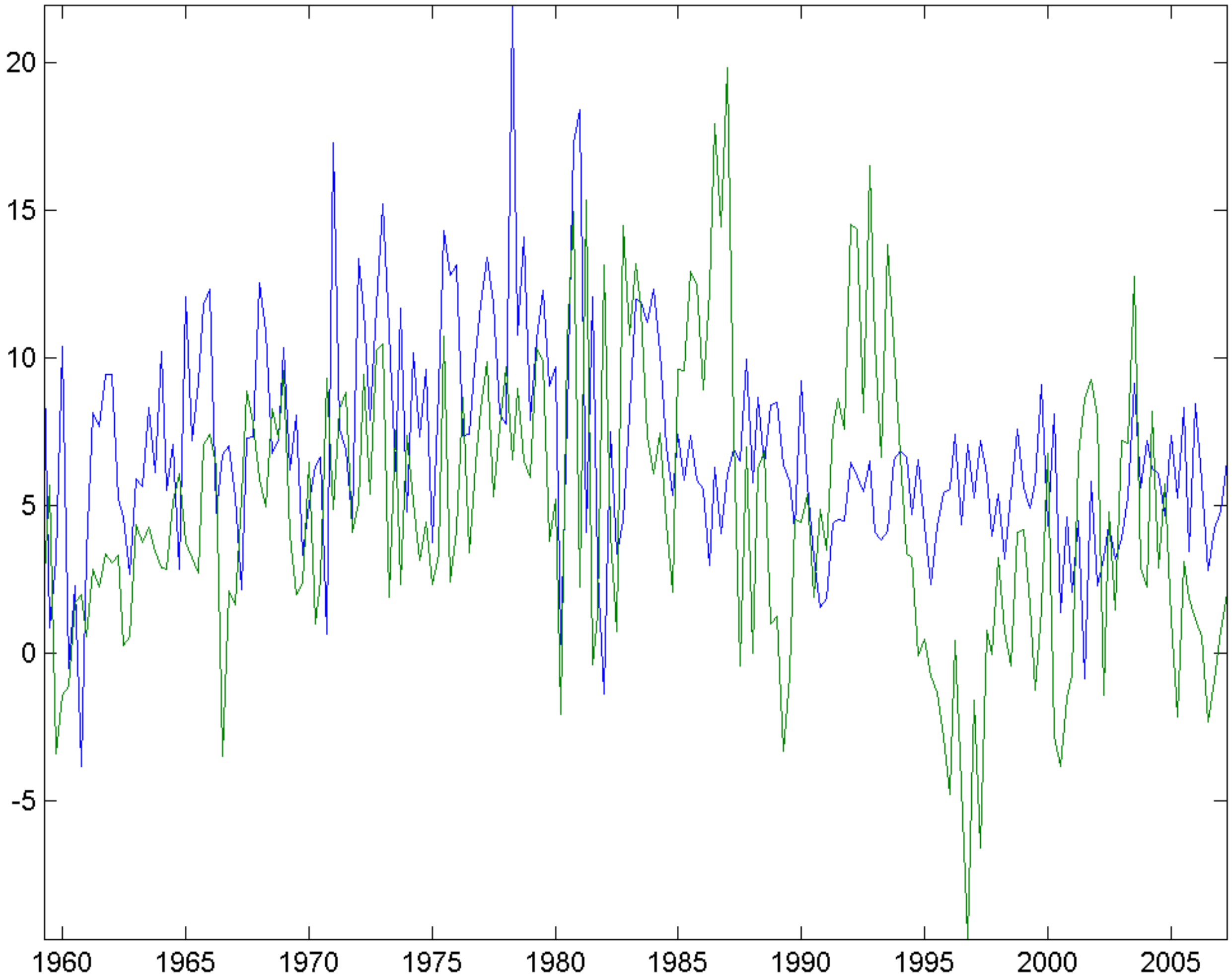


Figure 5: Blu: nominal gnp growth rates; green: M1 growth rates.

Application 1: Money, Income and Causality, Sims (1972)

Table 2: F-Tests of Granger Causality

	1959:II-1972:I	1959:II-2007:III
$M \to Y$	4.4440	2.2699
$Y \to M$	0.5695	3.5776
10%	2.0948	1.7071
5%	2.6123	1.9939
1%	3.8425	2.6187

In the first sample money Granger cause (at 5%) output but not the converse (Sims(72)'s result). In the second sample at the 5% both output Granger cause money and money Granger cause output.

Application 2: Output Growth and the Yield Curve

Many research papers have found that interest rate spread (difference in long and short yield) has been a good predictor, i.e. a variable that helps to forecast, for the real GDP growth in the US (Estrella, 2000,2005). Recessions are preceeded by sharp fall in the spread (short increases compared to long rates).

$\Rightarrow$ The spread should Granger cause output growth.

However recent evidence suggests that its predictive power has reduced since the beginning of the 80s (see D'Agostino, Giannone and Surico, 2006).

$\Rightarrow$ The spread should no longer cause output growth after mid 80's .

We estimate a bivariate VAR for the growth rates of the real GDP and the difference between the 10-year rate and the federal funds rate. Data are from FREDII StLouis Fed spanning from 1954:III-2007:III. The AIC criterion suggests $p = 6$.

Application 2: Output Growth and the Yield Curve

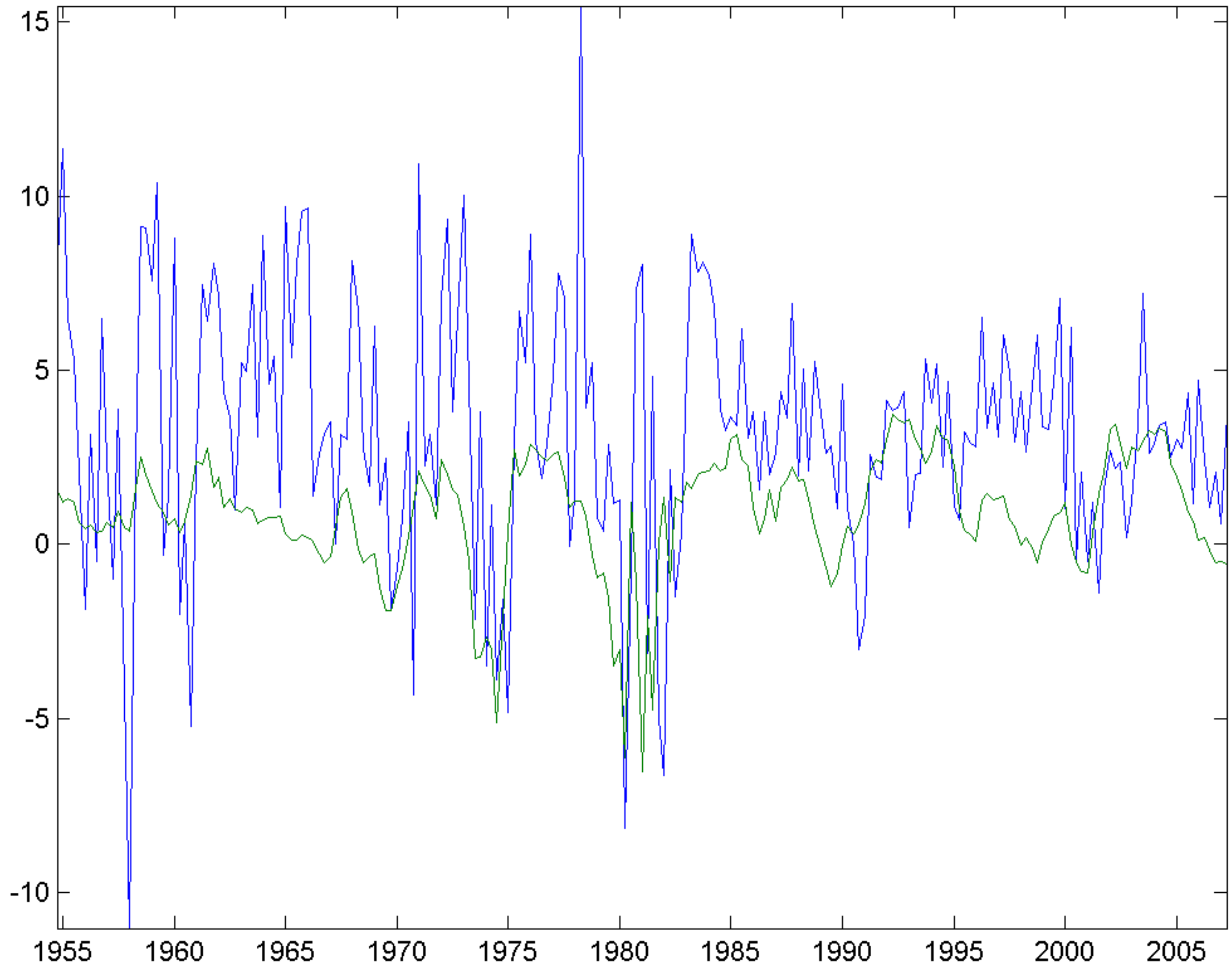


Figure 6: Blu: real gdp growth rates; green: spread long-short.

Application 2: Output Growth and the Yield Curve

Table 1: F-Tests of Granger Causality

	1954:IV-2007:III	1954:IV-1990:I	1990:I-2007:III
S_1	5.4233	6.0047	0.9687
10%	1.8050	1.8222	1.8954
5%	2.1460	2.1725	2.2864
1%	2.8971	2.9508	3.1864

We cannot reject the hypothesis that the spread does not Granger cause real output growth in the last period, while we reject the hypothesis for all the other sample. This can be explained by a change in the information content of private agents expectations, which is the information embedded in the yield curve.

Caveat: Granger Causality Tests and Forward Looking Behavior

Let us consider the following simple model of stock price determination where P_t is the price of one share of a stock, D_{t+1} are dividends payed at $t+1$ and r is the rate of return of the stock

$$(1+r)P_t = E_t(D_{t+1} + P_{t+1})$$

According to the theory stock price incorporates the market's best forecast of the present value of the future dividends. Solving forward we have

$$P_t = E_t \sum_{j=1}^{\infty} \left[\frac{1}{1+r}\right]^j D_{t+j}$$

Suppose

$$D_t = d + u_t + \delta u_{t-1} + v_t$$

where u_t, v_t are Gaussian WN and d is the mean dividend. The forecast of D_{t+j} based on this information is

$$E_t(D_{t+j}) = \begin{cases} d + \delta u_t & \text{for} \quad j = 1 \\ d & \text{for} \quad j = 2, 3, \ldots \end{cases}$$

Substituting in the stock price equation we have

$$P_t = d/r + \delta u_t/(1+r)$$

Thus the price is white noise and could not be forecast on the basis of lagged stock prices or dividends. No series should Granger cause prices. The value of u_{t-1} can be uncovered from the lagged stock price

$$\delta u_{t-1} = (1+r)P_{t-1} - (1+r)d/r$$

The bivariate VAR takes the form

$$\begin{pmatrix} P_t \\ D_t \end{pmatrix} = \begin{pmatrix} d/r \\ -d/r \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ (1+r) & 0 \end{pmatrix} \begin{pmatrix} P_{t-1} \\ D_{t-1} \end{pmatrix} + \begin{pmatrix} \delta u_t/(1+r) \\ u_t + v_t \end{pmatrix} \tag{26}$$

Granger causation runs in the opposite direction from the true causation. Dividends fail to Granger cause prices even though expected dividends are the only determinant of prices. On the other hand prices Granger cause dividends even though this is not the case in the true model.

2.4 Forecasting

Forecasting with VAR models

Let us consider the VAR(p) in companion form

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{e}_t \tag{27}$$

where $\mathbf{e}_t$ is white noise. The h-step ahead linear predictor of $\mathbf{Y}_{t+h}$ conditional on the information available at time t, $\mathbf{Y}_{t+h|t}$, is given by

$$\mathbf{Y}_{t+h|t} = \mathbf{A}^h\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t+h-1|t} \tag{28}$$

where the first n rows of $\mathbf{Y}_{t+h|t}$ represent the optimal forecast of Y_{t+h}. From (28) it is easy to compute recursively the forecast for Y_{t+h} at any horizon.

The predictor in the previous slide is optimal in the sense that, as we know, it delivers the minimum MSE among those that are linear functions of Y.

Using

$$\mathbf{Y}_{t+h} = \mathbf{A}^h \mathbf{Y}_t + \sum_{i=0}^{h-1} \mathbf{A}^i \mathbf{e}_{t+h-i} \tag{29}$$

we get the forecast error

$$\mathbf{Y}_{t+h} - \mathbf{Y}_{t+h|t} = \sum_{i=0}^{h-1} \mathbf{A}^i \mathbf{e}_{t+h-i} \tag{30}$$

From the forecast error it is easy to obtain the Mean Square Error, the covariance of the forecast error,

$$\begin{aligned} MSE[\mathbf{Y}_{t+h|t}] = E\left(\mathbf{Y}_{t+k} - \mathbf{Y}_{t+h|t}\right)(\mathbf{Y}_{t+k} - \mathbf{Y}_{t+h|t})' = \mathbf{\Sigma}(h) &= \sum_{i=0}^{h-1} \mathbf{A}^i \Omega \mathbf{A}^{i\prime} \\ &= \mathbf{\Sigma}(h-1) + \mathbf{A}^{h-1} \Omega \mathbf{A}^{h-1\prime} \end{aligned}$$

the MSE for will be the first upper left $n \times n$ matrix. Notice that the MSE is non decreasing and that as $h \to \infty$ will approach the variance of Y_t.

Application: Forecasting inflation and unemployment

D'Agostino Gambetti and Giannone (JAE 2012) consider a trivariate VAR(2) model including inflation unemployment and the short term interest rate.

The sample spans from 1948:I-2007:IV. Forecast up to 12 quarters ahead.

Estimation is in real time and is made using both VAR and AR for each of the series. Estimation is done both recursively snd with rolling window.

Table 1: Forecasting Accuracy over the sample 1970-2007: mean square forecast errors.

Horizon (quarters)	Variable	RW (MSFE)	AR-REC (RMSFE)	AR-ROL (RMSFE)	SV-AR (RMSFE)	TV-AR (RMSFE)	VAR-REC (RMSFE)	VAR-ROL (RMSFE)	SV-VAR (RMSFE)	TV-VAR (RMSFE)
	π	2.15	1.13	1.08	1.05	1.03	1.15	1.01	1.14	0.85
1	UR	0.15	1.00	1.08	0.98	1.00	0.99	1.18	0.96	1.04
	IR	0.87	1.12	1.23	1.01	1.04	0.99	1.09	0.94	0.99
	Avg.		1.08	1.13	1.01	1.02	1.04	1.09	1.01	0.96
	π	2.24	1.17	1.03	0.82	0.88	1.37	1.22	1.01	0.62
4	UR	1.07	1.03	1.24	0.95	1.01	0.67	0.91	0.67	0.77
	IR	3.46	1.05	1.20	0.93	0.95	0.96	1.39	0.93	0.93
	Avg.		1.08	1.16	0.90	0.95	1.00	1.17	0.87	0.77
	π	3.06	1.19	1.13	0.89	0.93	1.6	1.38	1.11	0.65
8	UR	2.39	0.95	1.14	0.88	0.95	0.45	0.63	0.42	0.61
	IR	7.54	1.05	1.18	1.11	0.92	0.99	1.44	0.85	0.89
	Avg.		1.06	1.15	0.88	0.93	1.01	1.15	0.80	0.72
	π	3.31	1.28	1.24	0.95	1.00	1.93	1.60	1.26	0.69
12	UR	3.22	0.85	1.12	0.79	0.86	0.47	0.85	0.40	0.51
	IR	10.28	1.08	1.15	0.82	0.91	1.03	1.32	0.79	0.86
	Avg.		1.07	1.17	0.85	0.92	1.14	1.26	0.82	0.72

The table reports the results relative to the forecasting accuracy using point forecasts. The variable we forecast are inflation (π_t), the unemployment rate (UR_t) and the interest rate (IR_t). The forecasting models are: RW - random walk; AR-REC - AR estimated recursively; AR-ROL - AR estimated with a rolling window; TV-VAR - time-varying VAR; VAR-REC - VAR estimated recursively; VAR-ROL - VAR estimated with a rolling window. For the random walk model we report the mean square forecast error (MSFE). For the other models we report the relative mean square forecast error (RMSFE), i.e. the ratio of the MSFE of a particular model to the MSFE of the naïve model. For each horizon it is also reported the average of the RMSFE across variables (Avg.).

Table 2: Forecasting Accuracy over the sample 1985-2007: Mean square forecast errors.

Horizon (quarters)	Variable	RW (MSFE)	AR-REC (RMSFE)	AR-ROL (RMSFE)	SV-AR (RMSFE)	TV-AR (RMSFE)	VAR-REC (RMSFE)	VAR-ROL (RMSFE)	SV-VAR (RMSFE)	TV-VAR (RMSFE)
	π	0.93	2.61	1.19	1.23	1.21	1.29	1.35	1.28	0.98
1	UR	0.05	2.80	1.16	1.05	1.07	1.09	1.17	0.99	1.03
	IR	0.27	3.64	1.08	0.85	0.83	0.87	1.02	0.77	0.83
	Avg.		3.02	1.14	1.05	1.04	1.08	1.18	1.01	0.94
	π	0.45	5.76	1.54	1.19	1.16	2.22	2.64	1.42	0.94
4	UR	0.37	3.00	1.15	0.82	0.82	0.97	1.23	0.77	0.89
	IR	2.09	1.74	1.17	0.78	0.81	0.78	1.20	0.74	0.81
	Avg.		3.50	1.29	0.93	0.93	1.32	1.69	0.97	0.87
	π	0.57	6.39	2.09	1.10	1.08	3.03	3.11	1.55	0.72
8	UR	1.33	1.72	0.86	0.61	0.56	0.42	0.72	0.38	0.57
	IR	5.16	1.53	1.05	0.68	0.74	0.67	1.20	0.63	0.75
	Avg.		3.21	1.33	0.80	0.79	1.37	1.68	0.85	0.69
	π	0.92	4.61	2.10	0.91	0.86	3.47	2.51	1.34	0.46
12	UR	2.25	1.22	0.72	0.48	0.43	0.35	0.73	0.27	0.48
	IR	7.69	1.44	0.89	0.55	0.63	0.70	1.13	0.51	0.61
	Avg.		2.42	1.24	0.65	0.64	1.51	1.46	0.71	0.51

The table reports the results relative to the forecasting accuracy using point forecasts. The variable we forecast are inflation (π_t), the unemployment rate (UR_t) and the interest rate (IR_t). The forecasting models are: RW - random walk; AR-REC - AR estimated recursively; AR-ROL - AR estimated with a rolling window; TV-VAR - time-varying VAR; VAR-REC - VAR estimated recursively; VAR-ROL - VAR estimated with a rolling window. For the random walk model we report the mean square forecast error (MSFE). For the other models we report the relative mean square forecast error (RMSFE), i.e. the ratio of the MSFE of a particular model to the MSFE of the naïve model. For each horizon it is also reported the average of the RMSFE across variables (Avg.).

3. Structural VARs - Theory

Structural Vector Autoregressions

Impulse response functions are interpreted under the assumption that *all the other shocks are held constant*. However in the Wold representation the shocks are not orthogonal. So the assumption is not very realistic!.

This is why we need Structural VAR in order to perform policy analysis. Ideally we would like to have

1) orthogonal shock

2) shocks with economic meaning (technology, demand, labor supply, monetary policy etc.)

Statistical Orthogonalizations

There are two easy way to orthogonalize shocks.

1) Cholesky decomposition

2) Spectral Decomposition

Cholesky decomposition Let us consider the matrix Ω. The Cholesky factor, S, of Ω is defined as the unique lower triangular matrix such that $SS' = \Omega$. This implies that we can rewrite the VAR in terms of orthogonal shocks $\eta = S^{-1}\epsilon_t$ with identity covariance matrix

$$A(L)Y_t = S\eta_t$$

Impulse response to orthogonalized shocks are found from the MA representation

$$\begin{aligned} Y_t &= C(L)S\eta_t \\ &= \sum_{j=0}^{\infty} C_j S \eta_{t-j} \end{aligned} \tag{31}$$

where $C_j S$ has the interpretation

$$\frac{\partial Y_{t+j}}{\partial \eta_t} = C_j S \tag{32}$$

That is, the row i, column k element of $C_j S$ identifies the consequences of a unit increase in η_k at date t for the value of the ith variable at time $t+j$ holding all other η_{-k} constant.

Spectral Decomposition Let V and be a matrix containing the eigenvectors of Ω and Λ a diagonal matrix with the eigenvalues of Ω on the main diagonal. Then we have that $V\Lambda V' = \Omega$. This implies that we can rewrite the VAR in terms of orthogonal shocks $\xi = (V\Lambda^{1/2})^{-1}\epsilon_t$ with identity covariance matrix

$$A(L)Y_t = V\Lambda^{1/2}\xi$$

Impulse response to orthogonalized shocks are found from the MA representation

$$\begin{aligned} Y_t &= C(L)V\Lambda^{1/2}\xi_t \\ &= \sum_{j=0}^{\infty} C_j V\Lambda^{1/2}\xi_{t-j} \end{aligned} \tag{33}$$

where $C_j V\Lambda^{1/2}$ has the interpretation

$$\frac{\partial Y_{t+j}}{\partial \xi_t} = C_j V\Lambda^{1/2} \tag{34}$$

That is, the row i, column k element of $C_j V\Lambda^{1/2}$ identifies the consequences of a unit increase in ξ_k at date t for the value of the ith variable at time $t+j$ holding all other η_{-k} constant.

The Class of Orthonormal Representations From the class of invertible MA representation of Y_t we can derive the class of orthonormal representation, i.e. the class of representations of Y_t in term of orthonormal shocks. Let H any orthogonal matrix, i.e. $HH' = I$. Defining $w_t = (SH)^{-1}\epsilon_t$ we can recover the general class of the orthonormal representation of Y_t

$$\begin{aligned} Y_t &= C(L)SHw_t \\ &= F(L)w_t \end{aligned}$$

where $F(L) = C(L)SH$ and $w_t \sim WN$ with

$$\begin{aligned} E(w_t w_t') &= E(HS^{-1}\epsilon_t\epsilon_t' S^{-1\prime}H') \\ &= HS^{-1}E(\epsilon_t\epsilon_t')S^{-1\prime}H' \\ &= HS^{-1}\Omega S^{-1\prime}H' \\ &= HS^{-1}SS'(S')^{-1}H' \\ &= I \end{aligned}$$

Problem H can be any, so how should we choose one?

The Identification Problem

Problem: what is the economic interpretation of the orthogonal shocks? What is the economic information contained in the impulse response functions to orthogonal shocks?

Except for special cases not clear.

The idea is that structural economic shocks are linear combinations of the VAR innovations.

Identifying the VAR means fixing a particular matrix H, i.e. choosing one particular representation of Y_t in order to recover the structural shocks from the VAR innovations

In order to choose a matrix H we have to fix $n(n-1)/2$ parameters since there is a total of n^2 parameters and a total of $n(n+1)/2$ restrictions implied by orthonormality.

The idea is to use economic theory in order to derive some restrictions on the effects of some shock on a particular variables to fix the remaining $n(n-1)/2$.

Zero restrictions: contemporaneous restrictions An identification scheme based on zero contemporaneous restrictions is a scheme which imposes restrictions to zero on the matrix F_0, the matrix of the impact effects.

Example. Let us consider a bivariate VAR. We have a total of $n^2 = 4$ parameters to fix. $n(n+1)/2 = 3$ are pinned down by the ortnonormality restrictions so that there are $n(n-1)/2 = 1$ free parameters. Suppose that the theory tells us that shock 2 has no effect on impact (contemporaneously) on Y_1 equal to 0, that is $F_0^{12} = 0$. This is the additional restriction that allows us to identify the shocks. In particular we will have the following restrictions:

$$HH' = I$$
$$S_{11}H_{12} + S_{12}H_{22} = 0$$

Since $S_{12} = 0$ the solution is $H_{11} = H_{22} = 1$ and $H_{12} = H_{21} = 0$.

A common identification scheme is the Cholesky scheme (like in this case). This implies setting $H = I$. Such an identification scheme creates a recursive contemporaneous ordering ordering among variables since S^{-1} is triangular.

This means that any variable in the vector Y_t does not depend contemporanously on the variables ordered after.

Results depend on the particular ordering of the variables.

Zero restrictions: long run restrictions An identification scheme based on zero long run restrictions is a scheme which imposes restrictions on the matrix $F(1) = F_0 + F_1 + F_2 + ...$, the matrix of the long run coefficients.

Example. Again let us consider a bivariate VAR. We have a total of $n^2 = 4$ parameters to fix. $n(n+1)/2 = 3$ are pinned down by the ortnonormality restrictions so that there are $n(n-1)/2 = 1$ free parameters. Suppose that the theory tells us that shock 2 does not affect Y_1 in the long run, i.e. $F_{12}(1) = 0$. This is the additional restriction that allows us to identify the shocks. In particular we will have the following restrictions:

$$HH' = I$$
$$D_{11}(1)H_{12} + D_{12}(1)H_{22} = 0$$

where $D(1) = C(1)S$ represents the long run effects of the Cholesky shocks.

Signs restrictions The previous two examples yield just identification in the sense that the shockis uniquely identified, there exists a unique matrix H yielding the structural shocks.

Sign identification is based on qualitative restriction involving the sign of some shocks on some variables. In this case we will have sets of consistent impulse response functions.

Example. Again let us consider a bivariate VAR. We have a total of $n^2 = 4$ parameters to fix. $n(n+1)/2 = 3$ are pinned down by the ortnonormality restrictions so that there are $n(n-1)/2 = 1$ free parameters. Suppose that the theory tells us that shock 2, which is the interesting one, produce a positive effect on Y_1 for k periods after the shock $F_j^{12} > 0$ for $j = 1, ..., k$. We will have the following restrictions:

$$
\begin{aligned}
HH' &= I \\
S_{11}H_{12} + S_{12}H_{22} &> 0 \\
D_{j,12}H_{12} + D_{j,22}H_{22} &> 0 \quad j = 1, ..., k
\end{aligned}
$$

where $D_j = C_j S$ represents the effects at horizon j.

In a classical statistics approach this delivers not exact identification since there can be many H consistent with such a restriction. That is for each parameter of the impulse response functions we will have an admissible set of values.

Increasing the number of restrictions can be helpful in reducing the number of H consistent with such restrictions.

Partial Identification In many cases we might be interested in identifying just a single shock and not all the n shocks.

Since the shock are orthogonal we can also partially identify the model, i.e. fix just one (or a subset of) column of H. In this case what we have to do is to fix $n-1$ elements of H, all but one elements of a column of the identifying matrix. The additional restriction is provided by the norm of the vector equal one.

Example Suppose $n = 3$. We want to identify a single shock using the restriction that such shock has no effects on the first variable on impact a positive effect on the second variable and negative on the third variable.

Impulse Response Functions

Impulse response to identified shocks are found from the structural MA representation

$$\begin{aligned} Y_t &= C(L)SHw_t \\ &= \sum_{j=0}^{\infty} C_j SHw_{t-j} \end{aligned} \tag{35}$$

where C_jSH has the interpretation

$$\frac{\partial Y_{t+j}}{\partial w_t} = C_jSH \tag{36}$$

That is, the row i, column k element of C_jSH identifies the consequences of a unit increase in w_k at date t for the value of the ith variable at time $t+j$.

Confidence bands can be obtained using the bootstrapping procedure described in lecture 3. Now the additional step is that for any draw of the reduced form impulse response functions we have to implement the identification scheme adopted.

Variance Decomposition

The second type of analysis which is usually done in SVAR is the variance decomposition analysis.

The idea is to decompose the total variance of a time series into the percentages attributable to each structural shock.

Variance decomposition analysis is useful in order to address questions like "What are the sources of the business cycle?" or "Is the shock important for economic fluctuations?".

Let us consider the MA representation of an identified SVAR

$$Y_t \quad = \quad F(L)w_t$$

The variance of Y_{it} is given by

$$\begin{aligned}\text{var}(Y_{it}) &= \sum_{k=1}^{n}\sum_{j=0}^{\infty} F_{ik}^{j2}\text{var}(w_{kt}) \\ &= \sum_{k=1}^{n}\sum_{j=0}^{\infty} F_{ik}^{j2}\end{aligned}$$

where $\sum_{j=0}^{\infty} F_{ik}^{j2}$ is the variance of Y_{it} generated by the kth shock. This implies that

$$\frac{\sum_{j=0}^{\infty} F_{ik}^{j2}}{\sum_{k=1}^{n}\sum_{j=0}^{\infty} F_{ik}^{j2}}$$

is the percentage of variance of Y_{it} explained by the kth shock.

It is also possible to study the of the series explained by the shock at different horizons, i.e. short vs. long run. Consider the forecast error in terms of structural shocks. The horizon h forecast error is given by

$$Y_{t+h} - Y_{t+h|t} = F_0 w_{t+1} + F_2 w_{t+2} + ... + F_k w_{t+h}$$

the variance of the forecast error is thus

$$\begin{aligned} \text{var}(Y_{t+h} - Y_{t+h|t}) &= \sum_{k=1}^{n} \sum_{j=0}^{h} F_{ik}^{j2} \text{var}(w_{kt}) \\ &= \sum_{k=1}^{n} \sum_{j=0}^{h} F_{ik}^{j2} \end{aligned}$$

Thus the percentage of variance of Y_{it} explained by the kth shock is

$$\frac{\sum_{j=0}^{h} F_{ik}^{j2}}{\sum_{k=1}^{n} \sum_{j=0}^{h} F_{ik}^{j2}}$$

4: Structural VARs - Applications

Monetary Policy Shocks (Christiano Eichenbaum and Evans, 1998)

Monetary policy shocks is the unexpected part of the equation for the monetary policy instrument (S_t).

$$S_t = f(\mathcal{I}_t) + w_t^{mp}$$

$f(\mathcal{I}_t)$ represents the systematic response of the monetary policy to economic conditions, $\mathcal{I}_t$ is the information set at time t and w_t^{mp} is the monetary policy shock.

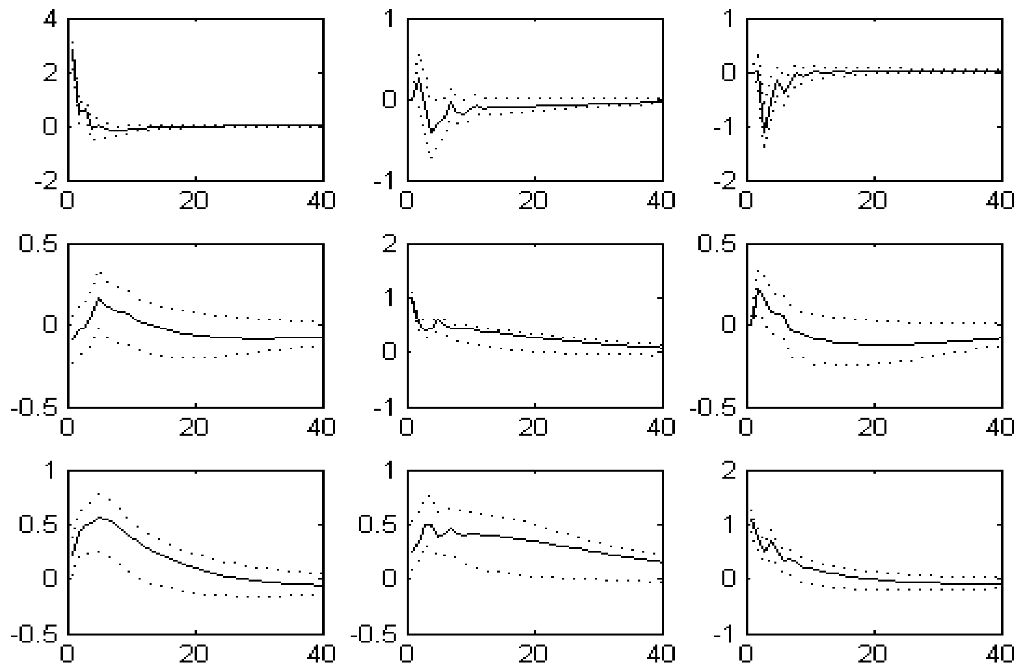
The "standard" way to identify monetary policy shock is through zero contemporaneous restrictions. Using the standard trivariate monetary VAR (a simplified version of the CEE 98 VAR) including output growth, inflation and the federal funds rate we identify the monetary policy shock using the following restrictions:

1) Monetary policy shocks do not affect output within the same quarter

2) Monetary policy shocks do not affect inflation within the same quarter

These two restrictions are not sufficient to identify all the shocks but are sufficient to identify the monetary policy shock.

A simple way to implement the restrictions is to take simply the Cholesky decomposition of the variance covariance matrix in a system in which the federal funds rate is ordered last. The last column of the impulse response functions is the column of the monetary policy shock.



Cholesky impulse response functions of a system with GDP inflation and the federal funds rate. Monetary shock is in the third column.

Notice that after a monetary tightening inflation goes up which is completely counterintuitive according to the standard transmission mechanism. This phenomenon if known as the *price puzzle*. Why is this the case?.

”*Sims (1992) conjectured that prices appeared to rise after certain measures of a contractionary policy shock because those measures were based on specifications of $\mathcal{I}_t$ that did not include information about future inflation that was available to the Fed. Put differently, the conjecture is that policy shocks which are associated with substantial price puzzles are actually confounded with non-policy disturbances that signal future increases in prices.*” (CEE 98)

Sims shows that including commodity prices (signaling future inflation increases) may solve the puzzle.

Uhlig (2005) JME's monetary policy shocks

Uhlig (2005 JME) proposes a very different method to identify monetary policy shocks. Instead of using zero restrictions as in CEE he uses sign restrictions.

He identifies the effects of a monetary policy shocks using restrictions which are implied by several economic models.

In particular a contractionary monetary policy shock:

1. does not increase prices for k periods after the shock

2. does not increase money or monetary aggregates (i.e. reserves) for k periods after the shock

3. does not reduce short term interest rate for k periods after the shock.

Since just one shock is identified only a column of H has to be identified, say column one.

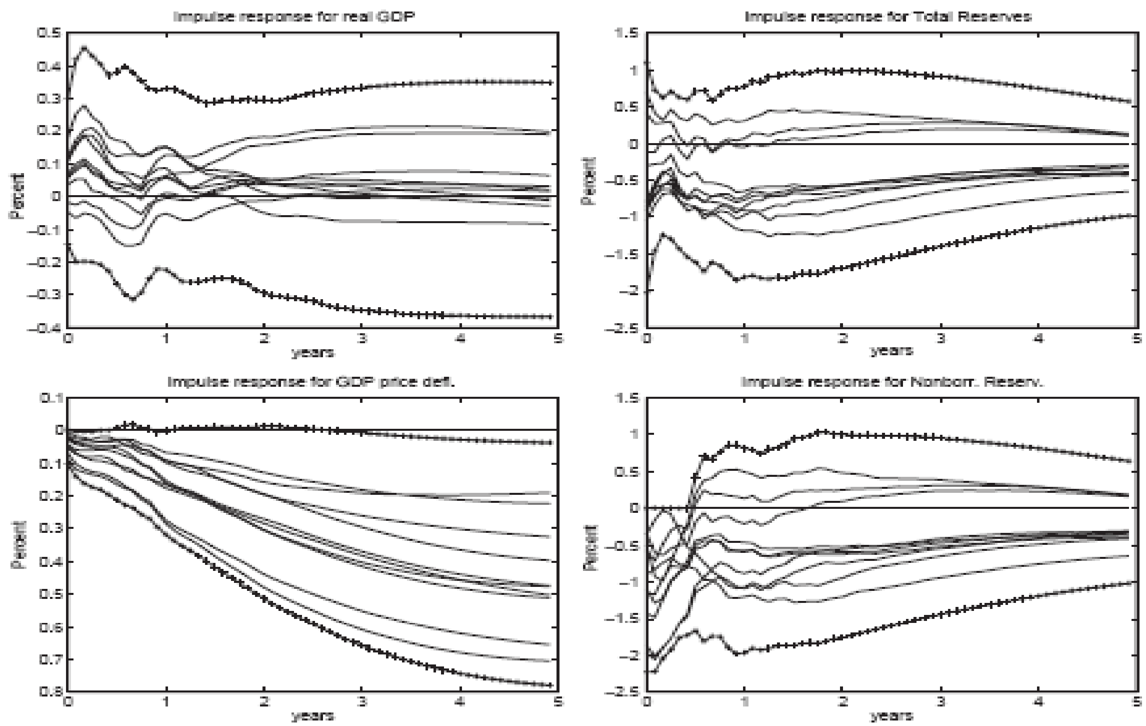
If we order the variables in vector Y_t as follows: GDP inflation, money growth and the interest rate the restrictions imply $F_k^{i1} < 0$ for $i = 2, 3$ and $F_k^{41} > 0$.

In order to draw impulse response functions he applies the following algorithm:

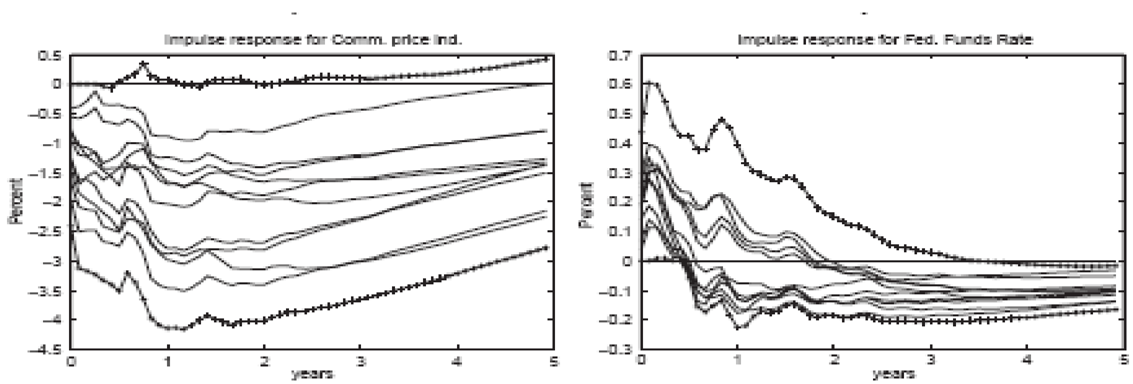
1. He assumes that the column of H, H_1, represents the coordinate of a point uniformly distributed over the unit hypersphere (in case of bivariate VAR it represents a point in a circle). To draw such point he draws from a $N(0, I)$ and divide by the norm of the vector.

2. Compute the impulse response functions $C_j S H_1$ for j=1,..,k.

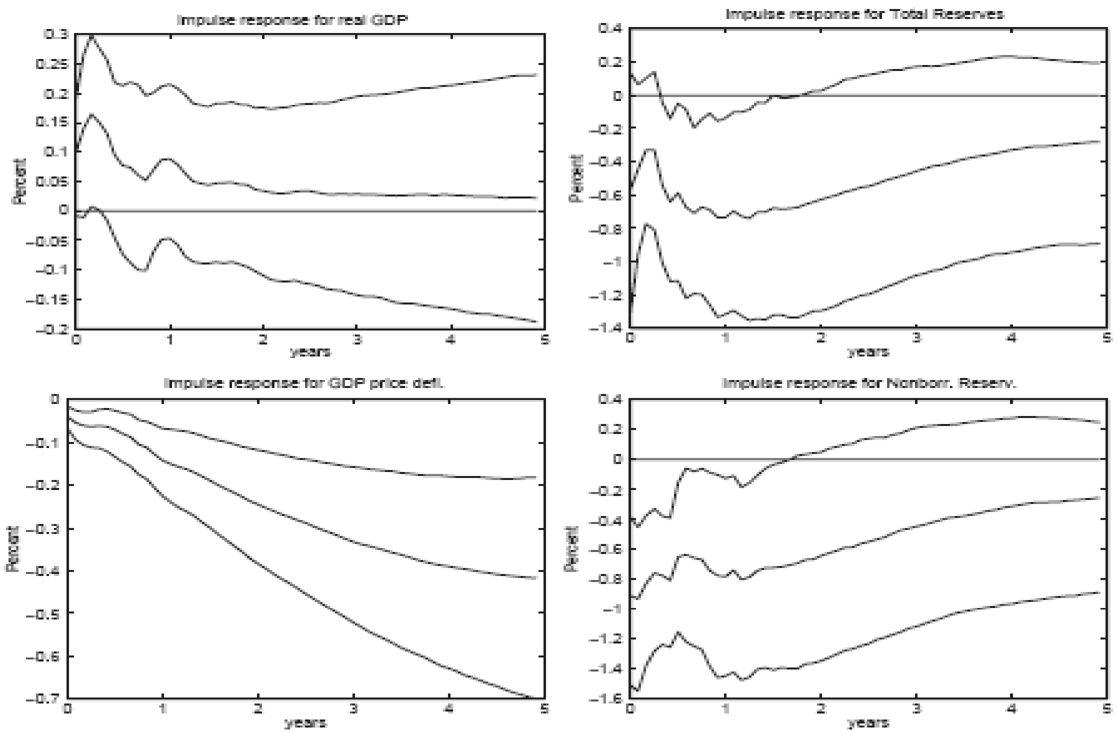
3. If the draw satisfies the restrictions keep it and go to 1), otherwise discard it and go to 1). Repeat 1)-3) a big number of timef L.



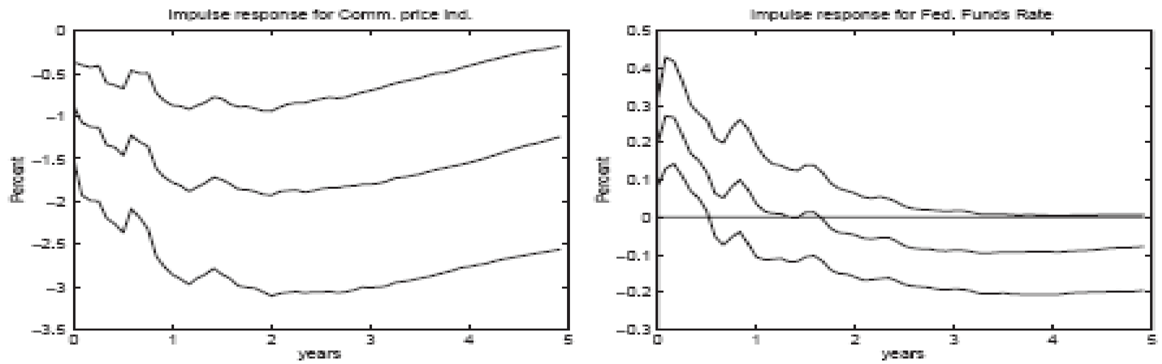
Source: What are the effects of a monetary policy shock... JME H. Uhlig (2006)



Source: What are the effects of a monetary policy shock... JME H. Uhlig (2006)



Source: What are the effects of a monetary policy shock... JME H. Uhlig (2006)



Source: What are the effects of a monetary policy shock... JME H. Uhlig (2006)

Monetary policy and housing

Central question: how does monetary policy affects house prices?

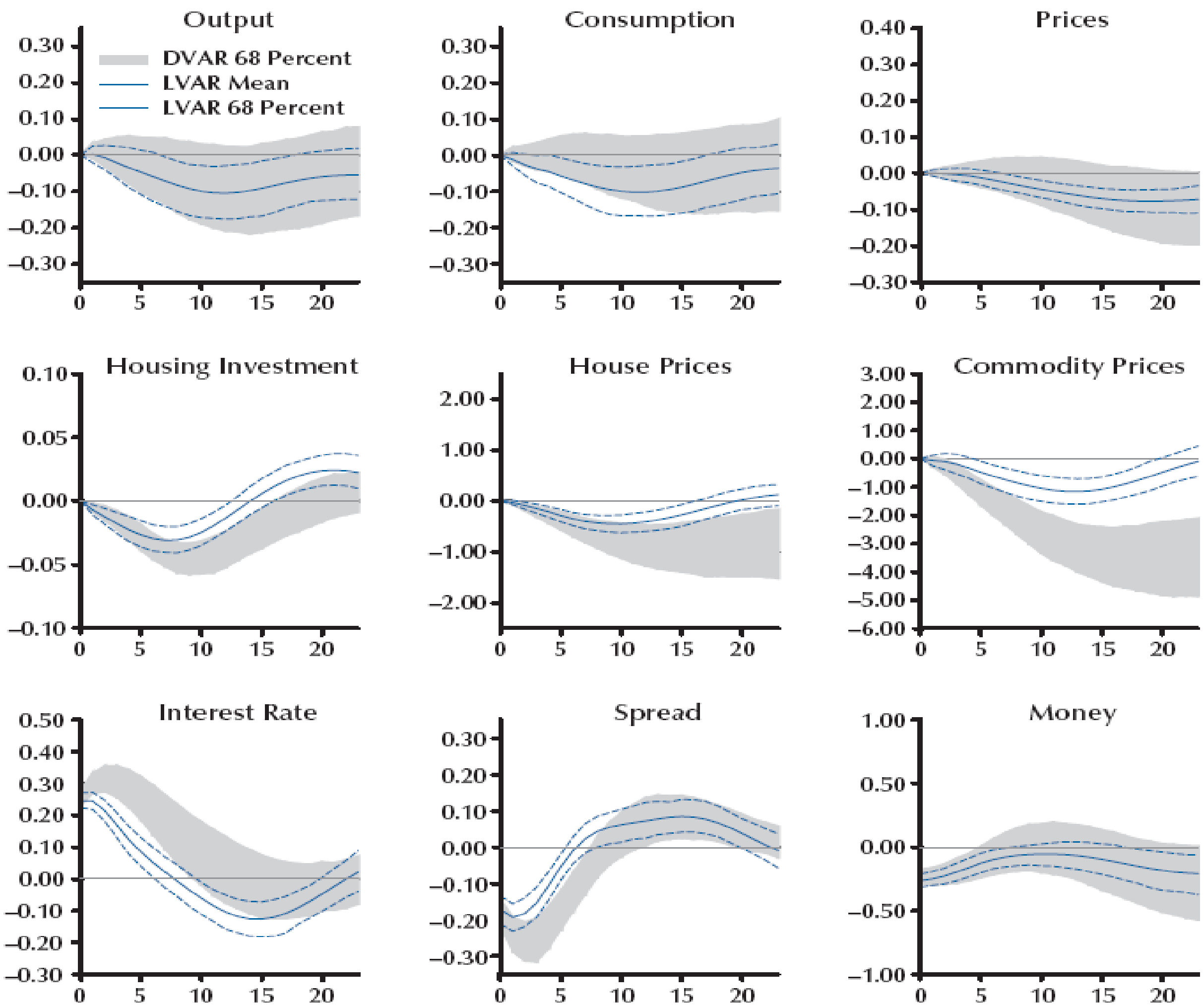
Jarocinski and Smets (2008) addresses this question.

Strategy:

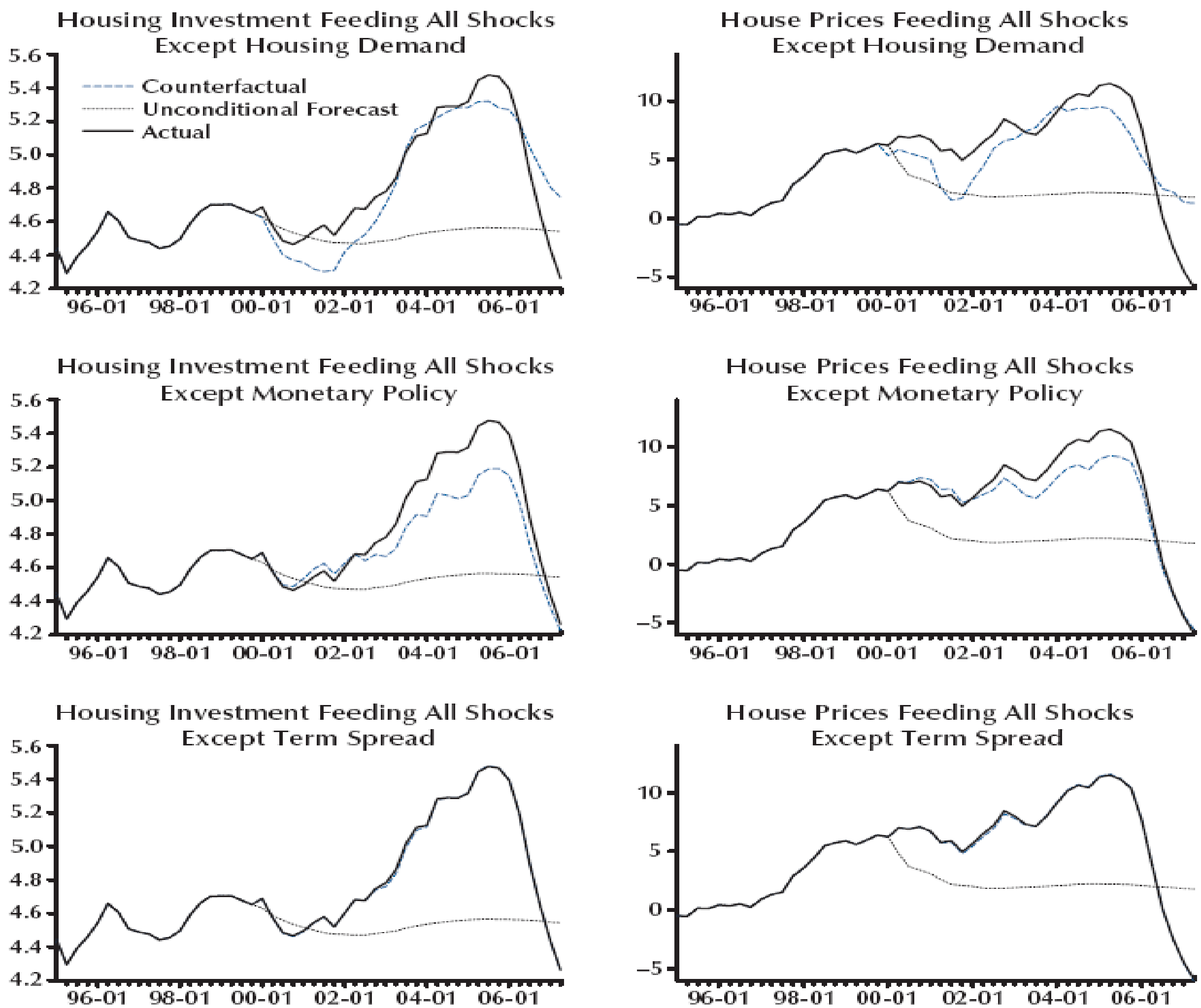
1. Estimate a VAR nine variables (including: short term interest rate, interest rate spread, housing investment share of GDP, real GDP, real consumption, real hours prices, prices, commodity price index and a money indicator.

2. Identify the monetary policy shock using the restriction that the shock does not affect prices and output contemporaneously but affect the short term interest rate, the spread and the money stock and analyze the impulse response functions.

3. Shut down the identified shock and study the counterfactual path of housing prices over time.



Source: Jarocinki and Smets (2008)



Source: Jarocinki and Smets (2008)

Table 2A

Shares of Housing Demand, Monetary Policy, and Term Spread Shocks in Variance Decompositions, DVAR

Variable	Shock	Horizon 0	3	11	23
Output	Housing	0.016	0.034	0.052	0.062
	Monetary policy	0.000	0.004	0.021	0.039
	Term premium	0.000	0.003	0.015	0.028
Consumption	Housing	0.005	0.018	0.033	0.055
	Monetary policy	0.000	0.003	0.015	0.029
	Term premium	0.000	0.005	0.034	0.063
Prices	Housing	0.002	0.013	0.120	0.166
	Monetary policy	0.000	0.003	0.014	0.037
	Term premium	0.000	0.006	0.034	0.046
Housing investment	Housing	0.521	0.579	0.382	0.291
	Monetary policy	0.000	0.015	0.175	0.136
	Term premium	0.000	0.005	0.023	0.062
House prices	Housing	0.535	0.554	0.410	0.242
	Monetary policy	0.000	0.010	0.068	0.083
	Term premium	0.000	0.002	0.021	0.060
Commodity prices	Housing	0.027	0.028	0.041	0.085
	Monetary Policy	0.000	0.012	0.167	0.222
	Term premium	0.000	0.004	0.018	0.055
Interest rate	Housing	0.037	0.061	0.165	0.178
	Monetary policy	0.752	0.496	0.192	0.166
	Term premium	0.000	0.023	0.076	0.088
Spread	Housing	0.090	0.050	0.177	0.186
	Monetary policy	0.223	0.303	0.214	0.206
	Term premium	0.336	0.245	0.146	0.134
Money	Housing	0.060	0.044	0.062	0.099
	Monetary policy	0.204	0.141	0.044	0.045
	Term premium	0.013	0.042	0.129	0.135

NOTE: The reported shares are averages over the posterior distribution and relate to the (log) level variables.

Source: Jarocinki and Smets (2008)

Blanchard Quah (1989) aggregate demand and supply shocks

Blanchard and Quah proposed an identification scheme based on long run restrictions.

In their model there are two shocks: an aggregate demand and an aggregate supply disturbance.

The restriction used to identify is that aggregate demand shocks have no effects on the long run levels of output, i.e. demand shocks are transitory on output. The idea behind of such a restriction is the existence of a vertical aggregate supply curve.

Let us consider the following bivariate VAR

$$\begin{pmatrix} \Delta Y_t \\ U_t \end{pmatrix} = \begin{pmatrix} F_{11}(L) & F_{12}(L) \\ F_{21}(L) & F_{22}(L) \end{pmatrix} \begin{pmatrix} w_t^s \\ w_t^d \end{pmatrix}$$

where Y_t is output, U_t is the unemployment rate and w_t^s, w_t^d are two aggregate supply and demand disturbances respectively.

The identification restriction is given by $F_{12}(1) = 0$.

The restriction can be implemented in the following way. Let us consider the reduced form VAR

$$\begin{pmatrix} \Delta Y_t \\ U_t \end{pmatrix} = \begin{pmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

where $E(\epsilon_t \epsilon_t') = \Omega$.

Let $S = chol(A(1)\Omega A(1)')$ and $K = A(1)^{-1}S$. The identified shocks are

$$w_t = K^{-1}\epsilon_t$$

and the resulting impulse response to structural shocks are

$$F(L) = A(L)K$$

notice that the restrictions are satisfied

$$\begin{aligned} F(1) &= A(1)K \\ &= A(1)A(1)^{-1}S \\ &= S \end{aligned}$$

which is lower triangular implying that $F_{12}(1) = 0$.

Moreover we have that shocks are orthogonal since

$$
\begin{aligned}
KK' &= A(1)^{-1}SS'A(1)^{-1\prime} && (37) \\
&= A(1)^{-1}A(1)\Omega A(1)'A(1)^{-1\prime} \\
&= \Omega \\
& && (38)
\end{aligned}
$$

And

$$
\begin{aligned}
E(w_t w_t') &= E(K^{-1}\epsilon_t \epsilon_t' K^{-1\prime}) \\
&= K^{-1}\Omega K^{-1\prime} \\
&= K^{-1}KK'K^{-1\prime}
\end{aligned}
$$

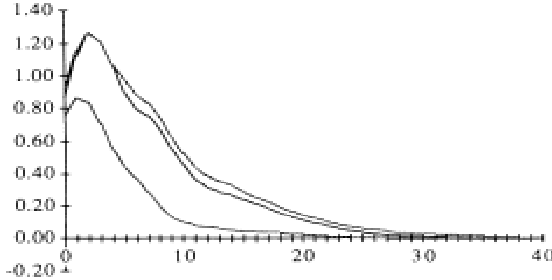


FIGURE 3. OUTPUT RESPONSE TO DEMAND

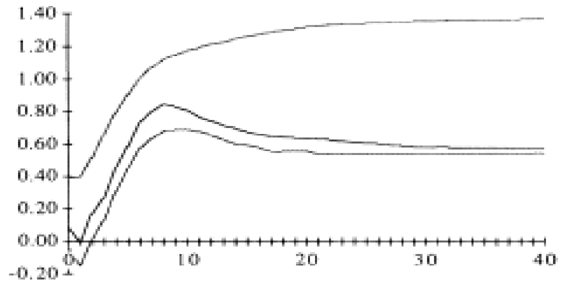


FIGURE 4. OUTPUT RESPONSE TO SUPPLY

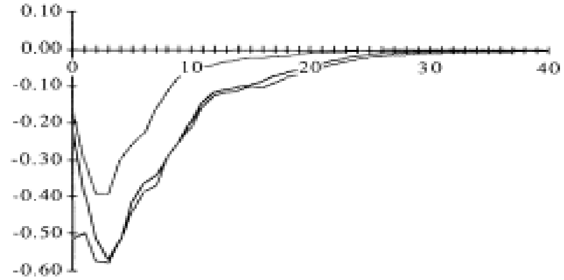


FIGURE 5. UNEMPLOYMENT RESPONSE TO DEMAND

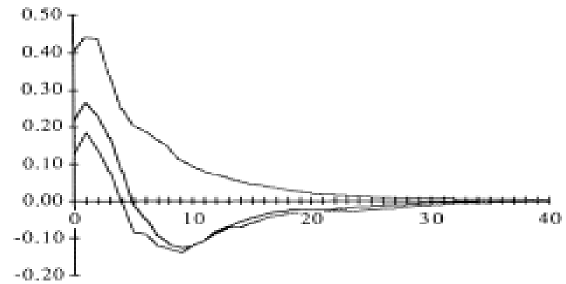


FIGURE 6. UNEMPLOYMENT RESPONSE TO SUPPLY

Source: The Dynamic Effects of Aggregate Demand and Supply Disturbances, (AER) Blanchard and Quah (1989):

TABLE 2—VARIANCE DECOMPOSITION OF OUTPUT AND UNEMPLOYMENT
(CHANGE IN OUTPUT GROWTH AT 1973/1974; UNEMPLOYMENT DETRENDED)

Percentage of Variance Due to Demand:		
Horizon (Quarters)	Output	Unemployment
1	99.0	51.9
	(76.9, 99.7)	(35.8, 77.6)
2	99.6	63.9
	(78.4, 99.9)	(41.8, 80.3)
3	99.0	73.8
	(76.0, 99.6)	(46.2, 85.6)
4	97.9	80.2
	(71.0, 98.9)	(49.7, 89.5)
8	81.7	87.3
	(46.3, 87.0)	(53.6, 92.9)
12	67.6	86.2
	(30.9, 73.9)	(52.9, 92.1)
40	39.3	85.6
	(7.5, 39.3)	(52.6, 91.6)

Source: The Dynamic Effects of Aggregate Demand and Supply Disturbances, (AER) Blanchard and Quah (1989):

TABLE 2A—VARIANCE DECOMPOSITION OF OUTPUT AND UNEMPLOYMENT
(NO DUMMY BREAK, TIME TREND IN UNEMPLOYMENT)

Percentage of Variance Due to Demand:		
Horizon (Quarters)	Output	Unemployment
1	83.8	79.7
	(59.4, 93.9)	(55,3, 92.0)
2	87.5	88.2
	(62.8, 95.4)	(58.9, 95.2)
3	83.4	93.5
	(58.8, 93.3)	(61.3, 97.5)
4	78.9	95.7
	(53.5, 90.0)	(63.9, 98.2)
8	52.5	88.9
	(31.4, 68.6)	(63.5, 94.5)
12	37.8	79.7
	(21.3, 51.4)	(58.8, 90.3)
40	18.7	75.9
	(7.4, 23.5)	(56.9, 88.6)

Source: The Dynamic Effects of Aggregate Demand and Supply Disturbances, (AER) Blanchard and Quah (1989):

The technology shocks and hours debate

This is a nice example of how SVAR models can be used in order to distinguish among competing models of the business cycles.

1) RBC technology important source of business cycles.

2) Other models (sticky prices) tech shocks not so important.

Response of hours worked very important in distinguish among theories

1) RBC hours increase.

2) Other hours fall

The model Technology shock: $z_t = z_{t-1} + \eta_t$ $\eta_t =$ technology shock

Monetary Policy: $m_t = m_{t-1} + \xi_t + \gamma\eta_t$ where $\xi_t =$ monetary policy shock.

Equilibrium:

$$\begin{aligned} \Delta x_t &= \left(1 - \frac{1}{\varphi}\right)\Delta\xi_t + \left(\frac{1-\gamma}{\varphi} + \gamma\right)\eta_t + (1-\gamma)\left(1 - \frac{1}{\varphi}\right)\eta_{t-1} \\ n_t &= \frac{1}{\varphi}\xi_t - \frac{(1-\gamma)}{\varphi}\eta_t \end{aligned}$$

or

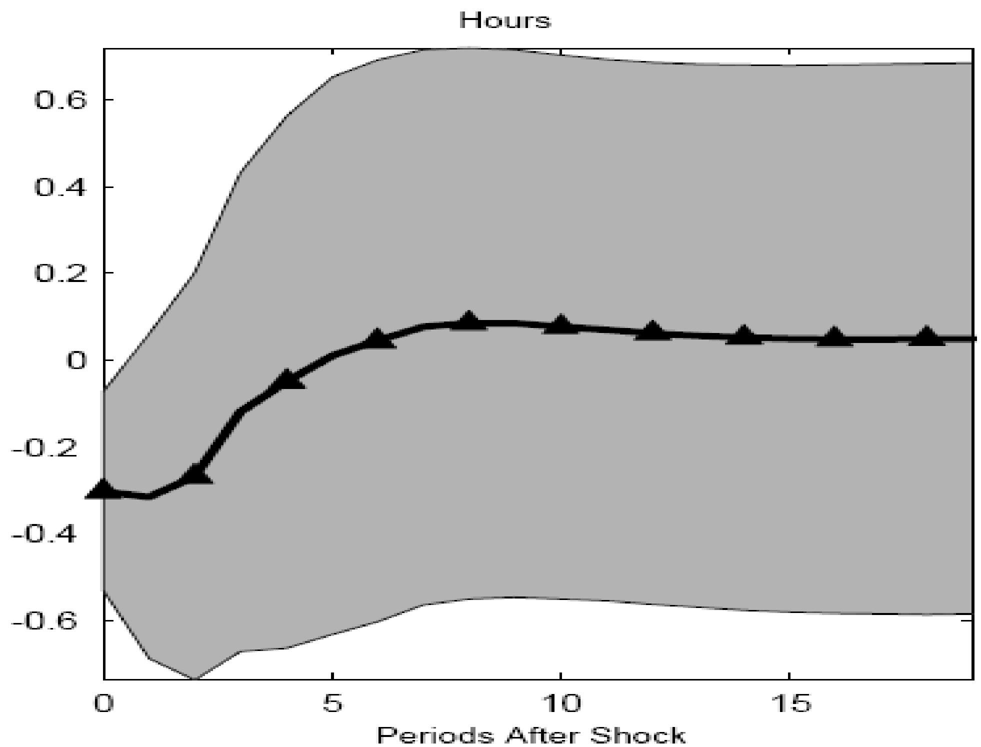
$$\begin{pmatrix} \Delta x_t \\ n_t \end{pmatrix} = \begin{pmatrix} \left(\frac{1-\gamma}{\varphi} + \gamma\right) + (1-\gamma)\left(1 - \frac{1}{\varphi}\right)L & \left(1 - \frac{1}{\varphi}\right)(1-L) \\ \frac{-(1-\gamma)}{\varphi} & \frac{1}{\varphi} \end{pmatrix} \begin{pmatrix} \eta_t \\ \xi_t \end{pmatrix} \tag{39}$$

In the long run $L = 1$

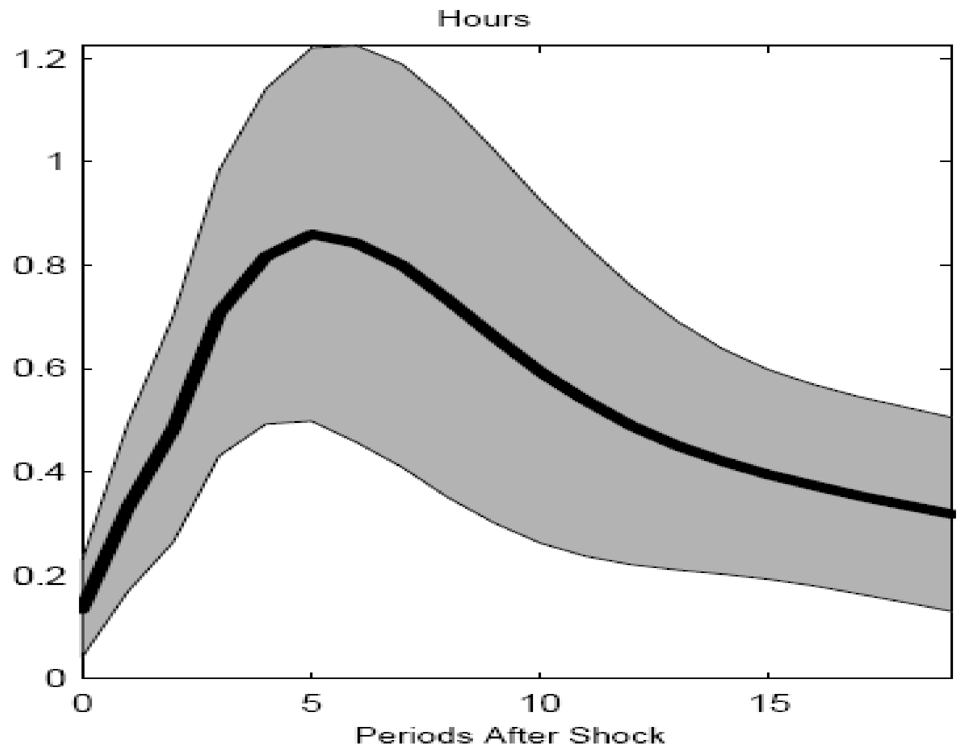
$$\begin{pmatrix} \Delta x_t \\ n_t \end{pmatrix} = \begin{pmatrix} \left(\frac{1-\gamma}{\varphi} + \gamma\right) + (1-\gamma)\left(1 - \frac{1}{\varphi}\right) & 0 \\ \frac{-(1-\gamma)}{\varphi} & \frac{1}{\varphi} \end{pmatrix} \begin{pmatrix} \eta_t \\ \xi_t \end{pmatrix} \tag{40}$$

that is only the technology shocks affects labor productivity.

Note the model prediction. If monetary policy is not completely accomodative $\gamma < 1$ then the response of hours to a technology shock $\frac{-(1-\gamma)}{\varphi}$ is negative.



Source: What Happens After a Technology Shock?... Christiano Eichenbaum and Vigfusson NBER WK (2003)



Source: What Happens After a Technology Shock?... Christiano Eichenbaum and Vigfusson NBER WK (2003)

Table 1: Contribution of Technology Shocks to Variance, Bivariate System

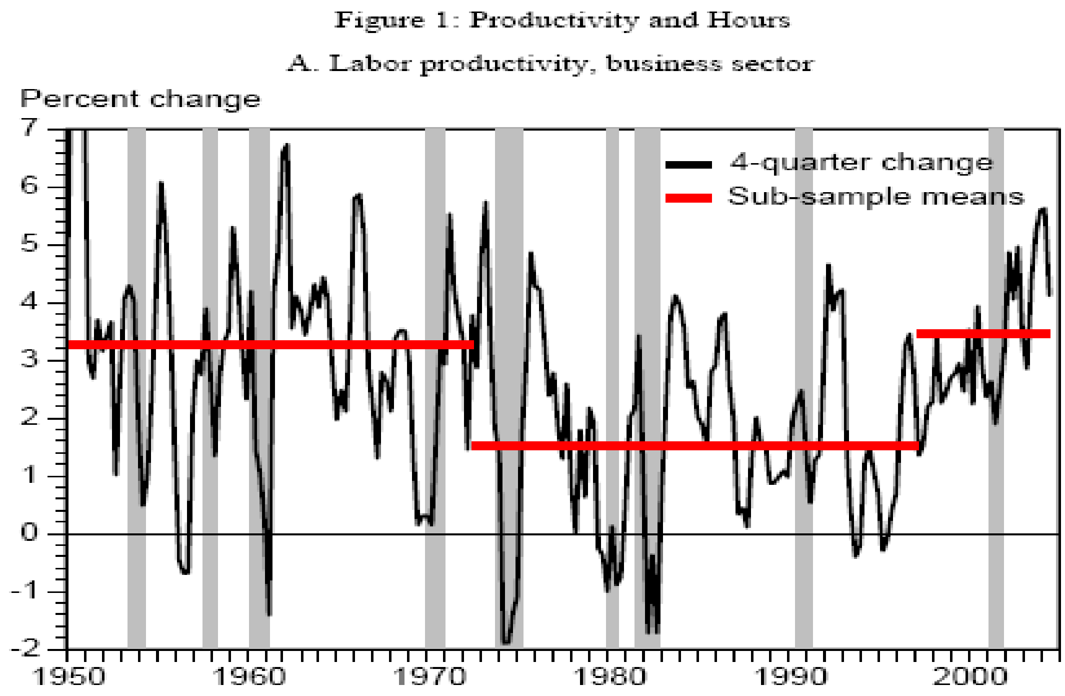
Level Specification

	Forecast Variance at Indicated Horizon					
Variable	1	4	8	12	20	50
Output	81.1	78.1	86.0	89.1	91.8	96
Hours	4.5	23.5	40.7	45.4	47.4	48.3

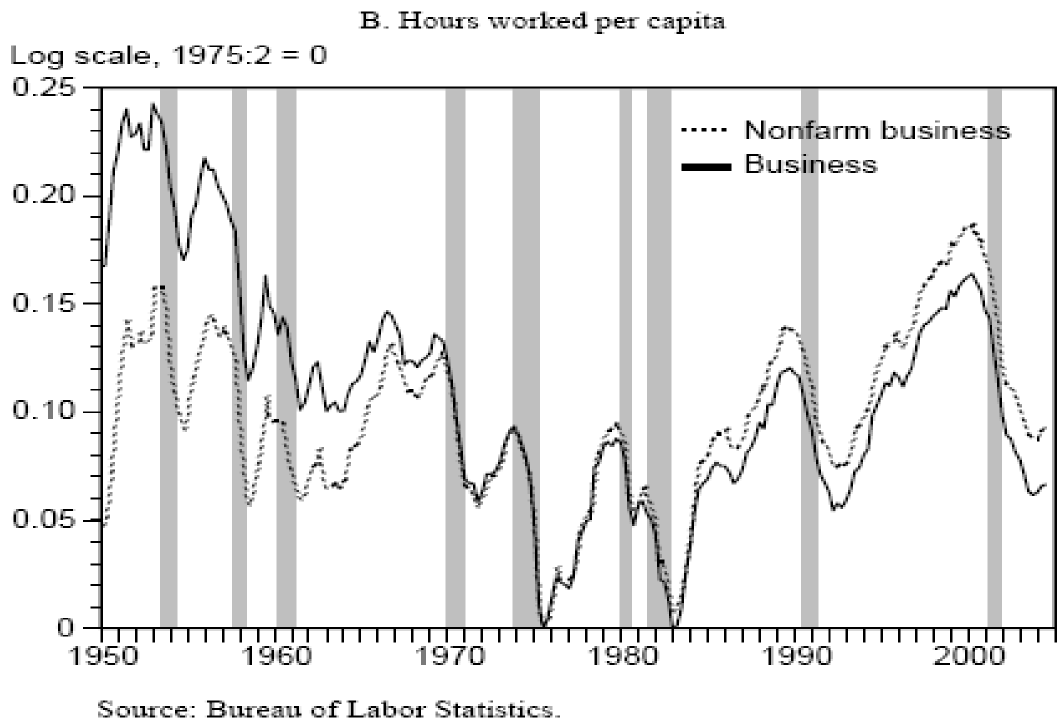
Difference Specification

	Forecast Variance at Indicated Horizon					
Variable	1	4	8	12	20	50
Output	16.5	11.7	17.9	20.7	22.3	23.8
Hours	21.3	6.4	2.3	1.6	1.0	0.5

Source: What Happens After a Technology Shock?... Christiano Eichenbaum and Vigfusson NBER WK (2003)



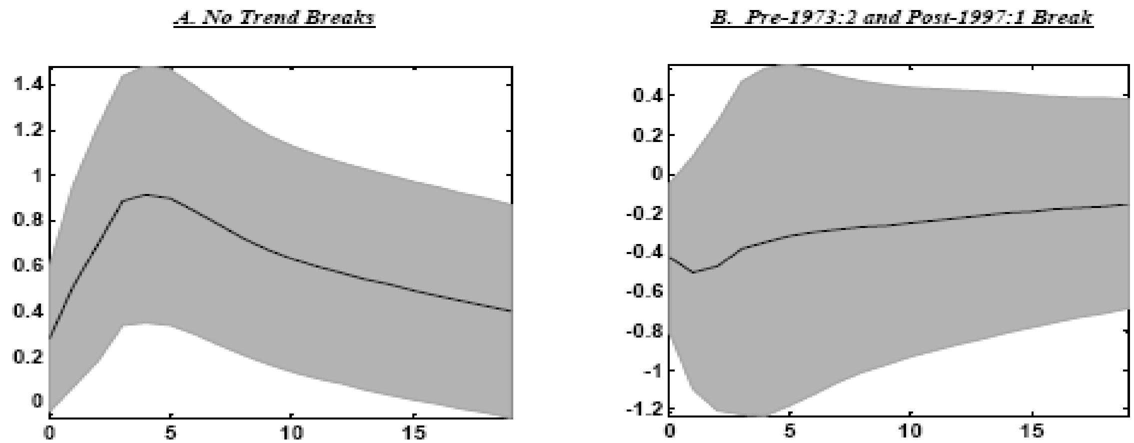
Source: Trend Breaks, Long-Run Restrictions, and Contractionary Technology Improvements, JME John Fernald (2007)



Source: Trend Breaks, Long-Run Restrictions, and Contractionary Technology Improvements, JME John Fernald (2007)

Figure 2. Impulse Responses from Bivariate Specification

Response of Hours to a Technology Shock



Source: Trend Breaks, Long-Run Restrictions, and Contractionary Technology Improvements, JME John Fernald (2007)

The effects of government spending shocks

Understanding the effects of government spending shocks is important for policy authorities but also to assess competing theories of the business cycle.

Keynesian theory: government spending, GDP, consumption and real wage $\uparrow$, (because of the government spending multiplier).

RBC theory: government spending $\uparrow$, but consumprion and the real weage $\downarrow$ because of a negative wealth effect.

Disagreement from the empirical point of view.

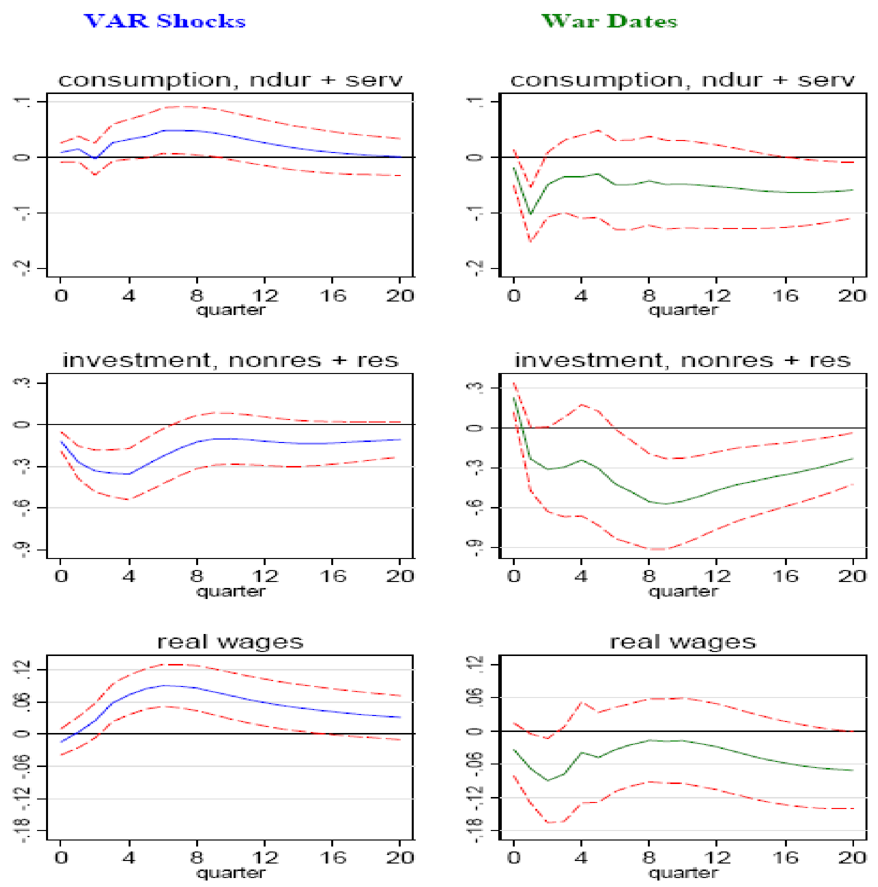
Government spending shocks: Blanchard and Perotti (2002) BP (originally) use a VAR for real per capita taxes, government spending, and GDP.

The shock is identified assuming that government spending does not react to taxes and GDP contemporaneously, Cholesky identification with government spending ordered first. The government spending shock is the first one (quadratic trend four lags).

When augmented with consumption consumption increases.

When augmented with investment, investment increases.

In a more recent version Perotti (2007) uses a larger VAR but the results are confirmed. Consumption and real wage $\uparrow$ but investment $\downarrow$



Source: IDENTIFYING GOVERNMENT SPENDING SHOCKS: IT'S ALL IN THE TIMING
Valerie A. Ramey NBER Working Paper 15464 (2009)

Government spending shocks: Ramey and Shapiro (1998) Ramey and Shapiro (1998) use a narrative approach to identify shocks to government spending.

Focus on episodes where Business Week suddenly forecast large rises in defense spending induced by major political events that were unrelated to the state of the U.S. economy (exogenous episodes of government spending).

Three of such episodes: Korean War, The Vietnam War and the Carter-Reagan Buildup + 9/11.

The military date variable takes a value of unity in 1950:3, 1965:1, 1980:1, and 2001:3, and zeros elsewhere.

To identify government spending shocks, the military date variable is embedded in the standard VAR, but ordered before the other variables.

Both methodologies have problems.

VARs: shocks are often anticipated (fiscal foresight shocks may be not invertible)

War Dummy: few observations, subjective, relies on the construction of an exogenous time series.

Possible extensions.

News Shocks, Beaudry and Portier (AER 2006)

- Main idea back to Pigou and Keynes: news about future productivity growth can generate business cycles since agents react to news by investing and consuming.

- Standard DSGE model have an hard time in generating these predictions because as agents feel richer they consume more work less and invest less. Aggregate variables move in opposite directions so news shocks cannot be the min source of fluctuations.

- BP uses a VECM for TFP and Stock prices plus other real variables to evaluate whether news shocks, in the data, do generate sizable fluctuations.

- Benchmark model a VECM for TFP and SP.

- Two identification procedures:

  1. Technology shocks is the only shock driving TFP in the long run.

  2. News shocks raise stock prices on impact but not TFP (lagged adjustment.

- Main finding:

  1. the two identified shocks are the same

  2. such shocks generate positive comovment in consumption, investment, output and hours (consistently with business cycles conmovements) and they explain a large portion of the variance of these series.

- Conclusion: news shocks can generate business cycles.

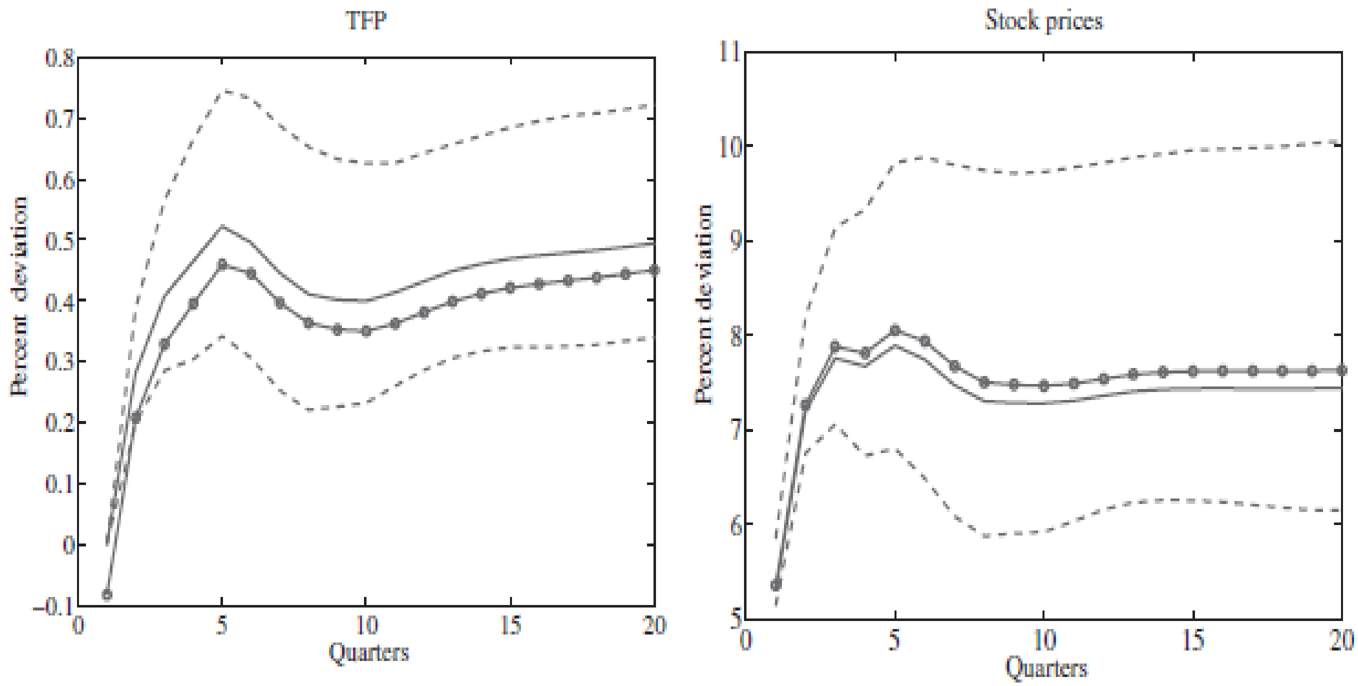


FIGURE 1. IMPULSE RESPONSES TO SHOCKS ε_2 AND $\tilde{\varepsilon}_1$ IN THE (*TFP*, *SP*) VECM

Source: Beaudry and Portier (AER 2006)

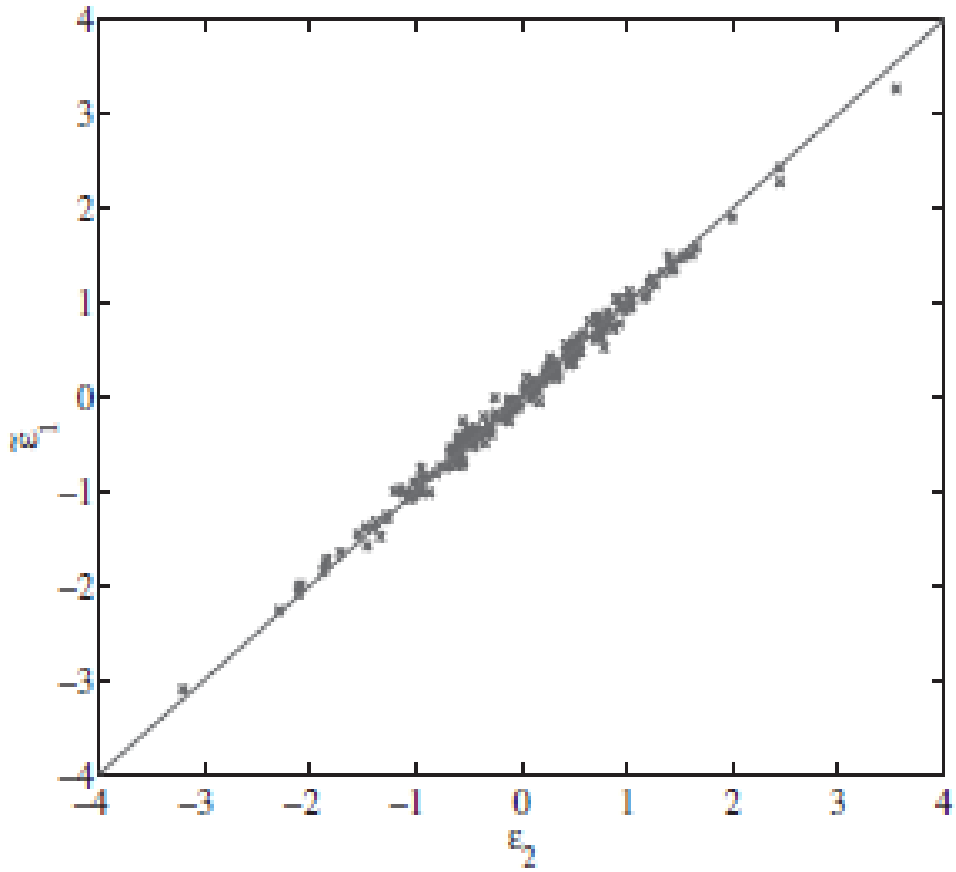


FIGURE 2. PLOT OF ε_2 AGAINST $\tilde{\varepsilon}_1$ IN THE (*TFP*, *SP*) VECM

Source: Beaudry and Portier (AER 2006)

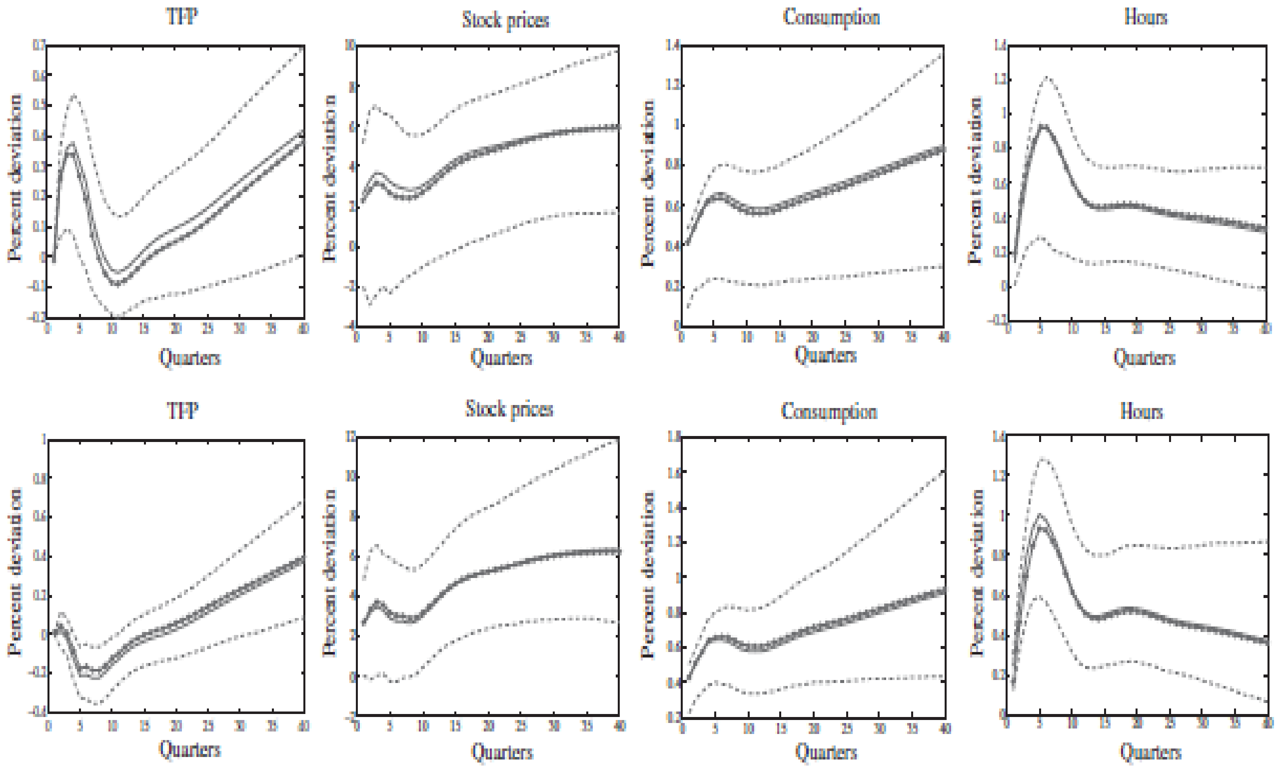


FIGURE 9. IMPULSE RESPONSES TO ε_2 AND $\tilde{\varepsilon}_1$ IN THE (*TFP*, *SP*, *C*, *H*) VECM, WITHOUT (UPPER PANELS) OR WITH (LOWER PANELS) ADJUSTING TFP FOR CAPACITY UTILIZATION

Source: Beaudry and Portier (AER 2006)

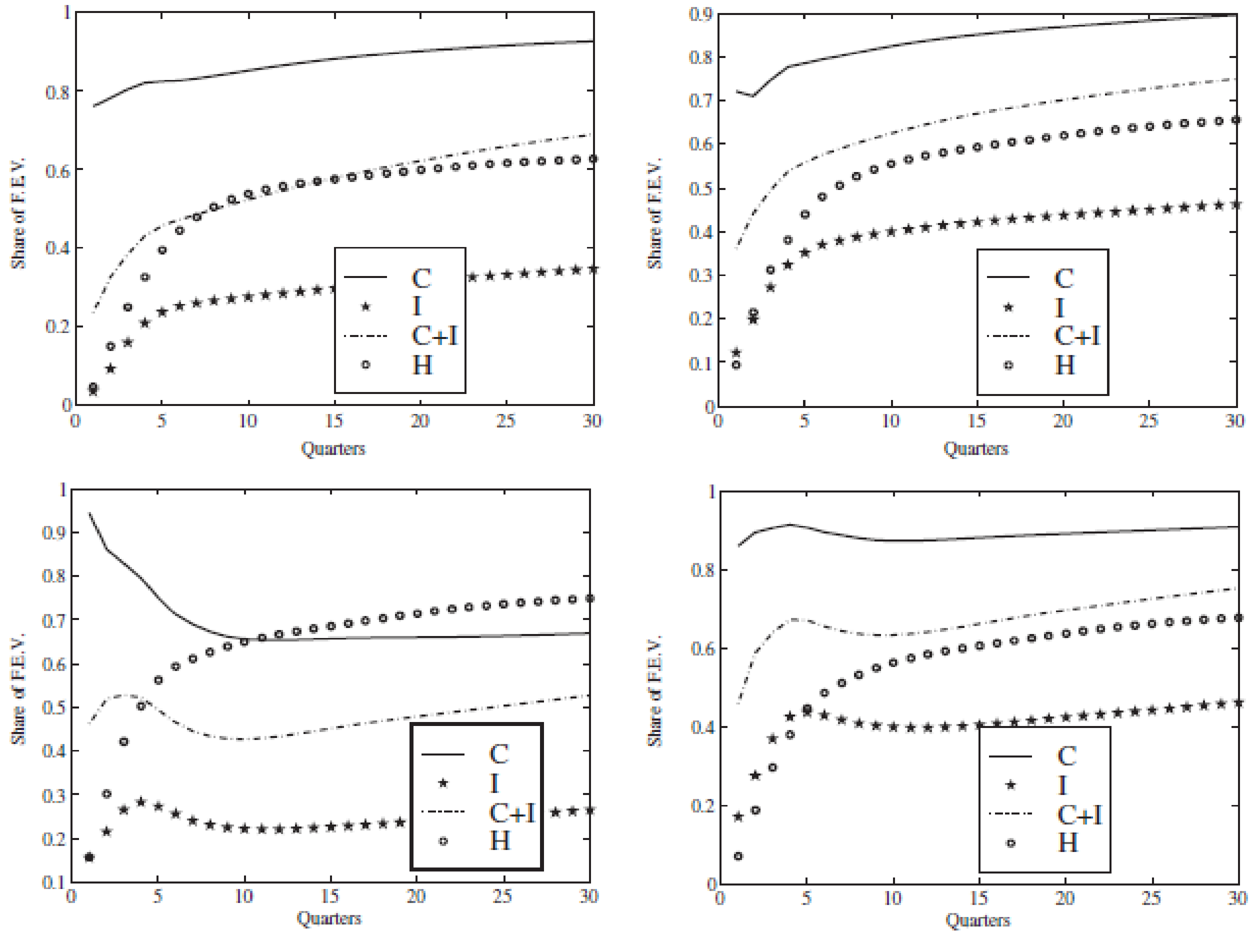


FIGURE 10. SHARE OF THE FORECAST ERROR VARIANCE (F.E.V.) OF CONSUMPTION (C), INVESTMENT I, OUTPUT ($C + I$), AND HOURS (H) ATTRIBUTABLE TO ε_2 (LEFT PANELS) AND TO $\tilde{\varepsilon}_1$ (RIGHT PANELS) IN VECMS, WITH NONADJUSTED TFP (TOP PANELS) OR ADJUSTED TFP (BOTTOM PANELS)

Source: Beaudry and Portier (AER 2006)

5 Cointegration

Cointegration

Definitions

• $I(1)$ *variables*. $z_t = (y_t\ x_t)$ is $I(1)$ (integrated of order 1) if it is not stationary but its first difference Δz_t is stationary.

It is easily seen that in general a linear combination of y_t and x_t is $I(1)$. However, in particular cases, it can be $I(0)$, i.e. stationary.

Example 1: Common trends Consider for instance the case

$$\begin{aligned} y_t &= T_t + c_t \\ x_t &= \beta T_t + e_t \end{aligned}$$

where T_t is a random walk process and c_t , e_t are co-stationary. $y_t - ax_t$ is $I(1)$ for $a \neq 1/\beta$ but is stationary for for $a = 1/\beta$.

• *Cointegration*. y_t and x_t are cointegrated of order 1 if and only if $z_t = (y_t\ x_t)'$ is $I(1)$ and there is a linear combination $y_t - ax_t$ which is stationary. The vector $\alpha = (1 \quad -a)'$ such that $\alpha' z_t$ is stationary is called the cointegrating vector.

• Examples are: money-prices, consumption-GDP labor productivity- real wages.

• *Example 2*. Consider the following system

$$\begin{aligned} Y_{1t} &= \gamma Y_{2t} + u_{1t} \\ Y_{2t} &= Y_{2t-1} + u_{2t} \end{aligned}$$

where u_{1t}, u_{2t} are WN. It easy to see that both Y_{1t}, Y_{2t} are I(1) but the linear combination $Y_{1t} - \gamma Y_{2t}$ is $I(0)$. Y_{1t}, Y_{2t} are cointegrated.

• *Cointegration of n variables*. The variables in $Y_t = (Y_{1t}, ..., Y_{nt})$ are cointegrated if they are jointly $I(1)$ and there exist a non-zero vector α such that $\alpha' Y_t \sim I(0)$.

Implications of cointegration

Example 2 contd. Considering again the processes described in Example 2, by taking the differences we have

$$\begin{aligned} \Delta Y_{1t} &= \gamma u_{2t} + u_{1t} - u_{1t-1} \\ \Delta Y_{2t} &= u_{2t} \end{aligned}$$

Let $\varepsilon_{2t} = u_{2t}$ and $\varepsilon_{1t} = \gamma u_{2t} + u_{1t}$ be the two forecast errors. Then

$$\Delta Y_t = C(L)\varepsilon_t$$

where

$$C(L) = \begin{pmatrix} 1-L & \gamma \\ 0 & 1 \end{pmatrix}$$

The problem is that the matrix associated with the moving average operator for this process $C(z)$ has a root at unity

$$|C(1)| = \left| \begin{pmatrix} 1-1 & \gamma \\ 0 & 1 \end{pmatrix} \right| = 0$$

implying that the MA representation is not invertible implying that no finite order VAR can describe ΔY_t. This is the general problem raised by cointegration.

Error correction mechanism

• A well-known model in time series macroeconometrics is the so called Error-Correction Mechanism (ECM):

$$A(L)\Delta Y_t = \beta\alpha' Y_{t-1} + \varepsilon_t \tag{41}$$

Engle and Granger two-step testing procedure

• The simplest way to test for cointegration was suggested in Engle and Granger (1987):

1. estimate by OLS the regression equation

$$x_t = b + a y_t + e_t$$

2. take the estimated residuals $\widehat{e}_t$ and test for stationarity by using the ADF test.

• Once the cointegrating vector has been estimated the remaining parameters can be estimated with OLS using $\widehat{\alpha}' Y_{t-1}$

6.Large-N models: Factor Models and FAVAR

Motivation

• VARs include usually a small amount of variables (6-8 at most)

• Problem: agents usually use much more information than the one included in the VAR to take decisions. Think of monetary policy authorities when deciding how to set the interest rate.

• What happens in that case? Identified shocks are combinations of innovations with respect to a *wrong* information set. So misleading impulse response analysis. An example: the price puzzle.

• Hansen and Sargent (1980) show that when economic agents have mnore information than that available to the econometrician (as may be in the case of VARs) a problem of non-fundamentalness can arise.

• To overcome the problem we would like to extend the information used to identify economic shocks. How? Two alternatives

1) Structural factors models

2) FAVAR models

Nonfundamentalness again Nonfundamentalness intuitively arises when economic agents have more information than the econometrician.

In factor models, the possibility of using large information sets halps in aligning the information set of the economic agents and that of the econometrician.

There is also a technical reason why nonfundamentalness should not be considered a serious threat in SFM. It will be clear later. So far it instructive to review the definition of fundamentalness.

Definition Assume that the n-dimensional stochastic vector μ_t admits a moving average representation

$$\mu_t = K(L)v_t$$

where $K(L)$ is a $n \times q$ $(q \leq n)$ polynomial matrix and v_t is a $q \times 1$ white noise. The above representation is fundamental if and only if the rank of $K(L)$ is q for all z such that $|z| < 1$. If it is fundamental all the fundamental WN vectors are linear transformations of v_t.

If $n = q$ (the SVAR case) if one root is smaller than one in absolute value the system is non fundamental.

If $n < q$ (the rectangular case) nonfundamentalness requires that there is a z smaller than one in absolute value for which the rank of $K(L)$ is smaller than q. But this requires that there is a common root in all the $q \times q$ submatrices of $K(L)$, which is a much more restrictive condition.

Example To understand how large information can mitigate the nonfundamentalness problem consider the two MA

$$X_t = (1 + 2L)\varepsilon_t \quad Y_t = L\varepsilon_t$$

both are nonfundamental because the absolute root in the first is 0.5 and in the second is 0. However the process

$$Z_t = X_t - 2Y_t = \varepsilon_t \tag{42}$$

is obviously fundamental.

Limited Information: Economic Examples

Fiscal Foresight (Leeper Yang and Walker, 2008). The model is a standard growth model with log preferences, inelastic labor supply and complete depreciation of capital. Log-linearized equilibrium solution for the three state variables of the model is

$$\begin{aligned} k_t &= \alpha k_{t-1} + a_t - (1-\theta)(\frac{\tau}{1-\tau})\sum_{i=0}^{\infty}\theta^i E_t \hat{\tau}_{t+i+1} \\ a_t &= \varepsilon_{A,t} \\ \tau_t &= \varepsilon_{\tau,t-q} \end{aligned}$$

where k_t is capital a_t is the technology shock and τ_t are taxes, $\varepsilon_{\tau,t-q}, \varepsilon_{A,t}$ are i.i.d. shocks to taxes and technolgy, $\theta = \alpha\beta(1-\tau) < 1$, where $0 < \alpha < 1$ is the $0 < \beta < 1$, and $0 \leq \tau < 1$ is the steady state tax rate. q is the period of foresight; for instance is $q = 1$ agents observe a shock which will affect taxes next period.

Let us consider the case when agents anticipate future tax changes. Suppose that $q = 2$. The capital transition equation becomes

$$k_t = \alpha k_{t-1} + a_t - (1-\theta)(\frac{\tau}{1-\tau})(\varepsilon_{\tau,t-1} + \theta\varepsilon_{\tau,t})$$

Suppose the econometrician wants to use data for capital and technology to estimate a VAR in order to identify the fiscal shock. The solution of the model for the two variables is

$$\begin{pmatrix} a_t \\ k_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{-\kappa(L+\theta)}{1-\alpha L} & \frac{1}{1-\alpha L} \end{pmatrix} \begin{pmatrix} \varepsilon_{\tau,t} \\ \varepsilon_{A,t} \end{pmatrix} = A_1(L)\varepsilon_t \tag{43}$$

where $\kappa = (1-\theta)(\tau/(1-\tau))$, $\theta = \alpha\beta(1-\tau)$. The determinant of $A_1(z)$ is $\frac{\kappa(z+\theta)}{1-\alpha z}$ which is zero for $z = -\theta$.

This implies that the shock cannot be recovered using a VAR with data for capital and technology.

Now suppose that the econometrician decides to use data for capital and taxes. The solution of the model for the two variables is

$$\begin{pmatrix} k_t \\ \tau_t \end{pmatrix} = \begin{pmatrix} \frac{-\kappa(L+\theta)}{1-\alpha L} & \frac{1}{1-\alpha L} \\ L^2 & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_{\tau,t} \\ \varepsilon_{A,t} \end{pmatrix} = A_2(L)\varepsilon_t \tag{44}$$

The determinant of $A_1(z)$ is $\frac{z}{1-\alpha z}$ which is zero for $z = 0$ meaning that the MA representation is non-invertible and the shock non-fundamental for τ_t and k_t. Again the shock cannot be recovered using a VAR with data for capital and taxes.

If the shocks are nonfundamental then SVAR models are not useful for structural analysis.

However if we consider the three variables

$$\begin{pmatrix} a_t \\ k_t \\ \tau_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{-\kappa(L+\theta)}{1-\alpha L} & \frac{1}{1-\alpha L} \\ L^2 & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_{\tau,t} \\ \varepsilon_{A,t} \end{pmatrix} = A_1(L)\varepsilon_t \tag{45}$$

the system is fundamental.

News shocks.

$$E_t \sum_{t=0}^{\infty} \beta^t C_t,$$

where C_t is consumption and β is a discount factor, subject to the constraint

$$C_t + P_t S_{t+1} \quad = \quad (P_t + \theta_t) S_t,$$

where P_t is the price of a share, S_t is the number of shares and $(P_t + \theta_t)S_t$ is the total amount of resources available at time t. The equilibrium value for asset prices is given by:

$$P_t = E_t \sum_{j=1}^{\infty} \beta^j \theta_{t+j}$$

Considering (1), the above equation can be solved to get the following structural MA representation

$$\begin{pmatrix} \Delta\theta_t \\ \Delta P_t \end{pmatrix} = \begin{pmatrix} L^2 & 1 \\ \frac{\beta^2}{1-\beta} + \beta L & \frac{\beta}{1-\beta} \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}. \tag{46}$$

The determinant is

$$-\frac{\beta^2}{1-\beta} - \beta z + \frac{\beta}{1-\beta} z^2$$

which vanishes for $z = 1$ and $z = -\beta$. As $\beta < 1$, the moving average is non invertible and the two shocks u_t and ε_t are non-fundamental for the variables ΔP_t and $\Delta\theta_t$.

Here the agents see the shocks. The econometrician only see the variables. In this case not even a very forward-looking variable like stock prices conveys enough information to recover the shock.

The role of information: introducing the factor model

To understand the role of information for fundamentalness, let us start from the $ABCD$ state-space representation of a macroeconomic equilibrium studied in Fernandez-Villaverde *et al.*, 2007:

$$\begin{aligned} s_t &= As_{t-1} + Bu_t & (47) \\ \chi_t &= Cs_{t-1} + Du_t & (48) \end{aligned}$$

where $\chi_t = (\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{nt})'$, is an n-dimensional vector of macroeconomic variables; s_t is an l-dimensional vector of "state" variables, $l \leq n$; u_t is a q-dimensional vector, $q \leq l$, of serially uncorrelated structural shocks, orthogonal to s_{t-k}, $k = 1, \ldots, \infty$; A, B, C and D are conformable matrices of parameters and B has a left inverse B^{-1} such that $B^{-1}B = I_q$.

Observe that the macroeconomic shocks u_t are linear combinations of present and lagged states, since, pre-multiplying (47) by B^{-1} we get

$$u_t = B^{-1}s_t - B^{-1}As_{t-1}. \tag{49}$$

Equation (49) shows that the states contain enough information to recover the shocks, or, in other words, that the shocks are always fundamental with respect to the states.

Substituting (49) into (48) and rearranging gives

$$\chi_t = DB^{-1}s_t + (C - DB^{-1}A)s_{t-1}. \tag{50}$$

Now let us assume that the econometrician observes $x_{it} = \chi_{it} + \xi_{it}$, ξ_{it} being a measurement error (which can be zero) and define $x_t = (x_{1t}\ x_{2t}\ \cdots\ x_{nt})'$, $\xi_t = (\xi_{1t}\ \xi_{2t}\ \cdots\ \xi_{nt})'$. From equation (50) it is seen that x_t follows the factor model

$$x_t = \Lambda f_t + \xi_t, \tag{51}$$

where $\Lambda = \begin{pmatrix} DB^{-1} & C - DB^{-1}A \end{pmatrix}$ and $f_t = \begin{pmatrix} s'_t & s'_{t-1} \end{pmatrix}'$. Since the factors f_t include the states, the structural shocks are always fundamental with respect to the factors as long as the macroeconomic equilibrium can be represented in the form (47)–(48).

The structural factor model (SFM)

Factor model have been originally introduced by Geweke (1977) and Sargent and Sims (1977). They have been used mainly for forecasting.

An alternative to SVAR models for structural analysis is represented by Structural Factor Models (SFM) (Forni, Lippi, Giannone and Rechlin, 2009, Stock and Watson, 2005).

The main difference with SVAR is represented by the fact that SFM can use a larger amount of information (100/200 time series vs 6/8).

Information is important for several reasons:

1. Better characterize the shocks

2. Less likely to be subject of misspecification

3. Response of many potentially interesting variables can be studied

4. Nonfundamentalness is not an issue in these models.

The Model

We assume that the panel has the following representation

$$X_t = A_n f_t + \xi_t, \tag{52}$$

$$D(L) f_t = \epsilon_t \tag{53}$$

$$\epsilon_t = R u_t$$

where

x_t − a vector containing the n variables of the panel.

$A_n f_t$ − the common component.

f_t − a vector containing $r < n$ unobserved factors (static factors, here r=16).

u_t − a vector containing $q < r$ structural macro shocks (dynamic factors, here q=4)).

R − a $r \times q$ matrix of coefficients.

$D(L)$ − a $r \times r$ matrix of polynomials in the lag operator.

ξ_t − a vector of n idiosyncratic components (orthogonal to the common one, poorly correlated in the cross-sectional dimension (the approximate factor model), and capturing sector specific variations, non-US shocks or measurement errors).

From (1)-(2) We can derive the dynamic representation of the model (in terms of structural shocks)

$$x_t = B(L)u_t + \xi_t \tag{54}$$

where $B_n(L) = A_n D(L)^{-1} R$ – a $n \times q$ matrix of impulse response functions to structural shocks.

Notice that the fact that $q < r$ makes $D(L)^{-1}$ a rectangular where the conditions for fundamentalness are those described before.

Identification

$B_n(L)$ is identified up to an orthogonal $(q \times q)$ matrix H (such that $HH' = I$) since $B_n(L)u_t = C_n(L)v_t$ where $B_n(L) = C_n(L)H$ and $v_t = H'u_t$.

In this context identification consists in imposing economically-based restrictions on $B_n(L)$ to determine a particular H. This is the same as in VAR but restriction can be imposed on a $n \times q$ matrix of responses.

In practice, given a matrix of nonstructural impulse response functions $\widehat{C_n(L)}$ obtained as described in the estimation one has to choosing H by imposing some restrictions on $B_n(L)$.

Same types of restrictions used in VAR: Cholesky, long run, signs etc.

Consistent Estimator of impulse response functions

The estimation is performed as follows.

1. Given $\widehat{r}$, the static factors are estimated by means of the first $\widehat{r}$ ordinary principal components of the variables in the data set. Let $\widehat{\Gamma}^x$ be the sample variance-covariance matrix of the data: the estimated loading matrix $\widehat{A}_n = (\widehat{a}_1' \widehat{a}_2' \cdots \widehat{a}_n')'$ is the $n \times r$ matrix having on the columns the normalized eigenvectors corresponding to the first largest $\widehat{r}$ eigenvalues of $\widehat{\Gamma}^x$, and the estimated factors are $\widehat{f}_t = \widehat{A}_n'(x_{1t} x_{2t} \cdots x_{nt})'$.

2. Second, a VAR(p) is run with $\widehat{f}_t$ to get estimates of $D(L)$ and the residuals ϵ_t, say $\widehat{D}(L)$ and $\widehat{\epsilon}_t$.

3. Now, let $\widehat{\Gamma}^\epsilon$ be the sample variance-covariance matrix of $\widehat{\epsilon}_t$. Given $\widehat{q}$, let $\widehat{\mu}_j^\epsilon$, $j = 1, \ldots, \widehat{q}$, be the j-th eigenvalue of $\widehat{\Gamma}^\epsilon$, in decreasing order, $\widehat{\mathcal{M}}$ the $q \times q$ diagonal matrix with $\sqrt{\widehat{\mu}_j^\epsilon}$ as its (j, j) entry, $\widehat{K}$ the $r \times q$ matrix with the corresponding normalized eigenvectors on the columns. Setting $\widehat{S} = \widehat{K}\widehat{\mathcal{M}}$, the estimated matrix of non-structural impulse response functions is
$$\widehat{C}_n(L) = \widehat{A}_n \widehat{D}(L)^{-1} \widehat{S}. \tag{55}$$

4. Finally, $\widehat{H}$ and $\widehat{b}_i(L) = \widehat{c}_i(L)\widehat{H}$ $i = 1, \ldots, n$ are obtained by imposing the identification restrictions on
$$\widehat{B}_m(L) = \widehat{C}_m(L)\widehat{H}. \tag{56}$$

Proposition 3 of FGLR states that $\widehat{b}_i(L)$, for a fixed i, is a consistent estimator of $b_i(L)$.

Inference

Confidence bands are obtained by a standard non-overlapping block bootstrap technique.

1. Let $X = [x_{it}]$ be the $T \times n$ matrix of data. Such matrix is partitioned into S sub-matrices X_s (blocks), $s = 1, \ldots, S$, of dimension $\tau \times n$, τ being the integer part of T/S.

2. An integer h_s between 1 and S is drawn randomly with reintroduction S times to obtain the sequence $h_1, \ldots, h_S$.

3. A new artificial sample of dimension $\tau S \times n$ is then generated as $X^* = \left[X'_{h_1} X'_{h_2} \cdots X'_{h_S}\right]'$ and the corresponding impulse response functions are estimated.

4. A distribution of impulse response functions is obtained by repeating drawing and estimation.

Determination of the number of factors

There are criteria available for the determination of the number of both static and dynamic factors.

# of static factors r Bai and Ng (2002) proposed a set of consistent criteria to determine the number of static factors. The idea is similar to the one behind information criteria to determine the number of lags in VAR. The most common one is the $IC_{p2}(r)$. r should be chosen in order to minimize

$$IC_{p2}(r) = \ln V(r, \hat{f}_t) + r\left(\frac{n+T}{nT}\right)\ln\left(Min(n,T)\right)$$

where $V(r, \hat{f}_t)$ is the sum of residuals (divided by (nT) from the regression of x_i on the r factors for all i,

$$V(r, \hat{f}_t) = \min_A \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - A_i^r f_t^r)^2$$

# of dynamic factors q Bai and Ng (2007) based on the rank of the residual covariance matrix.

Amengual and Watson (2008). Regress x on $\hat{f}_t$ and apply Bai and Ng (2002) to the new obtained residuals to study the number of dynamic factors.

Detecting nonfundamentalness in VAR

A strategy to detect nonfundamentalness in VAR is the following.

Suppose we want to know whether the past of q variables of interest, say x_q in the panel are sufficient to recover the structural shocks. That is if a VAR with these q variables can be used.

To test whether the shock is fundamental for x_q we can calculate the roots of the associated impulse response functions $C_q(L)$ and check whether there are roots smaller than one in modulus. If we find at least one this means that a VAR for x_q cannot be used to estimate the shock.

An application: Forni and Gambetti (2010, JME)

1. Motivation: standard theory of monetary policy predicts that after a contractionary policy shock:
   (a) Prices fall
   (b) The real exchange rate immediately appreciates and then depreciates (overshooting theory)
2. With VAR puzzling results:
   (a) Prices increase (price puzzle)
   (b) The real exchange rate appreciates with a long delay (delayed overshooting puzzle)
3. Here: we study the effects of monetary policy shocks whithin a SFM.
4. Why: information could be the key.
5. Main result: both of them solved IRF behave like theory predicts.

Data: 112 US monthly series from March 1973 to November 2007. Most series are those of the Stock-Watson, we added a few real exchange rates and short-term interest rate spreads between US and some foreign countries.

The monetary policy shock is identified by the following assumptions:

1. the monetary policy shock is orthogonal to all other structural shocks,
2. the monetary policy shock has no contemporaneous effect on prices and output (Cholesky scheme).

Table 1: Variance decomposition SVAR (*)

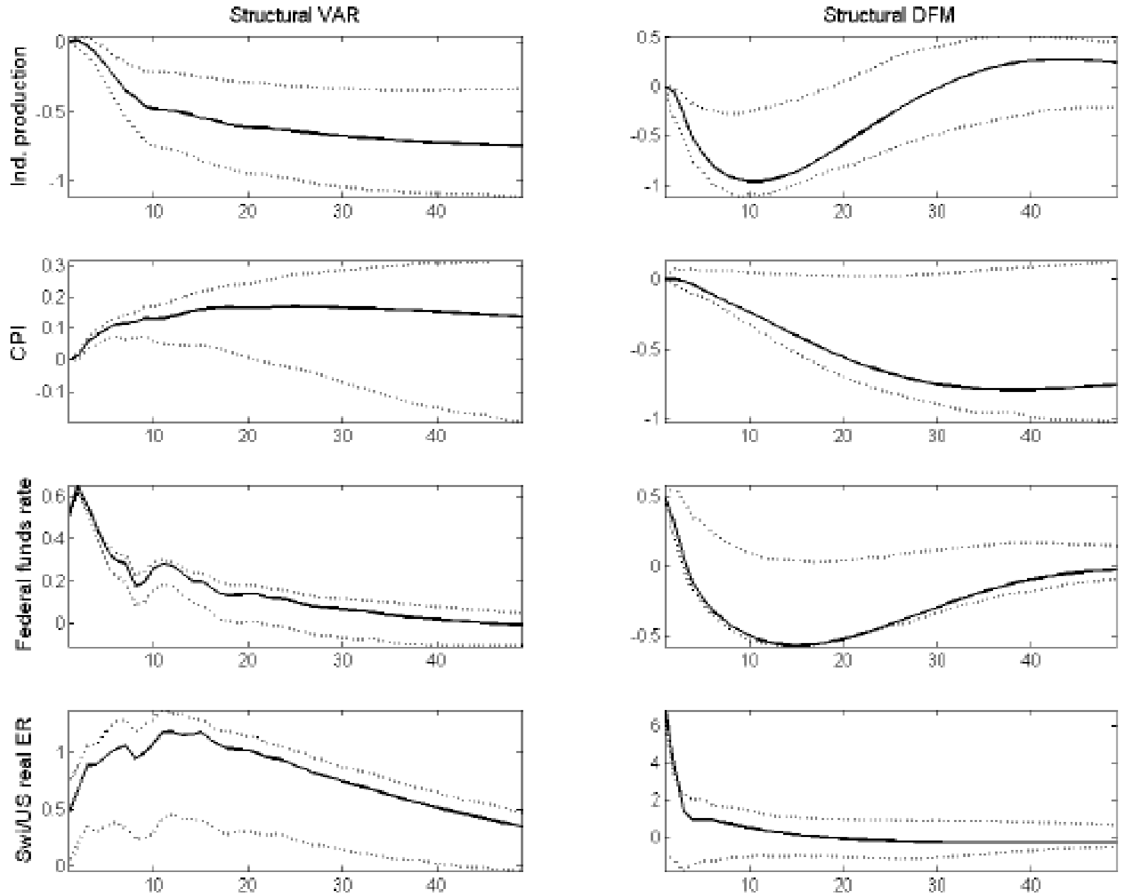
	k=0	k=6	k=12	k=48
Ind. production	0 (0)	0.0361 (0.0634)	0.1129 (0.1388)	0.3062 (0.1737)
CPI	0 (0)	0.0483 (0.0300)	0.0461 (0.0364)	0.0170 (0.0358)
Federal funds rate	0.9209 (0.0205)	0.5435 (0.0182)	0.3996 (0.0208)	0.1854 (0.0322)
Swi/US real ER	0.0275 0.0313	0.0685 (0.0420)	0.0923 (0.0497)	0.1434 (0.0607)

(*) Months after the shocks on the columns.

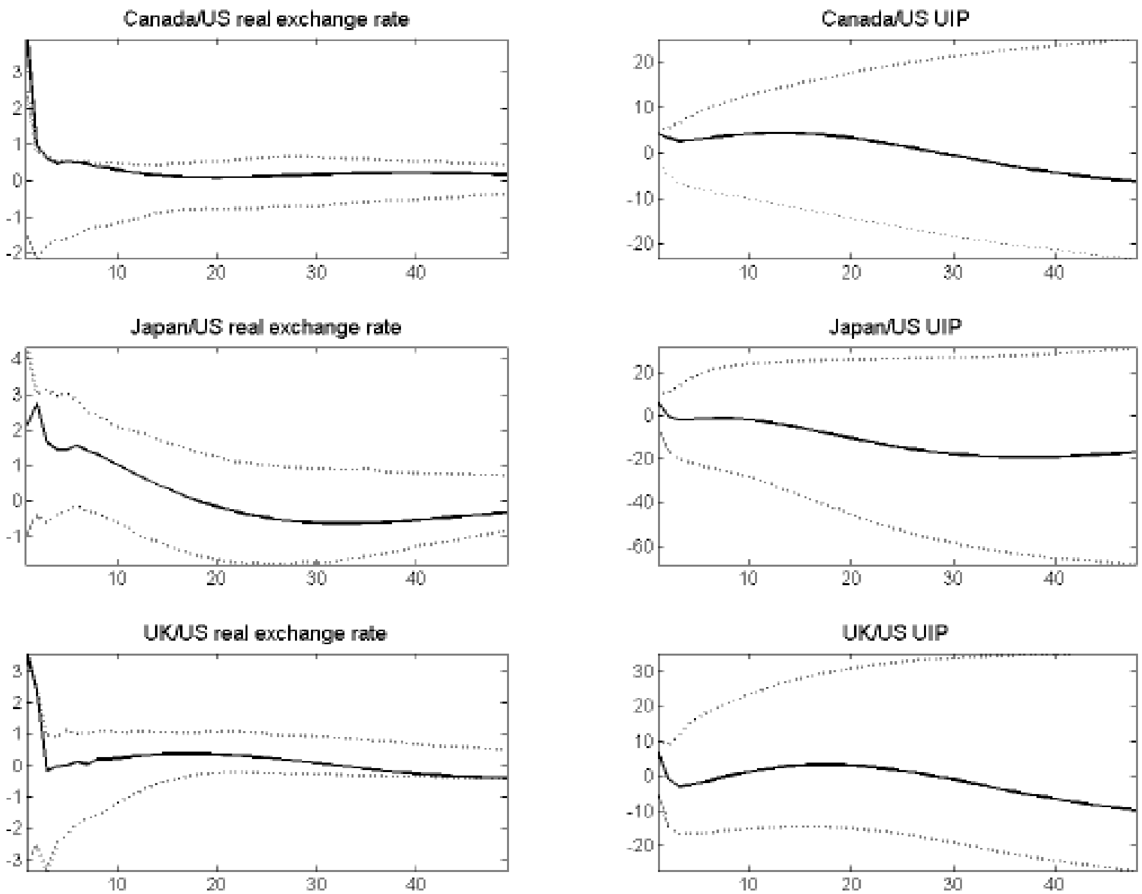
Table 2: Variance decomposition SDFM (*)

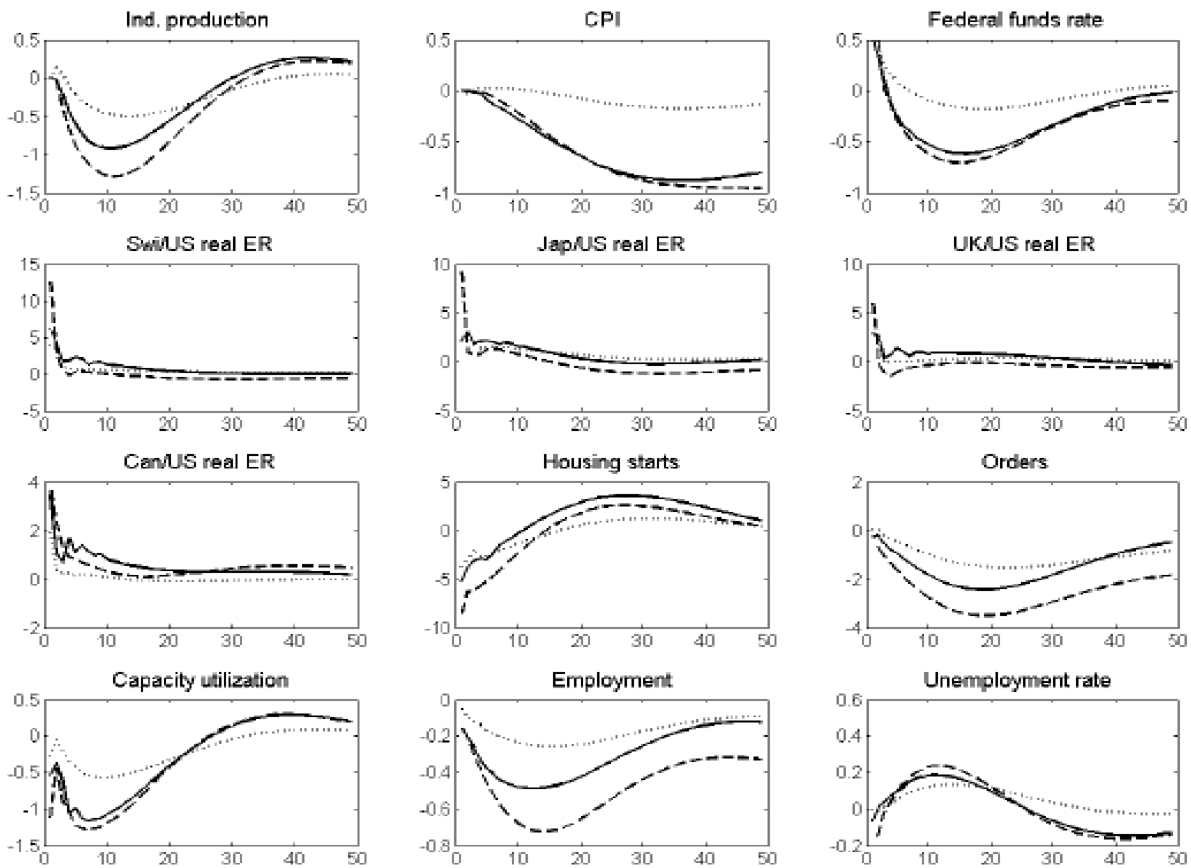
	k=0	k=6	k=12	k=48
Ind. production	0 (0)	0.0657 (0.0465)	0.1299 (0.0674)	0.1346 (0.0710)
CPI	0 (0)	0.0057 (0.0243)	0.0333 (0.0608)	0.1634 (0.1679)
Federal funds rate	0.5345 (0.2335)	0.1463 (0.2036)	0.1986 (0.1676)	0.2989 (0.1575)
Swi/US real ER	0.5227 (0.2704)	0.4330 (0.2123)	0.4041 (0.2028)	0.3836 (0.1666)
Can/US real ER	0.7541 (0.2605)	0.3474 (0.1825)	0.2523 (0.1794)	0.1643 (0.1580)
Jap/US real ER	0.1885 (0.2897)	0.2371 (0.2101)	0.2092 (0.2013)	0.1746 (0.1765)
UK/US real ER	0.2313 (0.2165)	0.1463 (0.1841)	0.1227 (0.1795)	0.1200 (0.1543)

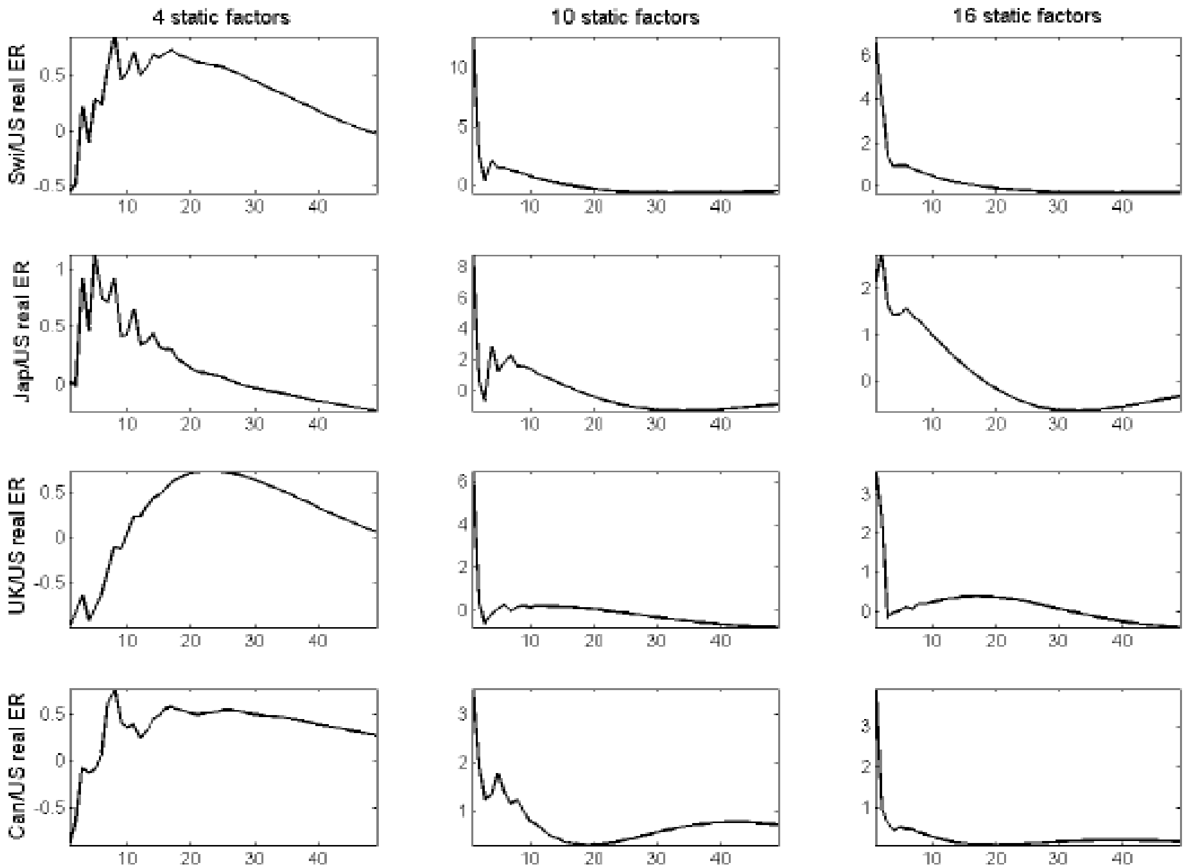
(*) Months after the shocks on the columns.



Structural VAR
Structural DFM
Ind. production
CPI
Federal funds rate
Swi/US real ER


Canada/US real exchange rate
Canada/US UIP
Japan/US real exchange rate
Japan/US UIP
UK/US real exchange rate
UK/US UIP
10
20
30
40


Ind. production
CPI
Federal funds rate
Swi/US real ER
Jap/US real ER
UK/US real ER
Can/US real ER
Housing starts
Orders
Capacity utilization
Employment
Unemployment rate
0 10 20 30 40 50


4 static factors
10 static factors
16 static factors
Swi/US real ER
Jap/US real ER
UK/US real ER
Can/US real ER

FAVAR

- Very similar to factors models.
- Two main differences:

1. Same number of dynamic and static factors $q = r$.

2. Possibility of including observed factors in the VAR for the factors.

FAVAR and Monetary policy shocks - BBE

Bernanke Boivin and Eliasz (2002) use a FAVAR model to study the effects of a monetary policy shock.

X_t consists of a panel of 120 monthly macroeconomic time series. The data span from January 1959 through August 2001.

The federal funds rate is the only observable factor, Y_t.

The model is estimated with 13 lags.

3 and 5 unobservable factors are used.

Identification of the monetary policy shock is achieved using a recursive ordering with the federal funds rate is ordered last in the vector of factors and and interpreting the last shock as the monetary policy one.

The ordering implies that all the other unobserved factors do not respond contemporaneously to the mp shock.

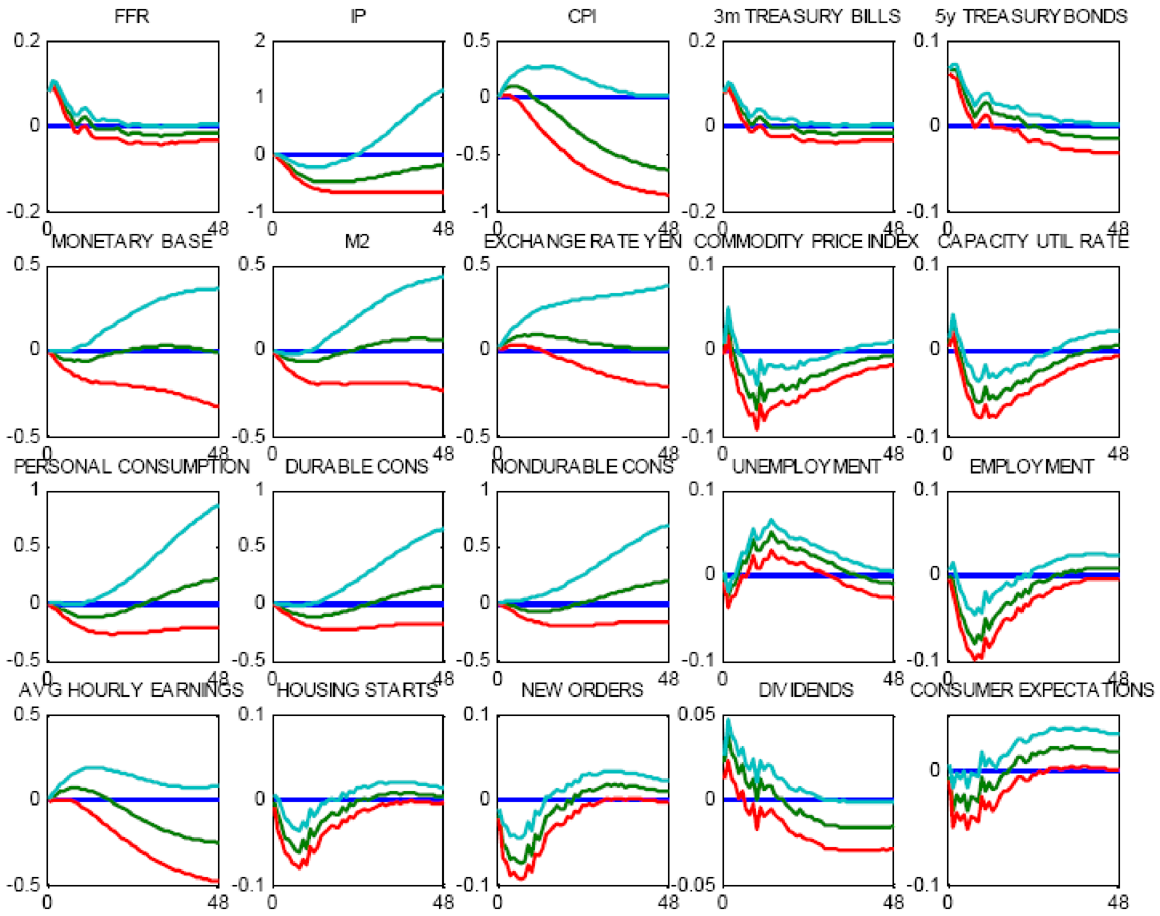


Figure 1. Impulse responses generated from FAVAR with 3 factors and FFR estimated by principal components with 2 step bootstrap.

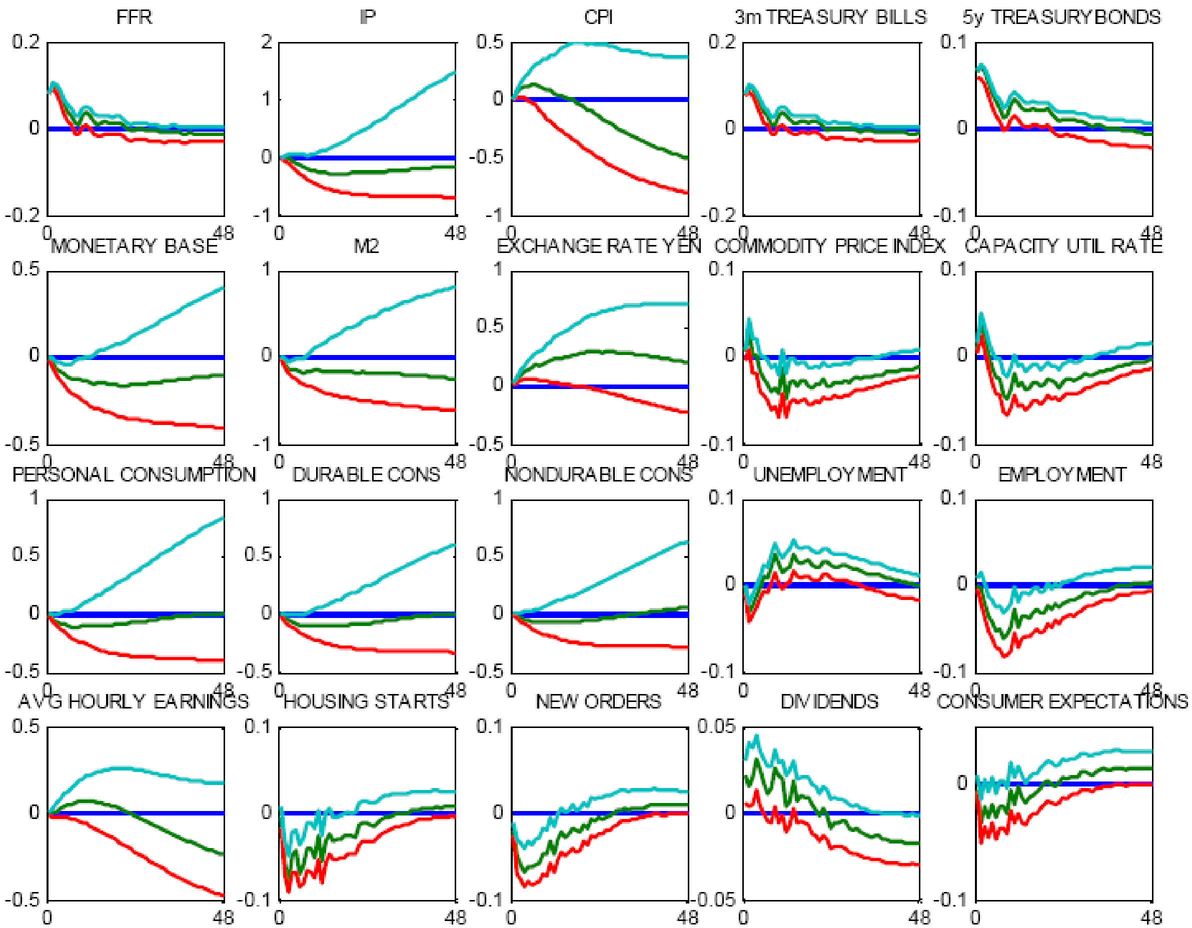


Figure 3. Impulse responses generated from FAVAR with 5 factors and FFR estimated by principal components with 2 step bootstrap.

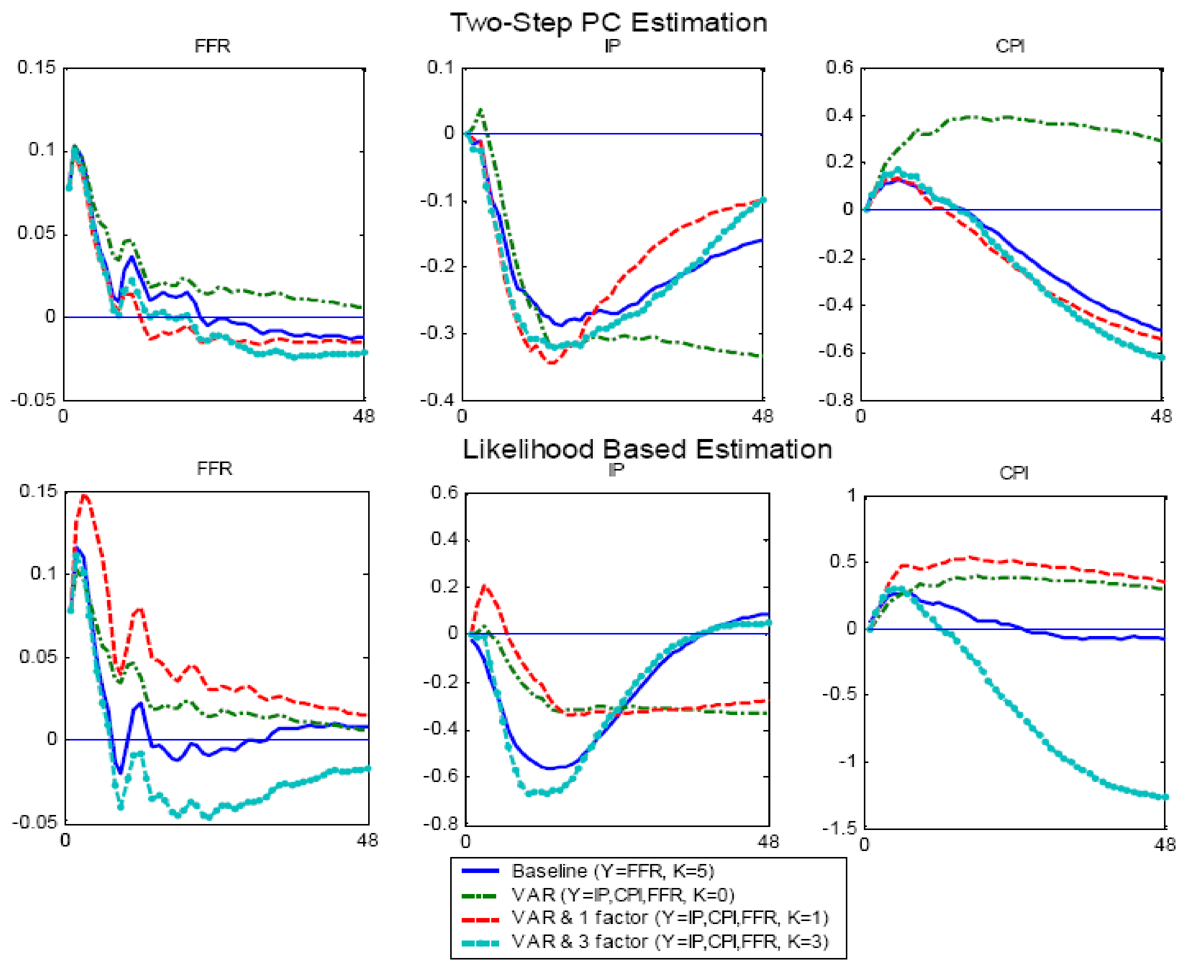


Figure 5. VAR – FAVAR comparison. The top panel displays estimated responses for the two-step principal component estimation and the bottom panel for the likelihood based estimation.

Table 1. Contribution of the policy shock to variance of the common component

Variables	**Variance Decomposition**	$\mathbf{R^2}$
Federal funds rate	0.4538	*1.0000
Industrial production	0.0763	0.7074
Consumer price index	0.0441	0.8699
3-month treasury bill	0.4440	0.9751
5-year bond	0.4354	0.9250
Monetary Base	0.0500	0.1039
M2	0.1035	0.0518
Exchange rate (Yen/$)	0.2816	0.0252
Commodity price Index	0.0750	0.6518
Capacity utilization	0.1328	0.7533
Personal consumption	0.0535	0.1076
Durable consumption	0.0850	0.0616
Non-durable cons.	0.0327	0.0621
Unemployment	0.1263	0.8168
Employment	0.0934	0.7073
Aver. Hourly Earnings	0.0965	0.0721
Housing Starts	0.0816	0.3872
New Orders	0.1291	0.6236
S&P dividend yield	0.1136	0.5486
Consumer Expectations	0.0514	0.7005

The column entitled "Variance Decomposition" reports the fraction of the variance of the forecast error of the common component, at the 60-month horizon, explained by the policy shock. "R^2" refers to the fraction of the variance of the variable explained by the common factors, $(\hat{F}_t, Y_t)$. See text for details.

*This is by construction.

No news in business cycle, Forni, Gambetti and Sala (2012)

- Beaudry and Portier (2006 AER) find news shocks are important for economic fluctuations. Output, investment, consumtpion and hours positively comove and the shocks explain a large fraction of the their variance.

- Use a VECM for TFPand stock prices.

- Standard model do not replicate the empirical finding since because of consumption comove negatively with investment and hours.

- Big effort in building models where news shocks generate business cycles.

– Motivation: news shocks can give raise to nonfundamentalness (recall example at the beginning).

– VAR models like Beaudry and Portier AER can have an hard time in estimating news shocks.

– Here:

  * Test whether the news shock is fundamental for TFP and stock prices, i.e. are fundamental for the variable in BP

  * Estimate the shocks using a FAVAR model.

- Data: 116 US quarterly series, covering the period 1959-I to 2007-IV.

- The news shock is identified imposing zero effect on TFP on impact and maximizing the effect of the TFP in the long run. Idea: is a shock that takes time to have effect on TFP but have important effects in the long run.

- Main results:

  1. News shocks are nonfundamental for the BP variables

  2. Explain little of the business cycle.

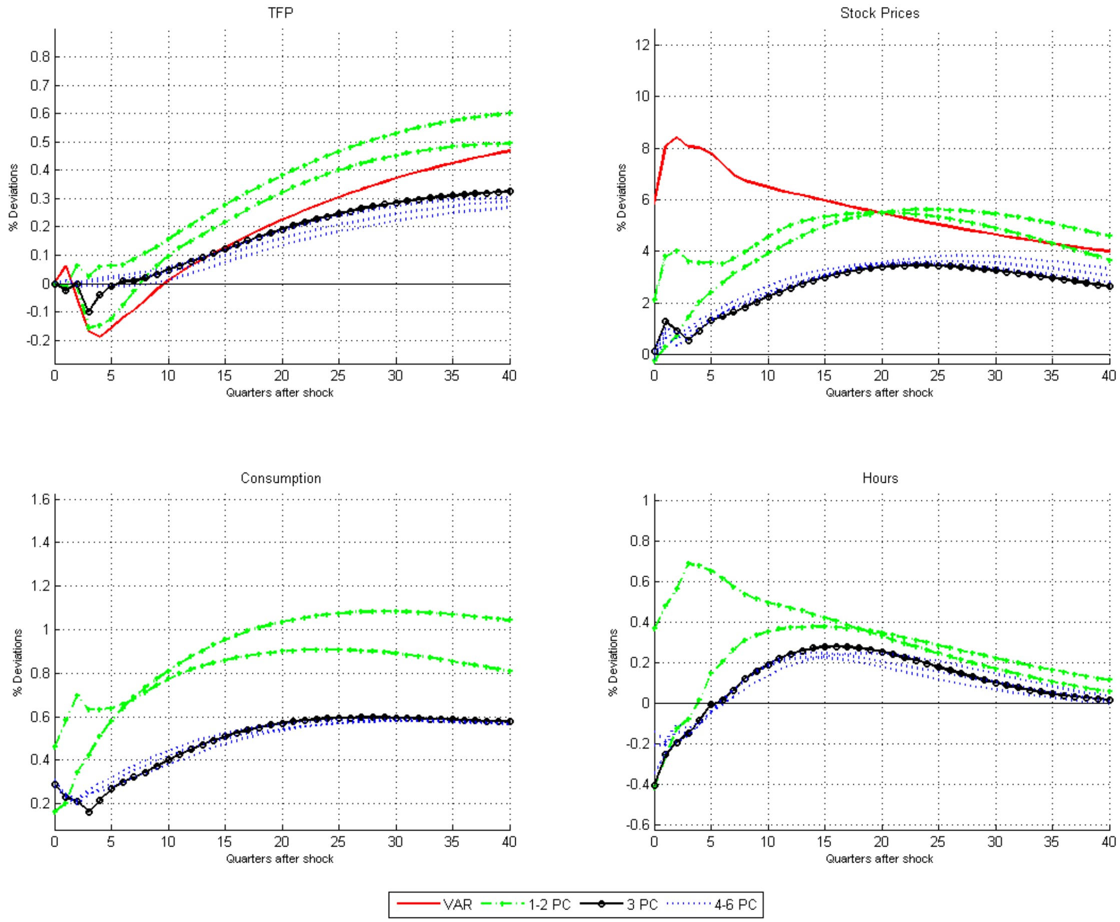
  3. Impulse response functions in line with standard models.

					lags
	2 variables ($S1$, $S2$: Beaudry and Portier, 2006)				lags
$S1$	TFP adj. (93)	Stock Prices (96)			5
$S2$	TFP (94)	Stock Prices (96)			5
	4 variables ($S3$, $S4$: Beaudry and Portier, 2006 - $S5$: Barsky and Sims, 2011)				
$S3$	TFP adj. (93)	Stock Prices (96)	Consumption (11)	Hours Worked (26)	5
$S4$	TFP (94)	Stock Prices (96)	Consumption (11)	Hours Worked (26)	5
$S5$	TFP adj. (93)	Output (5)	Consumption (11)	Hours Worked (26)	3
	7 variables ($S6$: Barsky and Sims, 2011)				
$S6$	TFP adj. (93)	Output (5)	Consumption (11)	Hours Worked (26)	
	Stock Prices (96)	Confidence (104)	Inflation (71)		3

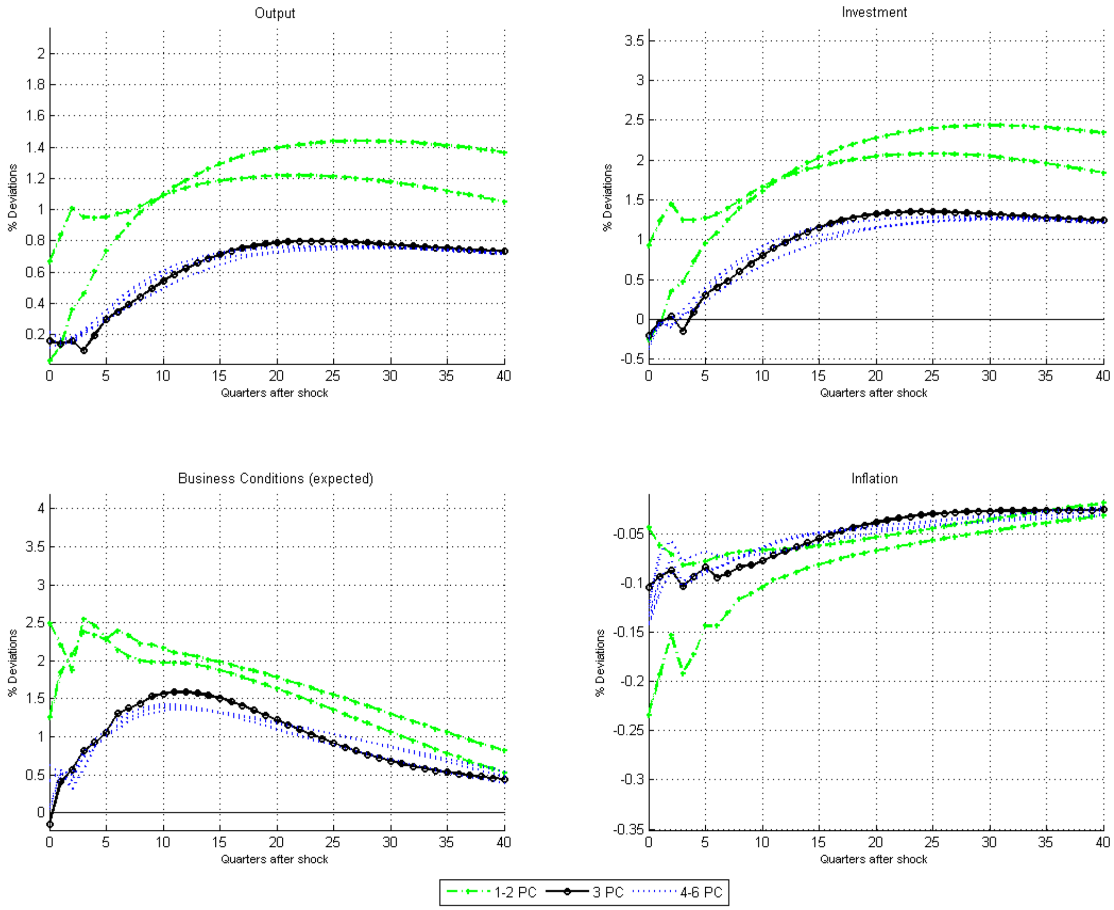
VAR specifications used to identify news shocks. Numbers in brackets correspond to the series in the data appendix.

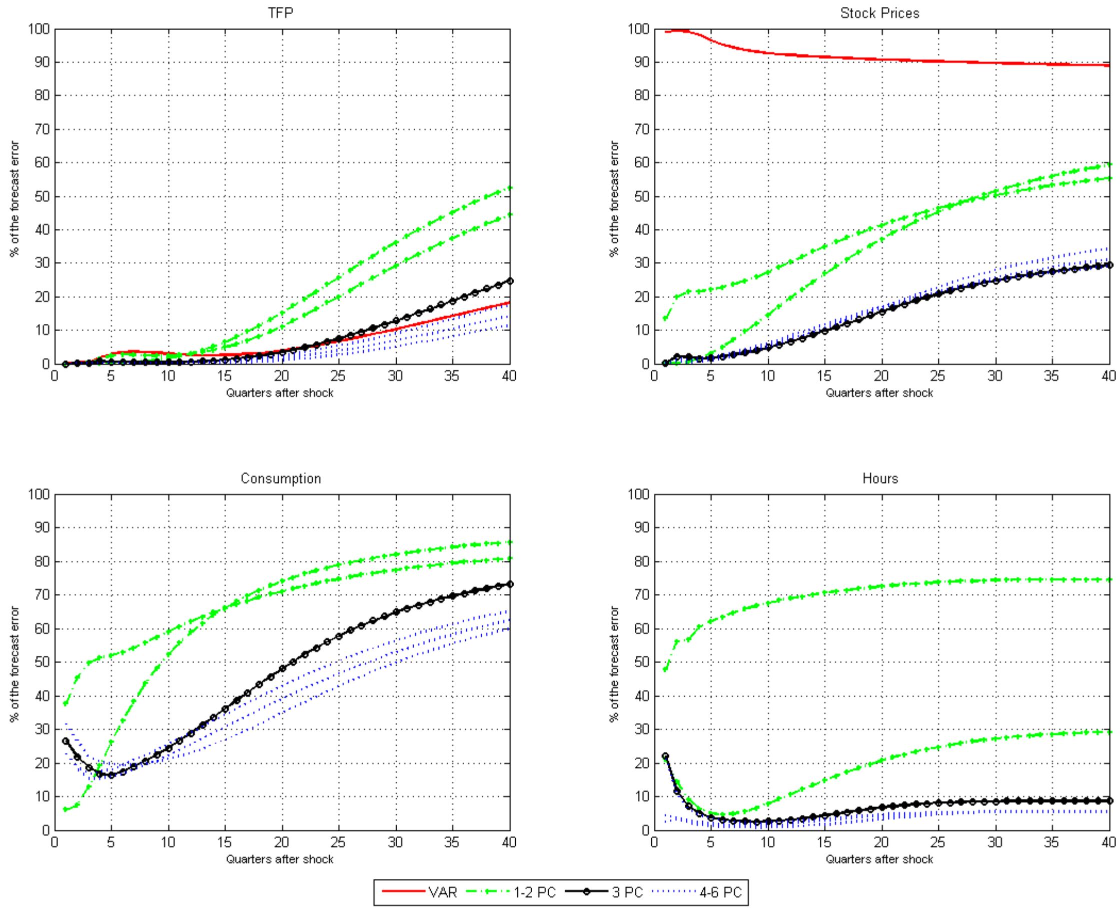
		Principal components (from 1 to j)									
spec	lags	1	2	3	4	5	6	7	8	9	10
S1	2	0.17	0.36	0.11	0.06	0.10	0.10	0.07	0.11	0.15	0.07
	4	0.37	0.19	0.04	0.06	0.10	0.03	0.06	0.09	0.12	0.02
S2	2	0.39	0.73	0.11	0.04	0.08	0.06	0.04	0.04	0.06	0.02
	4	0.56	0.61	0.09	0.06	0.12	0.04	0.08	0.11	0.13	0.04
S3	2	0.05	0.05	0.04	0.08	0.14	0.09	0.12	0.16	0.19	0.26
	4	0.20	0.09	0.07	0.20	0.12	0.08	0.08	0.10	0.12	0.21
S4	2	0.34	0.05	0.10	0.10	0.16	0.12	0.17	0.17	0.17	0.22
	4	0.48	0.03	0.08	0.13	0.03	0.03	0.08	0.08	0.06	0.08
S5	2	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
S6	2	0.79	0.49	0.76	0.72	0.85	0.86	0.84	0.91	0.92	0.85
	4	0.43	0.24	0.53	0.52	0.49	0.72	0.58	0.66	0.72	0.72

Results of the fundamentalness test described in Section **??**. Each entry of the table reports the p-value of the F-test in a regression of the news shock estimated using specifications $S1$ to $S6$ on 2 and 4 lags of the first differences of the first j principal components, $j = 1, \dots, 10$. The news shock is identified as the shock that does not move TFP on impact and (for specifications from $S3$ to $S6$) has maximal effect on TFP at horizon 60.

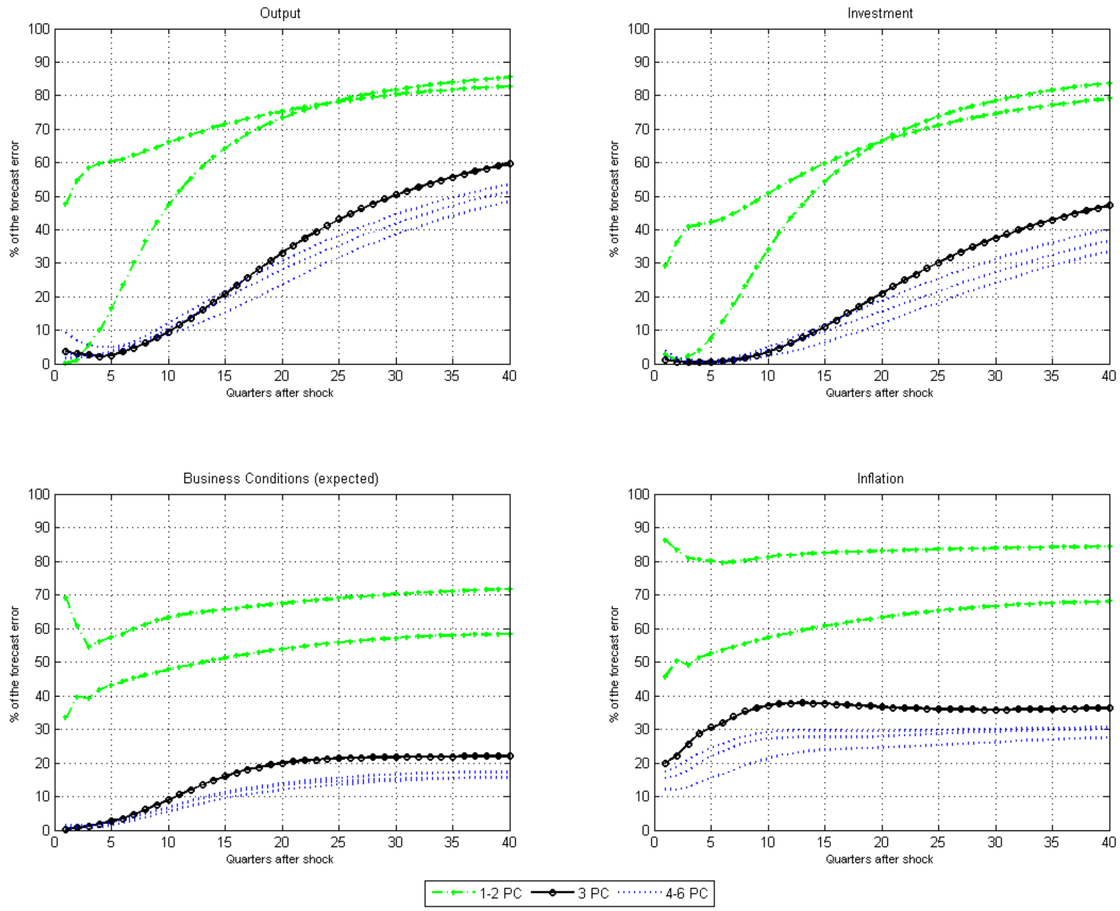


TFP
Stock Prices
Consumption
Hours
% Deviations
Quarters after shock
VAR
1-2 PC
3 PC
4-6 PC

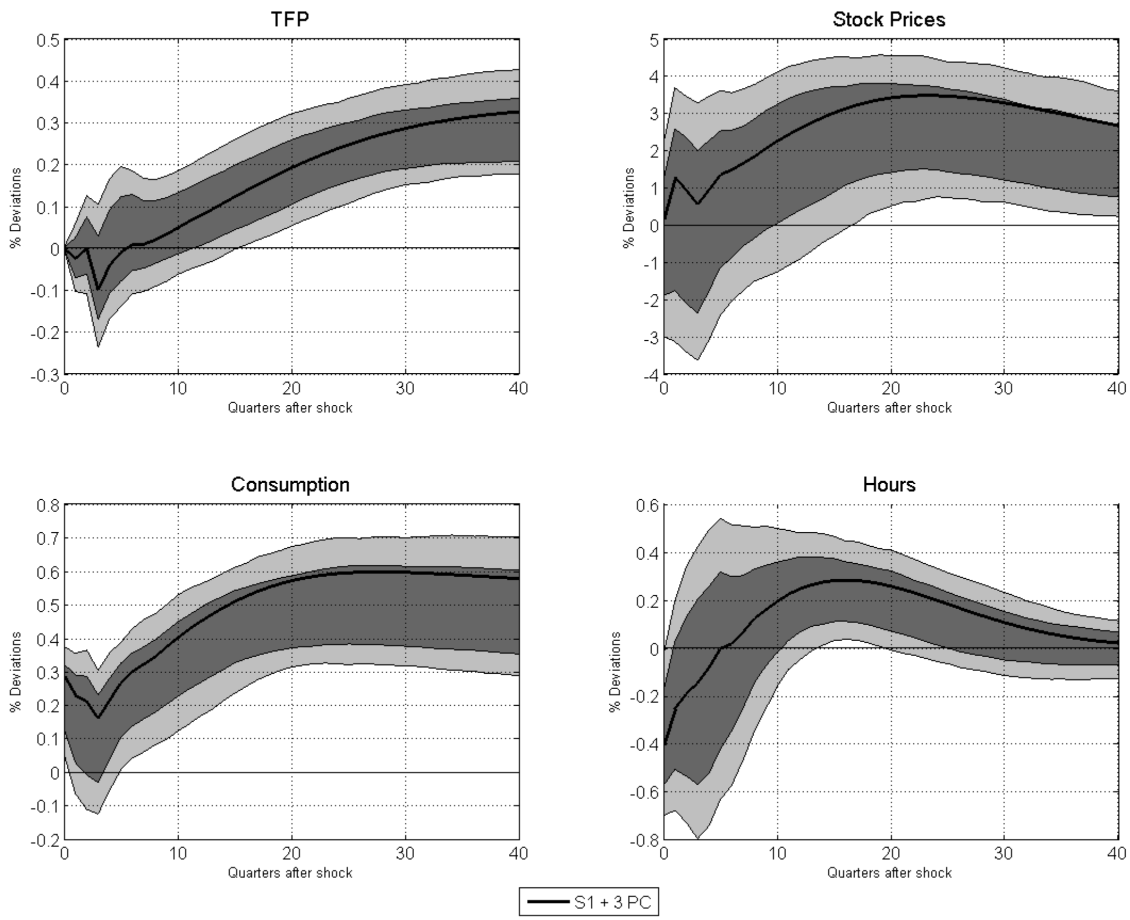

Output
Investment
Business Conditions (expected)
Inflation
% Deviations
Quarters after shock
1-2 PC
3 PC
4-6 PC



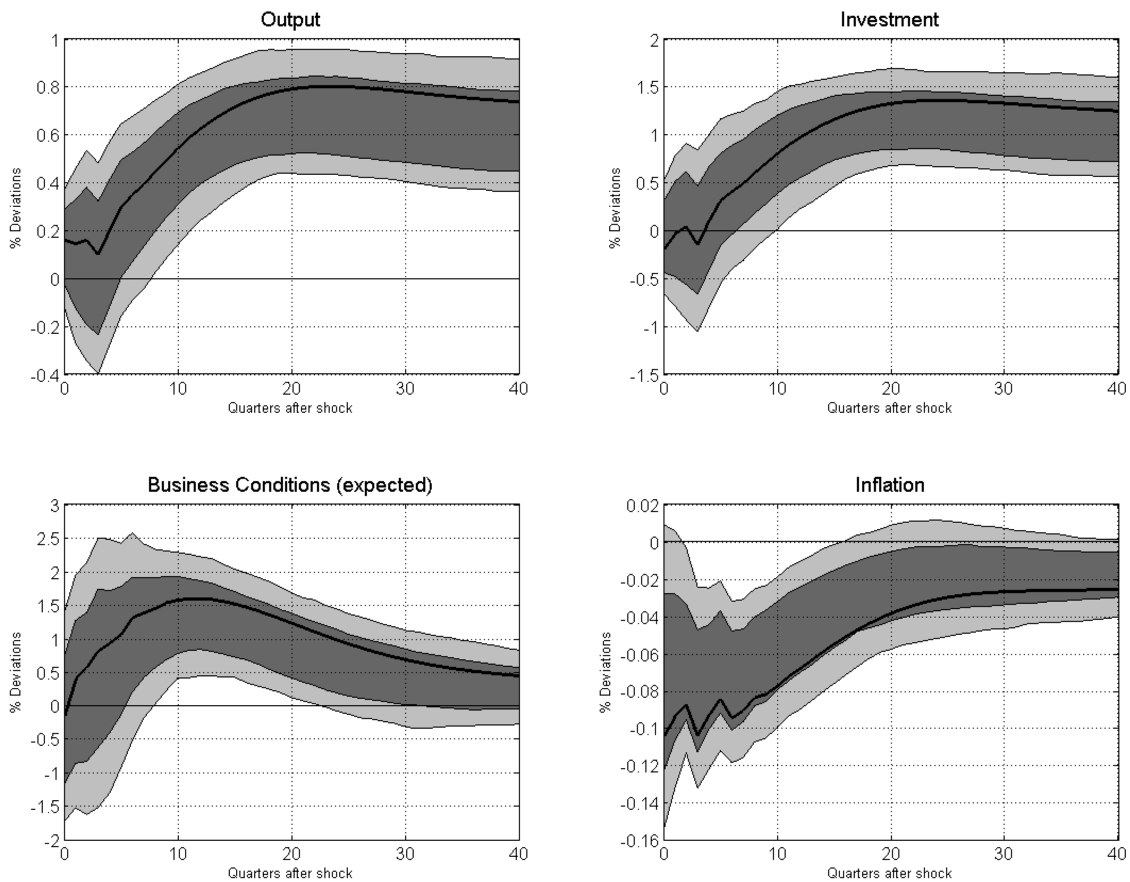
TFP
Stock Prices
Consumption
Hours
% of the forecast error
Quarters after shock
VAR
1-2 PC
3 PC
4-6 PC



Output
Investment
Business Conditions (expected)
Inflation
% of the forecast error
Quarters after shock
1-2 PC
3 PC
4-6 PC



TFP
Stock Prices
Consumption
Hours
% Deviations
Quarters after shock
S1 + 3 PC


Output
Investment
Business Conditions (expected)
Inflation
% Deviations
Quarters after shock

FAVAR and Disaggregated prices response - BGM

- Boivin Giannoni and Mihov in "Sticky Prices and Monetary Policy: Evidence from Disaggregated U.S. Data" study the response of disaggregate prices to monetary policy.

- Important because assessing of the speed of price adjustment is crucial to discriminate among competing models of business cycle (flexible vs sticky prices models).

- They asses the relative importance in disaggregated U.S. consumer and producer of aggregate macroeconomic factors and sectorial conditions.

Main results

1. Most of the fluctuations in disaggregated inflations are due to sector-specific factors. (only 15% due to macroeconomic factors).

2. Very little variations of most prices for several months after the shock. Very sluggish response to aggregate macroeconomic shocks such as monetary policy shocks.

3. The above can explain why, at the disaggregated level, individual prices are found to be adjusted relatively frequently, while in aggregate data estimates of the degree of price rigidity are much higher.

Implication. This might explain why models that assume considerable price stickiness have been successful.

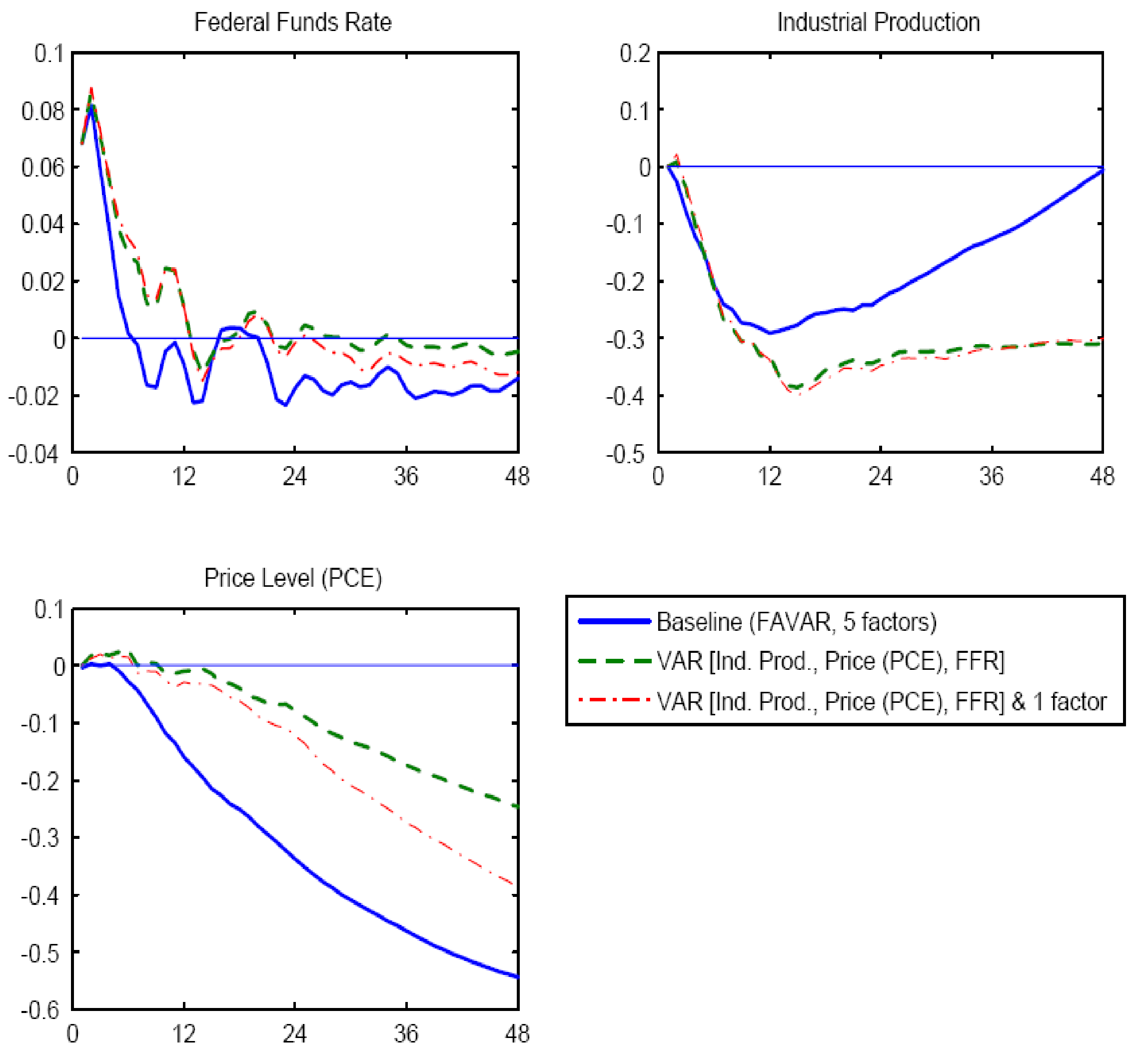


Figure 2a: Estimated impulse responses to an identified monetary policy shock

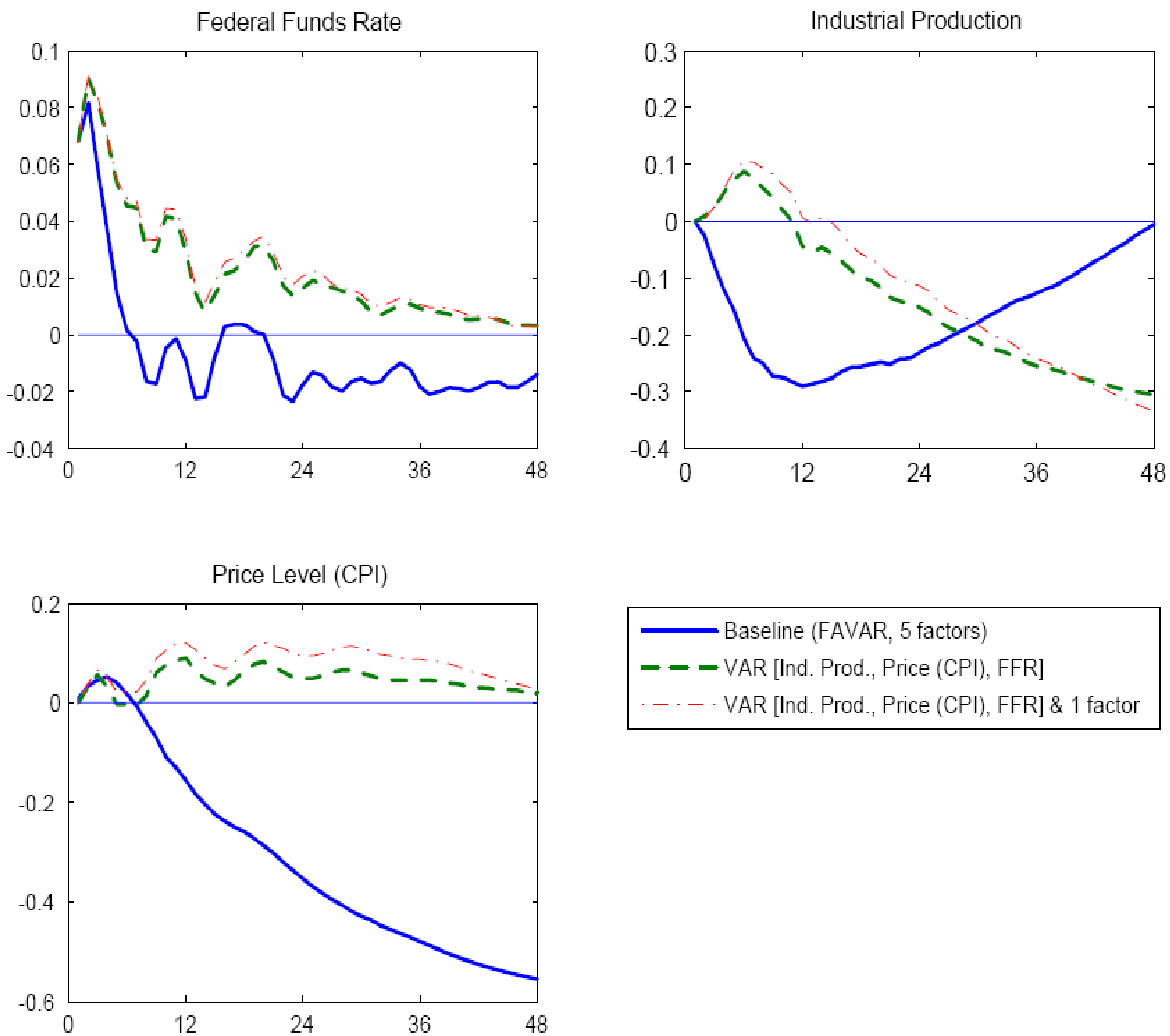


Figure 2b: Estimated impulse responses to an identified monetary policy shock (CPI)

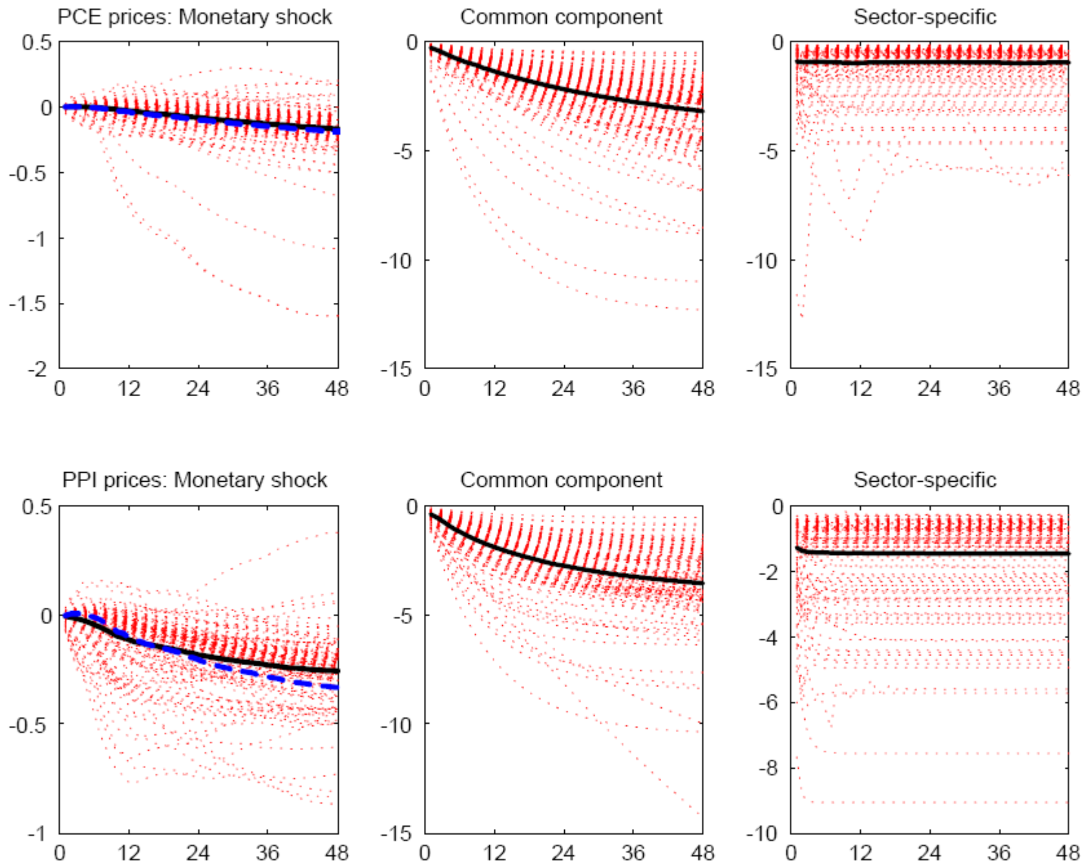


Figure 4a. Estimated impulse responses of (log) sectorial prices to an identified monetary policy shock (left panels), to a shock to the common component (middle panels), and to a sector-specific shock (right panels). Fat lines represent unweighted average responses. Fat dashed lines represent the response of the aggregate PCE and PPI (finished) price indices to a monetary policy shock.

Table 1: Volatility and persistence of inflation series

		Standard deviation				Persistence		
		Inflation	Common comp.	Sector-specific	R2	Inflation	Common comp.	Sector-specific
Aggregated series								
PCE	Total	0.24	0.21	0.11	0.77	0.90	0.95	0.13
	Durables	0.33	0.25	0.21	0.60	0.88	0.97	0.08
	Nondurables	0.42	0.30	0.30	0.50	0.44	0.91	0.22
	Services	0.24	0.19	0.14	0.63	0.91	0.98	0.01
Disaggregated series								
All	Average	1.15	0.33	1.08	0.15	0.29	0.91	-0.03
	Median	0.75	0.27	0.71	0.12	0.30	0.93	-0.02
	Minimum	0.23	0.08	0.13	0.01	-2.32	0.39	-1.83
	Maximum	11.67	1.85	11.59	0.68	0.96	0.99	0.87
	Std	1.14	0.22	1.13	0.12	0.39	0.06	0.33
PCE	Average	0.97	0.29	0.92	0.17	0.30	0.92	-0.05
	Average (weighted)	0.88	0.31	0.80	0.27	0.47	0.93	0.04
	Median	0.65	0.23	0.60	0.12	0.36	0.95	-0.02
	Minimum	0.23	0.08	0.13	0.01	-2.32	0.39	-1.83
	Maximum	11.67	1.85	11.59	0.68	0.96	0.99	0.87
	Std	1.10	0.23	1.09	0.15	0.44	0.07	0.37
PPI	Average	1.36	0.38	1.29	0.13	0.28	0.90	0.01
	Median	0.92	0.30	0.87	0.11	0.27	0.91	-0.01
	Minimum	0.35	0.08	0.29	0.01	-0.76	0.61	-0.93
	Maximum	7.73	1.15	7.66	0.43	0.91	0.98	0.63
	Std	1.15	0.21	1.15	0.08	0.31	0.06	0.27

Note: Weighted average of statistics for disaggregated PCE series is obtained using expenditure shares in year 2005 as weights.

7 Time-Varying Coefficients VAR

Motivation

- Economies and economic dynamics are evolving over-time.
- Parameters constancy probably not a too good idea.
- Better idea: allowing model dynamics to also vary over-time.
- Several ways to do it:
  - More or less smooth regime switches.
  - Continuously varying parameters
- Here we focus on the second.

The Model

Time-varying coefficients VAR (TVC-VAR) represent a generalization of VAR models in which the coefficients are allowed to change over time. Let Y_t be a vector of time series. We assume that Y_t satisfies

$$Y_t = A_{0,t} + A_{1,t}Y_{t-1} + ... + A_{p,t}Y_{t-p} + \varepsilon_t \tag{57}$$

where ε_t is a Gaussian white noise with zero mean and time-varying covariance matrix Σ_t.

Let $A_t = [A_{0,t}, A_{1,t}..., A_{p,t}]$, and $\theta_t = vec(A_t')$, where $vec(\cdot)$ is the column stacking operator. We postulate

$$\theta_t = \theta_{t-1} + \omega_t \tag{58}$$

where ω_t is a Gaussian white noise with zero mean and covariance Ω.

Let $\Sigma_t = F_t D_t F_t'$, where F_t is lower triangular, with ones on the main diagonal, and D_t a diagonal matrix. Let σ_t be the vector of the diagonal elements of $D_t^{1/2}$ and $\phi_{i,t}$, $i = 1, ..., n-1$ the column vector formed by the non-zero and non-one elements of the $(i+1)$-th row of F_t^{-1}. We assume that

$$\log \sigma_t = \log \sigma_{t-1} + \xi_t \tag{59}$$

$$\phi_{i,t} = \phi_{i,t-1} + \psi_{i,t} \tag{60}$$

where ξ_t and $\psi_{i,t}$ are Gaussian white noises with zero mean and covariance matrix Ξ and Ψ_i, respectively.

Let $\phi_t = [\phi'_{1,t}, \ldots, \phi'_{n-1,t}]$, $\psi_t = [\psi'_{1,t}, \ldots, \psi'_{n-1,t}]$, and Ψ be the covariance matrix of ψ_t. We assume that $\psi_{i,t}$ is independent of $\psi_{j,t}$, for $j \neq i$, and that ξ_t, ψ_t, ω_t, ε_t are mutually uncorrelated at all leads and lags.

Time-varying dynamics

A characteristic of the TVC-VAR is that impulse response functions and the second moments (variances and covariances) are time varying meaning that the effects and the contributions of a shock may change over time.

Example: TVC-VAR(1). consider the model

$$Y_t = A_t Y_{t-1} + \varepsilon_t \tag{61}$$

Ask: (i) what are the impulse response functions (effects of a shock)? (ii) and what is the variance of the process?

IRF measure the effects of a shock occurring today (time t) on future ($t+j$) time series. Substituting forward we obtain

$$\begin{aligned}
Y_{t+1} &= A_{t+1}A_tY_{t-1} + A_{t+1}\varepsilon_t + \varepsilon_{t+1} \\
Y_{t+2} &= A_{t+2}A_{t+1}A_tY_{t-1} + A_{t+2}A_{t+1}\varepsilon_t + A_{t+1}\varepsilon_{t+1} + \varepsilon_{t+2} \\
Y_{t+k} &= A_{t+k}...A_{t+1}A_tY_{t-1} + A_{t+k}...A_{t+2}A_{t+1}\varepsilon_t + ... + 1\varepsilon_{t+k}
\end{aligned}$$

so the collection $(A_{t+k}...A_{t+2}A_{t+1}), (A_{t+k-1}...A_{t+2}A_{t+1}), ..., (A_{t+2}A_{t+1}), A_{t+1}, I$ represent the impulse response functions of ε_t. Clearly these will be different for ε_{t-k}.

Note that if we ignore variations in future coefficients impulse response functions reduce to $A_t^k, A_t^{k-1}, ..., A_t, I$. That means that if we ignore future uncertainty in the parameters, the vector of time series, if stationarity condition are satisfied for every t, can be inverted at every t and

$$Y_t = C_t(L)\varepsilon_t$$

where $C_t(L) = I + C_{1t}L + C_{2t}L^2 + ...$ and $C_{kt} = A_t^k$.

Identifcation is similar to the fixed coefficients case. The difference is that both the Cholesky factor and the rotation matrix will be time varying S_t and H_t respectively. The structural model is

$$Y_t = C_t(L)S_tH_tw_t$$

Also the variance and the other second moments will be time varying. In fact the variance of the process at time t will be given by

$$Var_t(Y_t) = \sum_{j=0}^{\infty} A_t^j \Omega A_t'^j$$

Given that impulse response functions and the variance are time varying the contribution of each shock may change over time.

Estimation

• The easiest way to estimate the model is by using Bayesian MCMC methods, specifically the Gibbs sampler.

• Objective: we want to draws of the coefficients from the posterior distribution.

• The posterior distribution is unknown. What is known are the conditional posteriors so we can draw from

1. $p(\sigma^T|x^T, \theta^T, \phi^T, \Omega, \Xi, \Psi)$
2. $p(\phi^T|x^T, \theta^T, \sigma^T, \Omega, \Xi, \Psi)$
3. $p(\theta^T|x^T, \sigma^T, \phi^T, \Omega, \Xi, \Psi)$
4. $p(\Omega|x^T, \theta^T, \sigma^T, \phi^T, \Xi, \Psi)$
5. $p(\Xi|x^T, \theta^T, \sigma^T, \phi^T, \Omega, \Psi)$
6. $p(\Psi|x^T, \theta^T, \sigma^T, \phi^T, \Omega, \Xi)$

• After a large number N the draws obtained are draws from the joint posterior.

• The objects of interests (IRF, variance decomposition, etc.) can be computed from the draws obtained.

Application 1: Cogley and Sargent (2002) on unemployment-inflation dynamics)

• Very important paper. The first paper using a version of the model seen above.

• Aim: to provide evidence about the evolution of measures of the persistence of inflation, prospective long-horizon forecasts (means) of inflation and unemployment, statistics for a version of a Taylor rule.

• VAR for inflation, unemployment and the real interest rate.

• Main results: long-run mean, persistence and variance of inflation have changed. The Taylor principle was violated before Volcker (pre-1980). Monetary policy too loose.

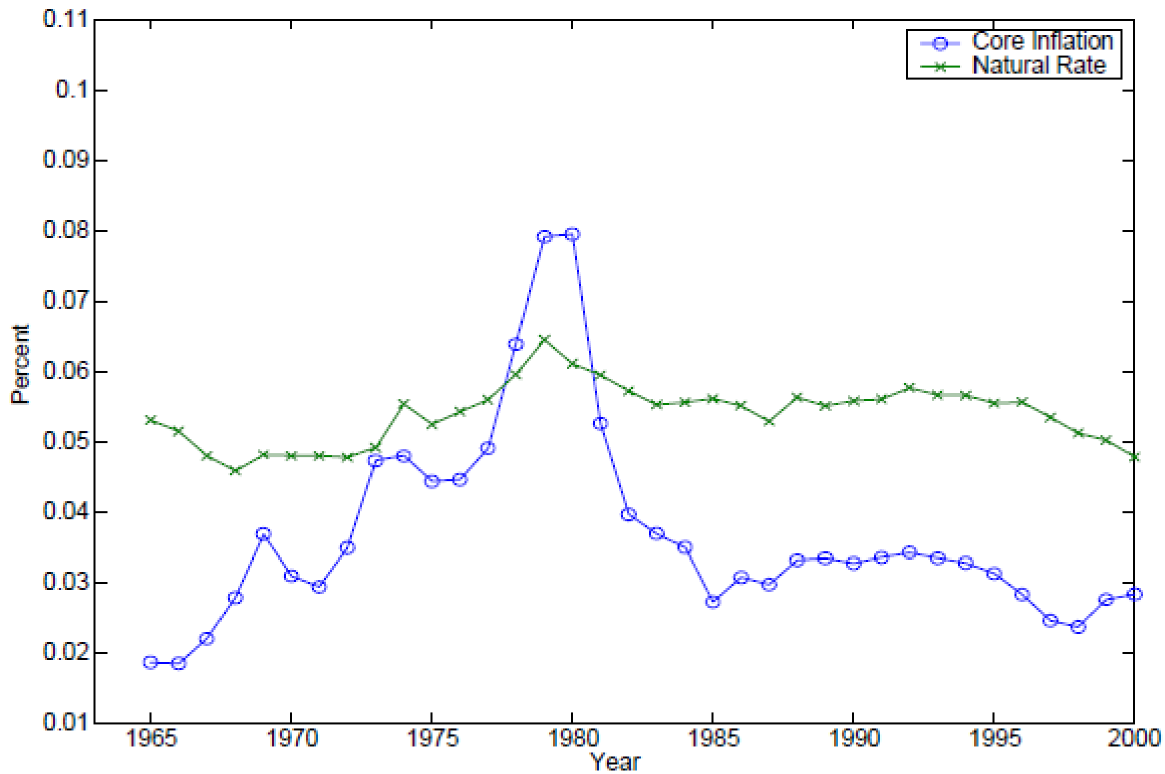


Figure 3.1: Core Inflation and the Natural Rate of Unemployment

T. Cogley and T.J. Sargent, (2002). "Evolving Post-World War II U.S. Inflation Dynamics," NBER Macroeconomics Annual 2001, Volume 16, pages 331-388.

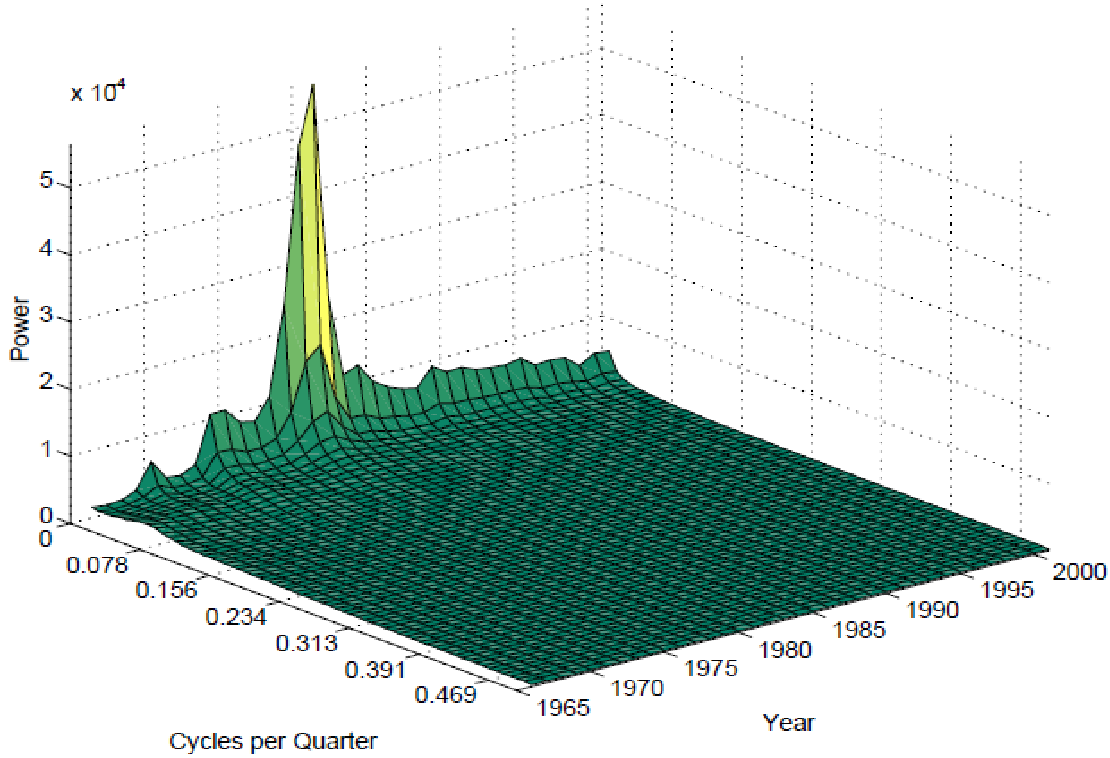


Figure 3.5: Median Posterior Spectrum for Inflation. Power is measured in basis points, the units of measurement for the variance of inflation.

T. Cogley and T.J. Sargent, (2002). "Evolving Post-World War II U.S. Inflation Dynamics," NBER Macroeconomics Annual 2001, Volume 16, pages 331-388.

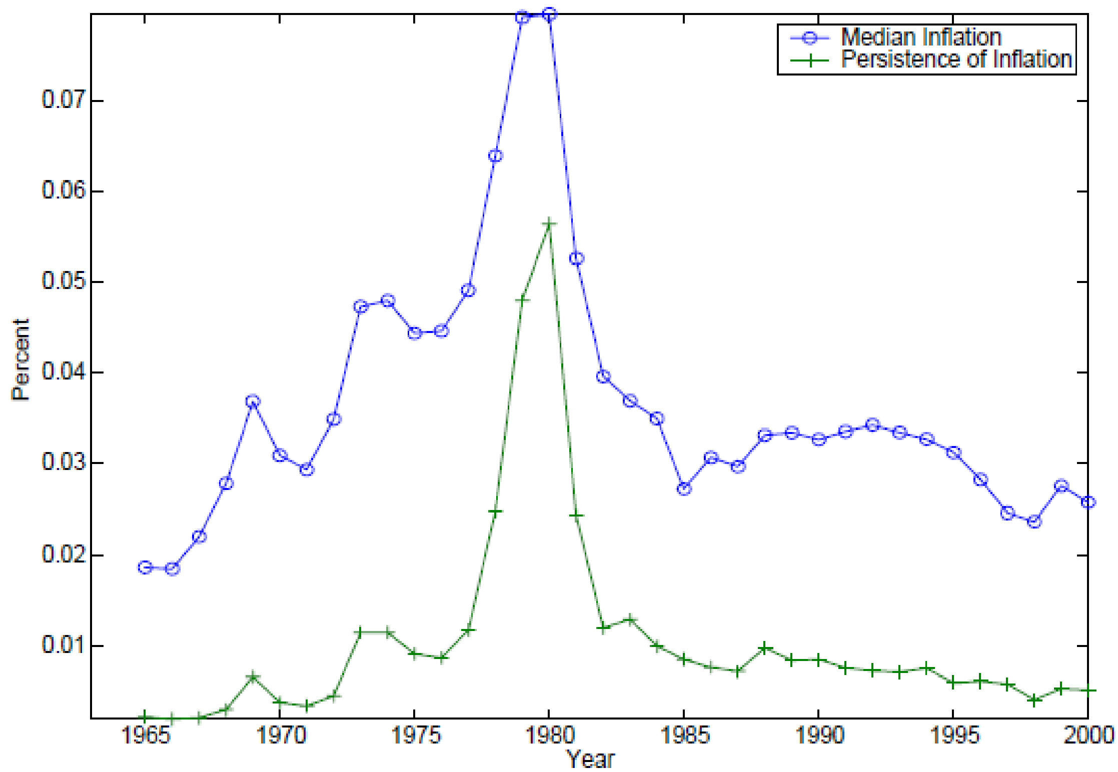


Figure 3.10: Core Inflation and Inflation Persistence

T. Cogley and T.J. Sargent, (2002). "Evolving Post-World War II U.S. Inflation Dynamics," NBER Macroeconomics Annual 2001, Volume 16, pages 331-388.

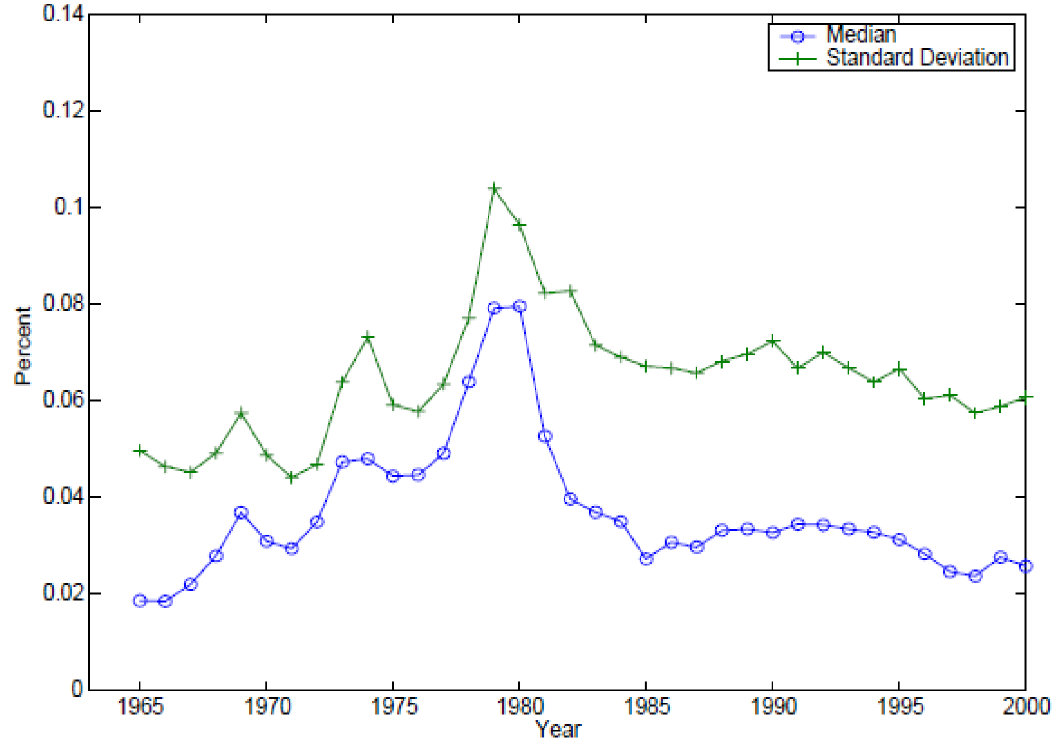


Figure 3.11: Core Inflation and the Standard Deviation of Inflation, 30 Years Ahead

T. Cogley and T.J. Sargent, (2002). "Evolving Post-World War II U.S. Inflation Dynamics," NBER Macroeconomics Annual 2001, Volume 16, pages 331-388.

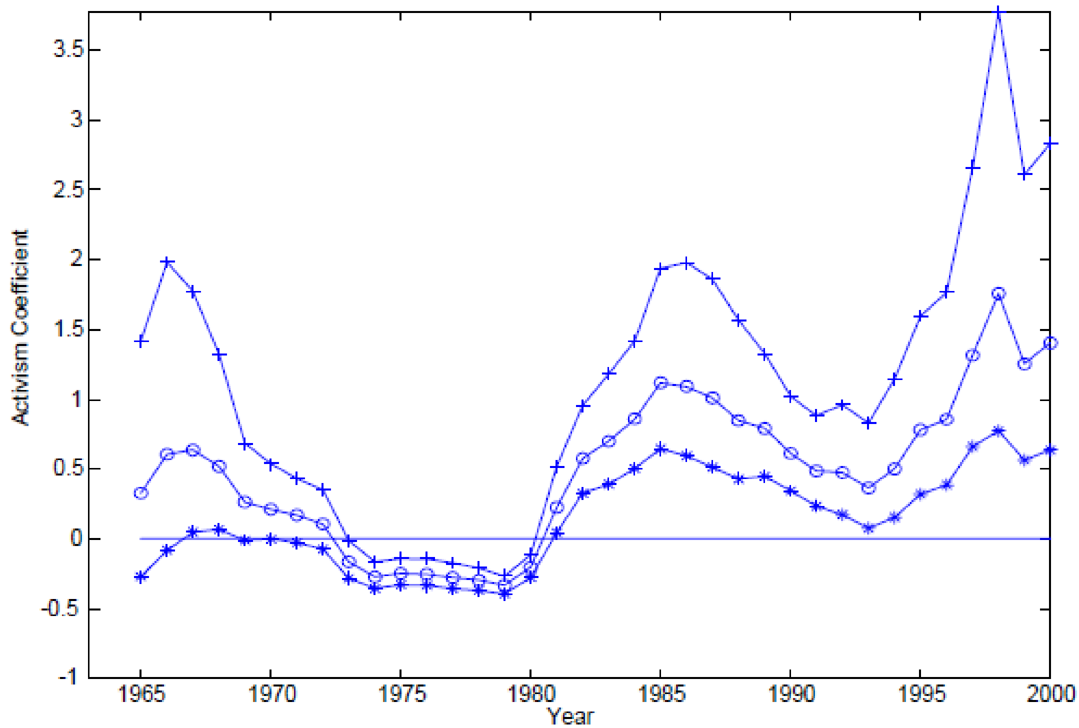


Figure 3.12: Posterior Median and Interquartile Range for the Activism Coefficient

T. Cogley and T.J. Sargent, (2002). "Evolving Post-World War II U.S. Inflation Dynamics," NBER Macroeconomics Annual 2001, Volume 16, pages 331-388.

Application 2: Primiceri (2005, ReStud) on monetary policy

- Very important paper: the first paper adding stochastic volatility.
- The paper studies changes in the monetary policy in the US over the postwar period.
- VAR for inflation, unemployment and the real interest rate.
- Main results:
  - systematic responses of the interest rate to inflation and unemployment exhibit a trend toward a more aggressive behavior,
  - this has had a negligible effect on the rest of the economy

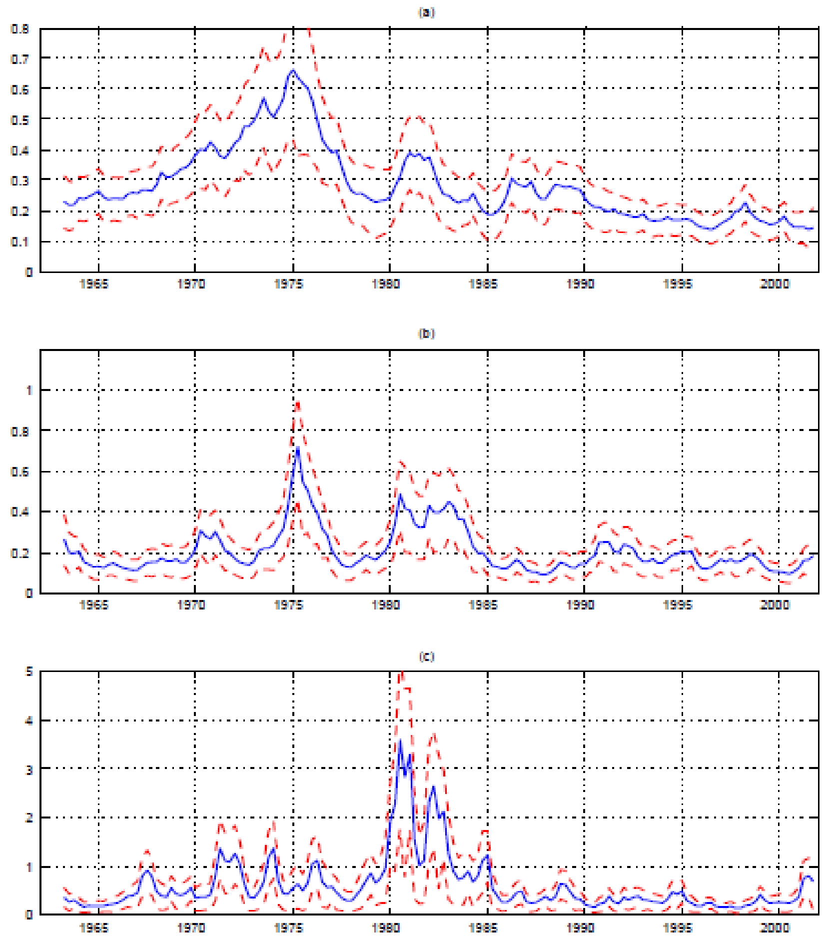


Figure 1: Posterior mean, 16th and 84th percentiles of the standard deviation of (a) residuals of the inflation equation, (b) residuals of the unemployment equation and (c) residuals of the interest rate equation or monetary policy shocks.

Source: G. Primiceri "Time Varying Structural Vector Autoregressions and Monetary Policy", The Review of Economic Studies, 72, July 2005, pp. 821-852

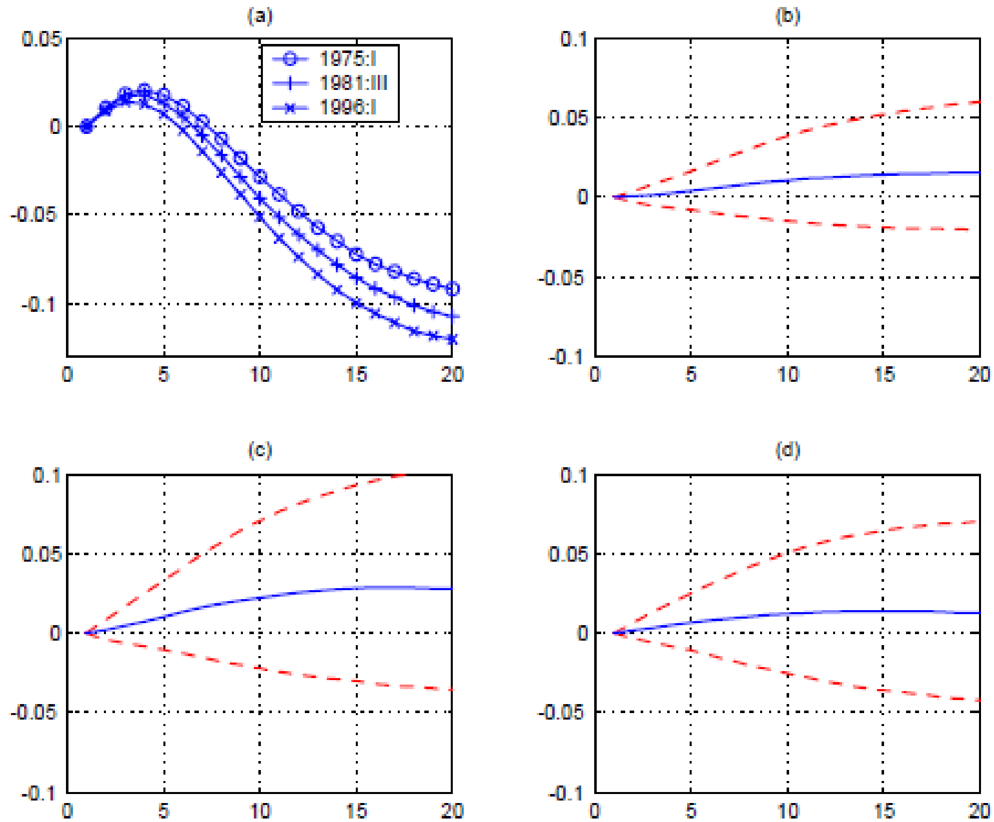


Figure 2: (a) impulse responses of inflation to monetary policy shocks in 1975:I, 1981:III and 1996:I, (b) difference between the responses in 1975:I and 1981:III with 16th and 84th percentiles, (c) difference between the responses in 1975:I and 1996:I with 16th and 84th percentiles, (d) difference between the responses in 1981:III and 1996:I with 16th and 84th percentiles.

Source: G. Primiceri "Time Varying Structural Vector Autoregressions and Monetary Policy", The Review of Economic Studies, 72, July 2005, pp. 821-852

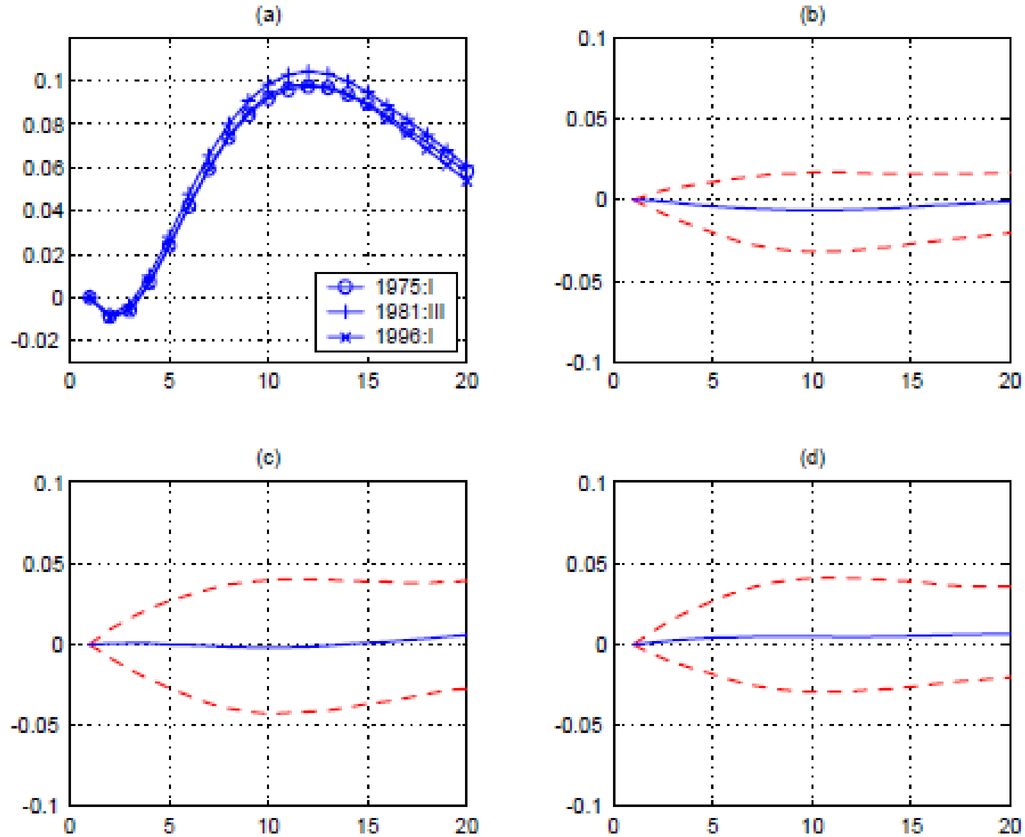


Figure 3: (a) impulse responses of unemployment to monetary policy shocks in 1975:I, 1981:III and 1996:I, (b) difference between the responses in 1975:I and 1981:III with 16th and 84th percentiles, (c) difference between the responses in 1975:I and 1996:I with 16th and 84th percentiles, (d) difference between the responses in 1981:III and 1996:I with 16th and 84th percentiles.

Source: G. Primiceri "Time Varying Structural Vector Autoregressions and Monetary Policy", The Review of Economic Studies, 72, July 2005, pp. 821-852

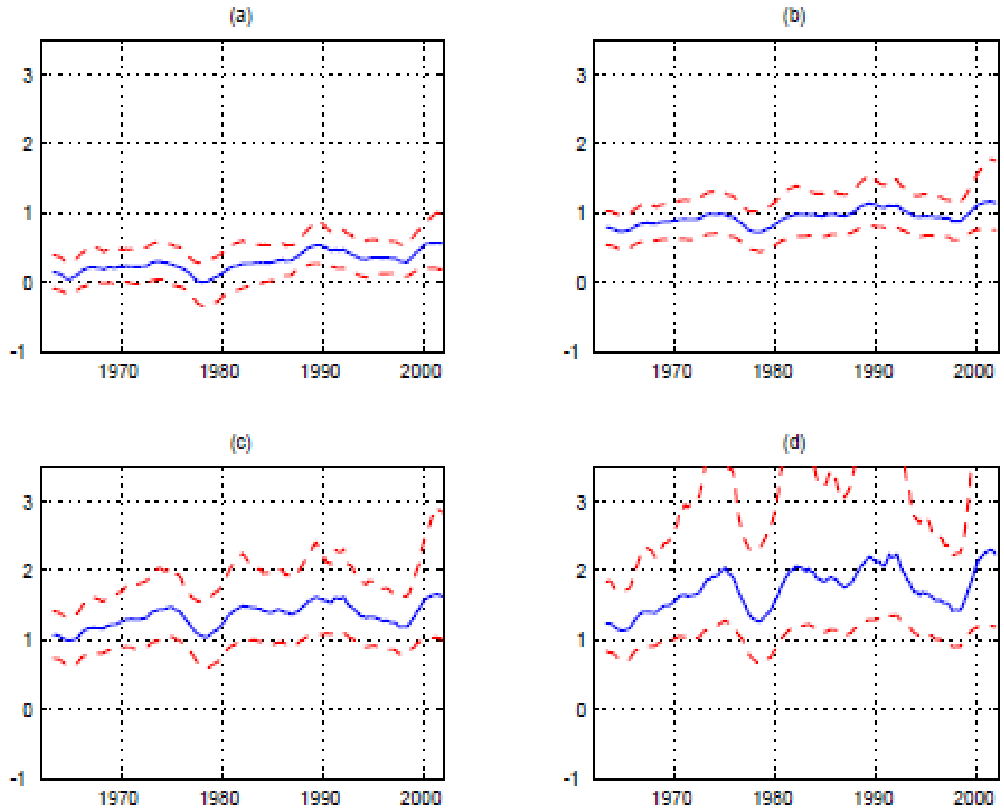


Figure 4: Interest rate response to a 1% permanent increase of inflation with 16th and 84th percentiles. (a) Simultaneous response, (b) response after 10 quarters, (c) response after 20 quarters, (d) response after 60 quarters.

Source: G. Primiceri "Time Varying Structural Vector Autoregressions and Monetary Policy", The Review of Economic Studies, 72, July 2005, pp. 821-852

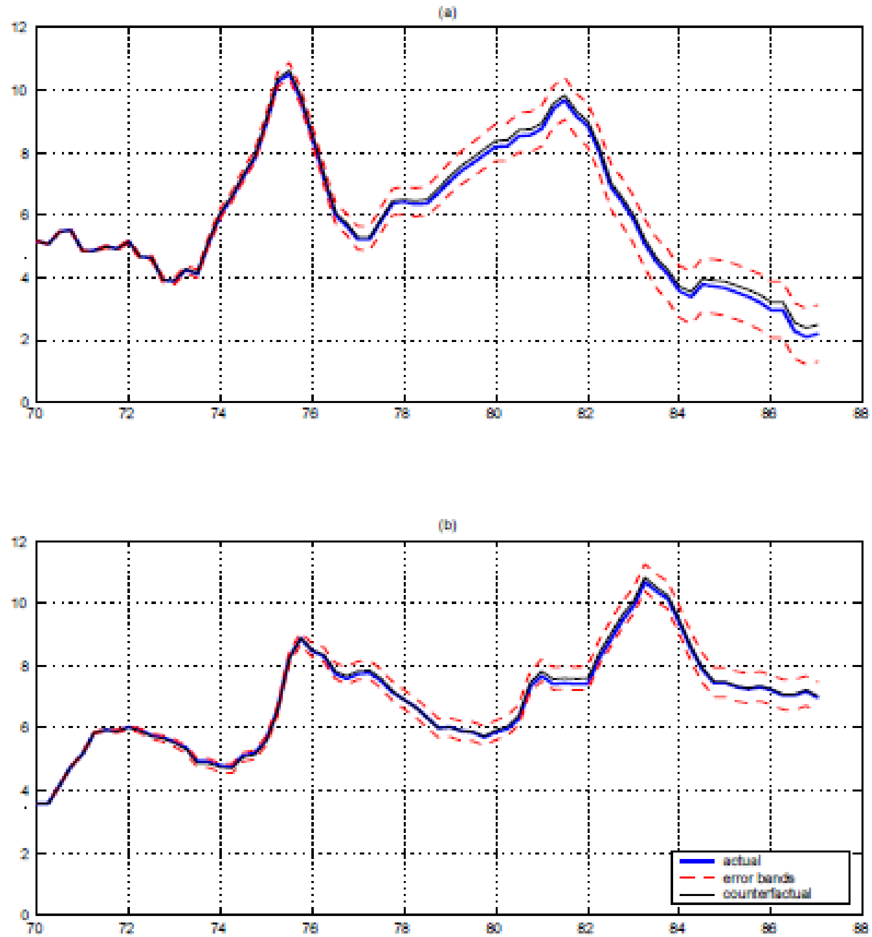


Figure 8: Counterfactual historical simulation drawing the parameters of the monetary policy rule from their 1991-1992 posterior. (a) Inflation, (b) unemployment.

Source: G. Primiceri "Time Varying Structural Vector Autoregressions and Monetary Policy", The Review of Economic Studies, 72, July 2005, pp. 821-852

Application 3: Gali and Gambetti (2009, AEJ-Macro) on the Great Moderation

Sharp reduction in the volatility of US output growth starting from mid 80's.

Kim and Nelson, (REStat, 99).

McConnel and Perez-Quiros, (AER, 00).

Blanchard and Simon, (BPEA 01).

Stock and Watson (NBER MA 02, JEEA 05).

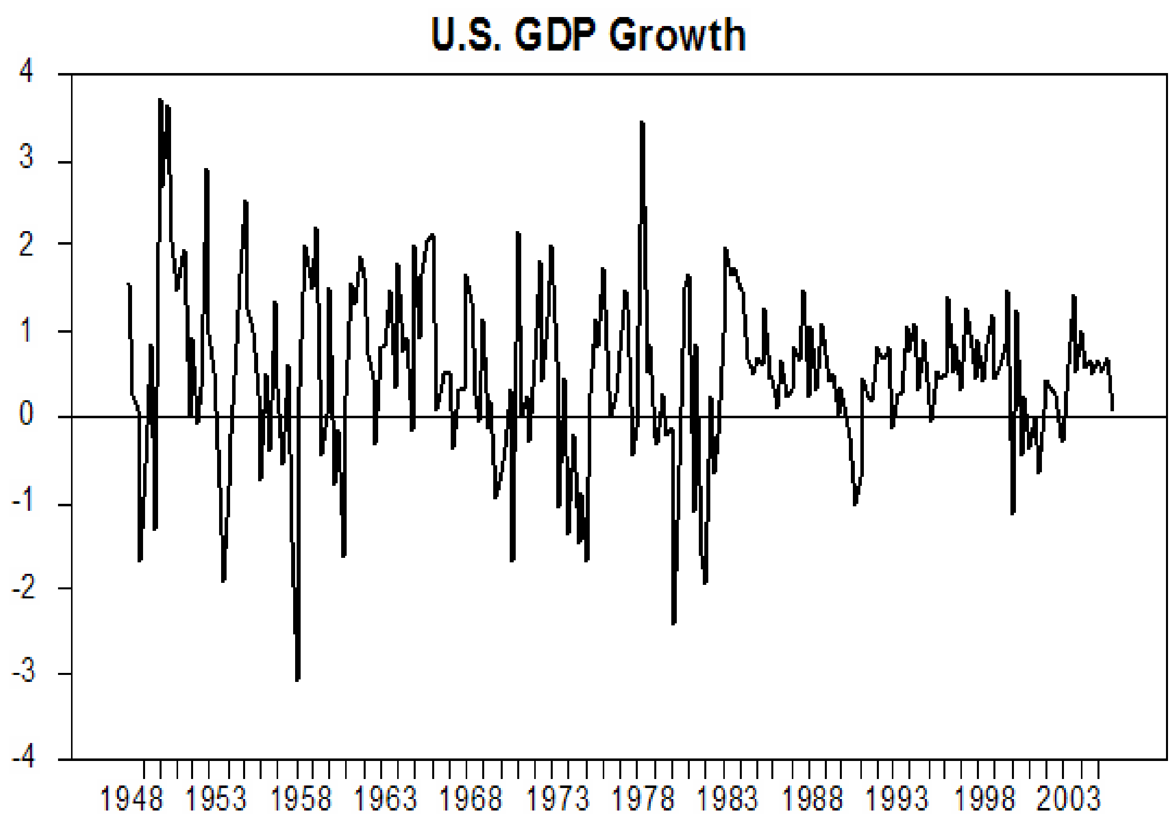

U.S. GDP Growth
4
3
2
1
0
-1
-2
-3
-4
1948
1953
1958
1963
1968
1973
1978
1983
1988
1993
1998
2003

Table 1. The Great Moderation

	Standard Deviation		
	Pre-84	Post-84	$\frac{\text{Post-84}}{\text{Pre-84}}$
First-Difference			
GDP	1.21	0.54	0.44
Nonfarm Business Output	1.57	0.68	0.43
BP-Filter			
GDP	2.01	0.93	0.46
Nonfarm Business Output	2.61	1.21	0.46

The literature has provided three different explanations

1. Strong good luck hypothesis $\Rightarrow$ same reduction in the variance of all shocks (Ahmed, Levin and Wilson, 2002).

2. Weak good luck hypotesis $\Rightarrow$ reduction of the variance of some shocks (Arias, Hansen and Ohanian, 2006, Justiniano and Primiceri, 2005).

3. Structural change hypothesis $\Rightarrow$ policy or non-policy changes (monetary policy, inventories management).

Gali and Gambetti (2009, AEJM) using this class of model try to assess the causes of this reduction in volatility

Idea of the paper very simple. Exploit different implications in terms of conditional and unconditional second moments of the different explanations.

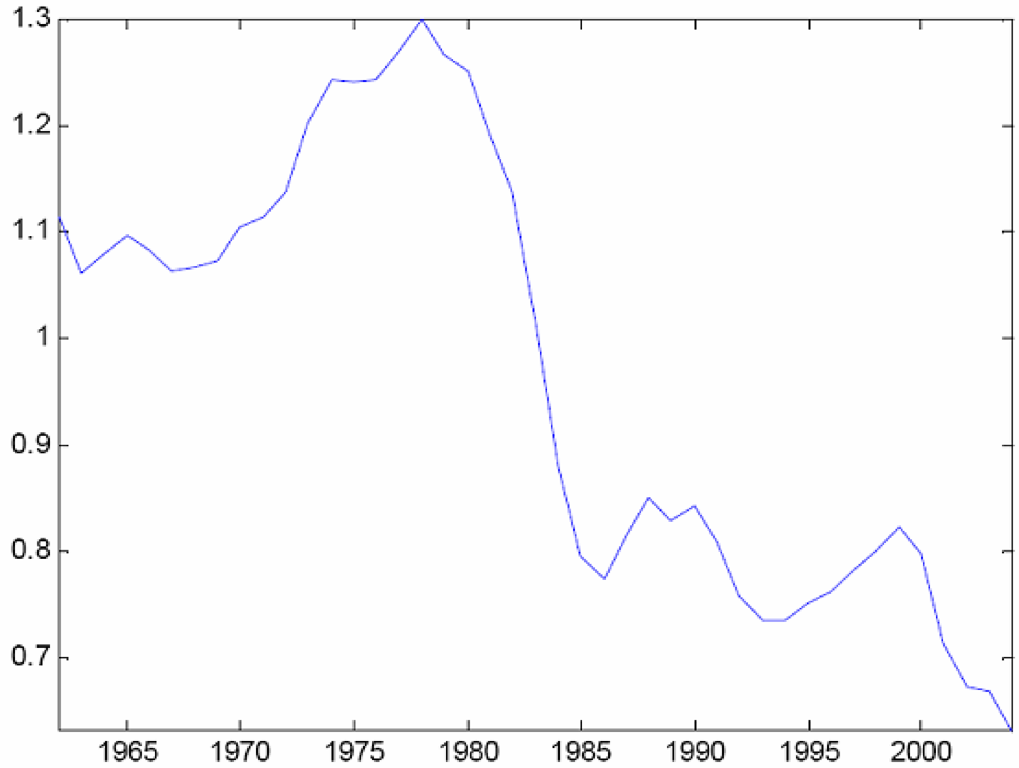
1. Strong good luck hypothesis $\Rightarrow$ scaling down of all shocks variances, no change in conditional (to a specific shock) and unconditional correlations.

2. Weak good luck hypotesis $\Rightarrow$ change in the pattern of unconditional correlations, no change in conditional correlations.

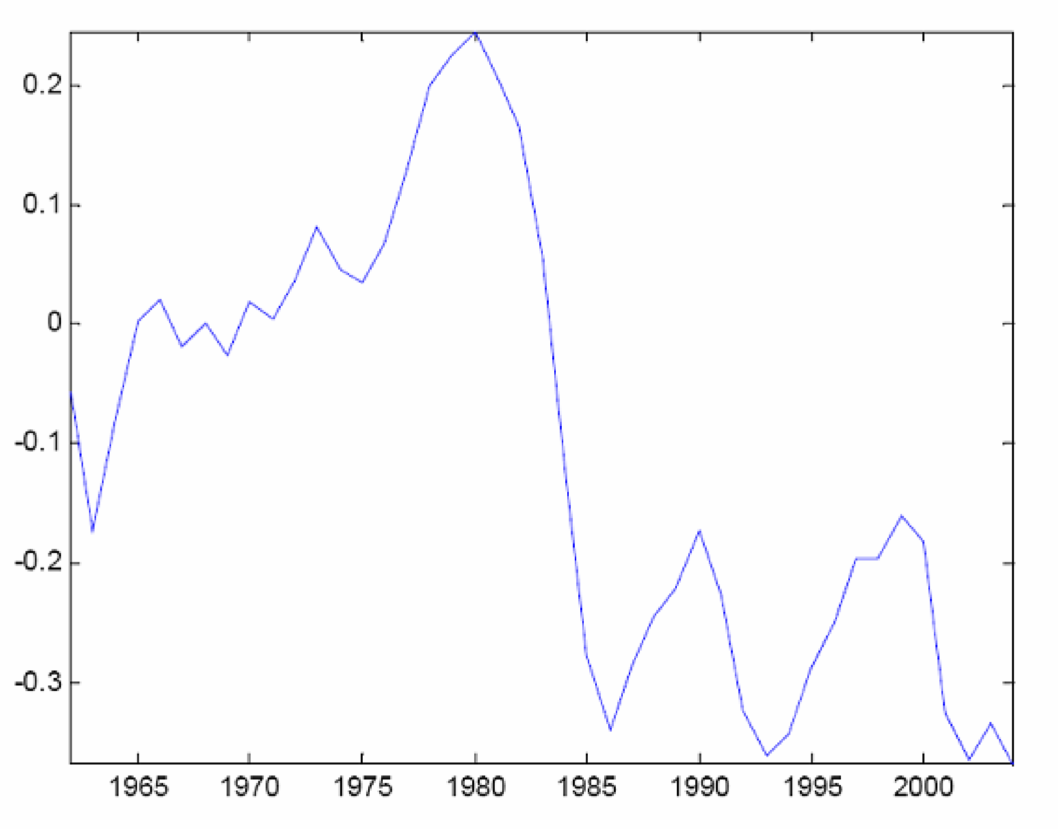
3. Structural change hypothesis $\Rightarrow$ changes in both unconditional and conditional correlations.

Estimate a TVC-VAR for labor productivity growth and hours worked identify a technology and non-technology shock and study model implied second moments.

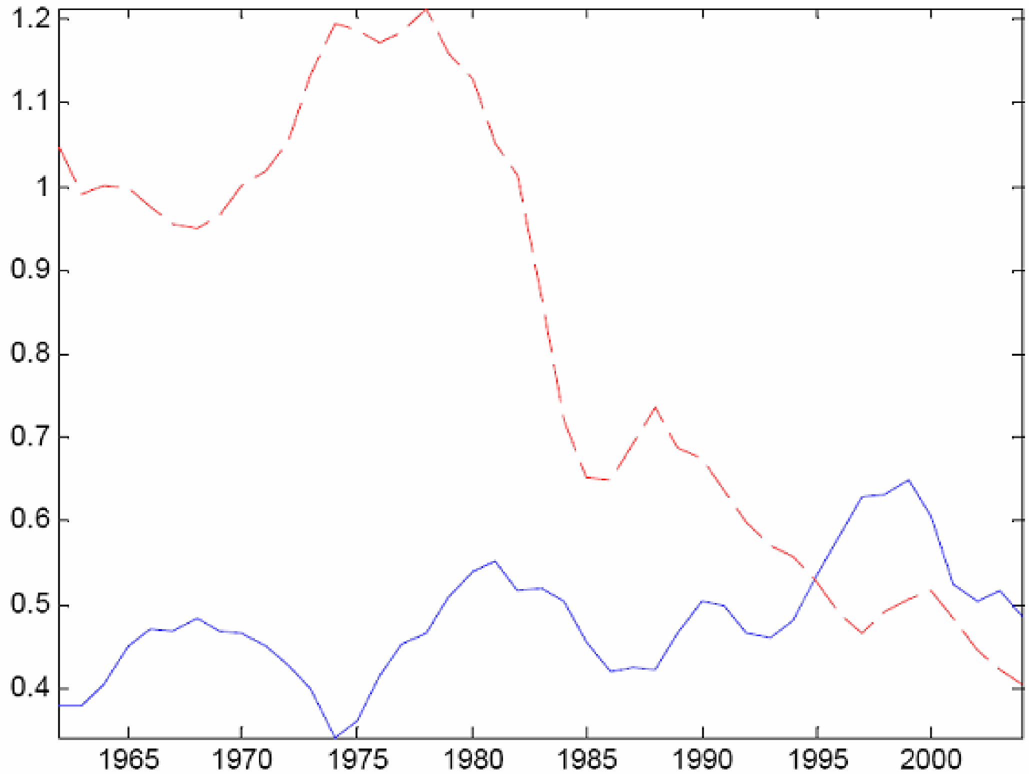
Standard deviation of output growth



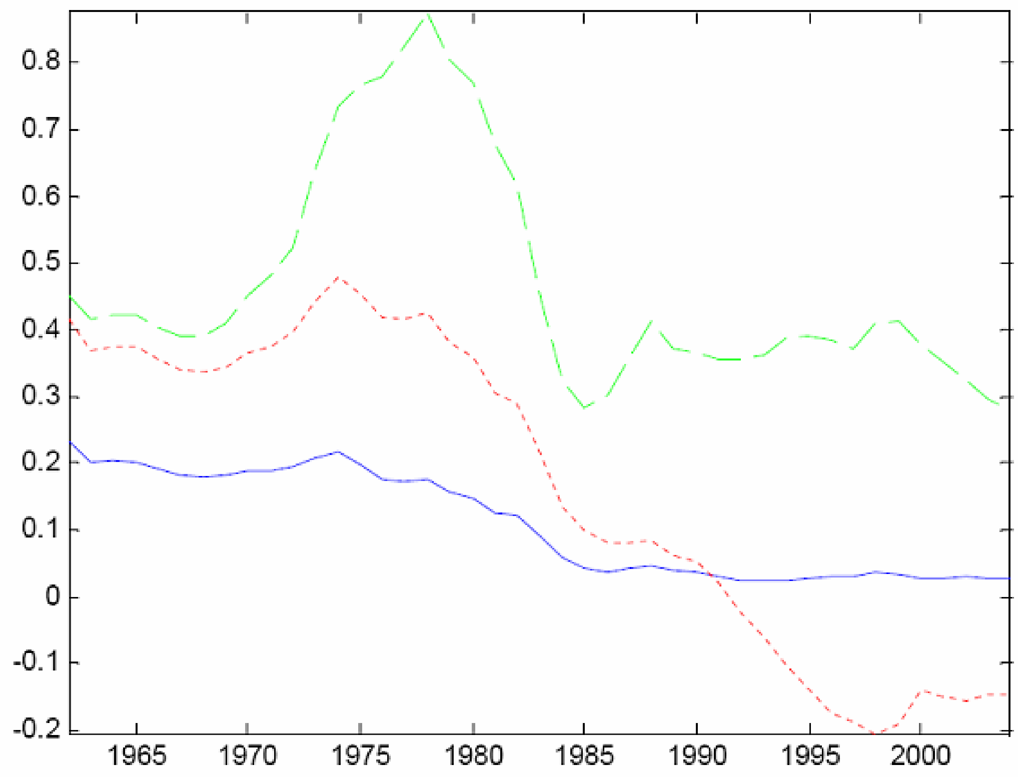
Unconditional moments: correlation of hours and labor productivity growth



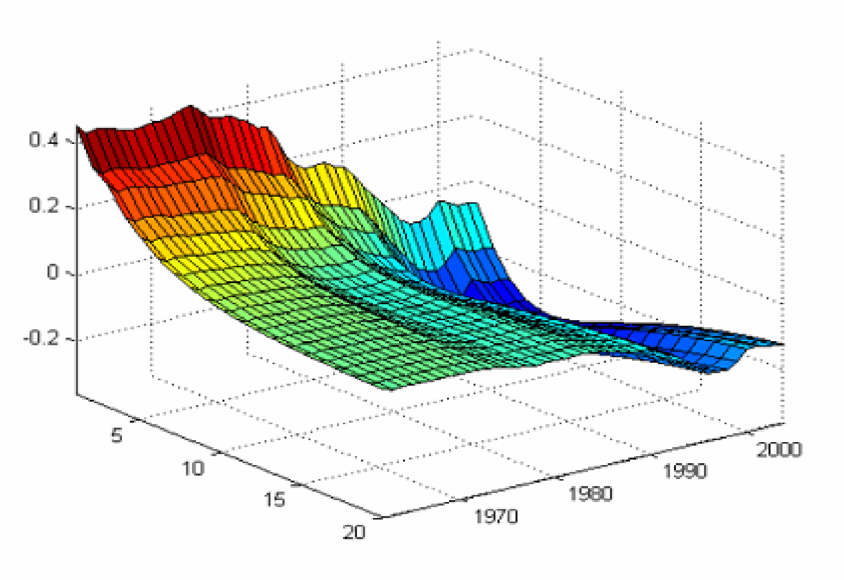
Technology and non-technology components of output growth volatility



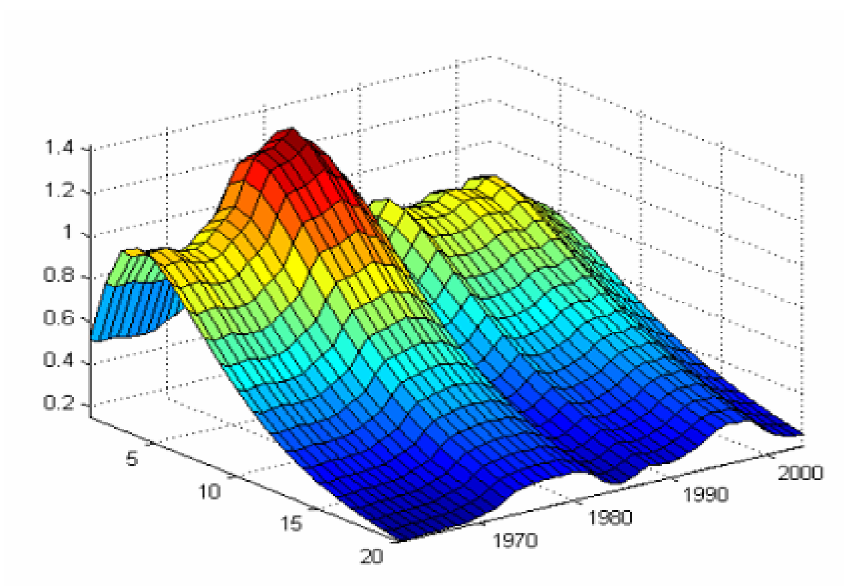
Non technology shock: variance decomposition of output growth



Non technology shock: labor productivity response



Non technology shock: hours response



Application 4: D'Agostino Gambetti and Giannone (forthcoming JAE) on the forecasting

- D'Agostino Gambetti and Giannone (JAE forthcoming) consider whether time-variations can improve upon the forecast made with standard VAR models.

- The result is not trivial: time variations helpful but quite high number of parameter could worsen the forecasts.

- The sample spans from 1948:I-2007:IV. Forecast up to 12 quarters ahead.

- Estimation is in real time.

Table 1: Forecasting Accuracy over the sample 1970-2007: mean square forecast errors.

Horizon (quarters)	Variable	RW (MSFE)	AR-REC (RMSFE)	AR-ROL (RMSFE)	SV-AR (RMSFE)	TV-AR (RMSFE)	VAR-REC (RMSFE)	VAR-ROL (RMSFE)	SV-VAR (RMSFE)	TV-VAR (RMSFE)
	π	2.15	1.13	1.08	1.05	1.03	1.15	1.01	1.14	0.85
1	UR	0.15	1.00	1.08	0.98	1.00	0.99	1.18	0.96	1.04
	IR	0.87	1.12	1.23	1.01	1.04	0.99	1.09	0.94	0.99
	Avg.		1.08	1.13	1.01	1.02	1.04	1.09	1.01	0.96
	π	2.24	1.17	1.03	0.82	0.88	1.37	1.22	1.01	0.62
4	UR	1.07	1.03	1.24	0.95	1.01	0.67	0.91	0.67	0.77
	IR	3.46	1.05	1.20	0.93	0.95	0.96	1.39	0.93	0.93
	Avg.		1.08	1.16	0.90	0.95	1.00	1.17	0.87	0.77
	π	3.06	1.19	1.13	0.89	0.93	1.6	1.38	1.11	0.65
8	UR	2.39	0.95	1.14	0.88	0.95	0.45	0.63	0.42	0.61
	IR	7.54	1.05	1.18	1.11	0.92	0.99	1.44	0.85	0.89
	Avg.		1.06	1.15	0.88	0.93	1.01	1.15	0.80	0.72
	π	3.31	1.28	1.24	0.95	1.00	1.93	1.60	1.26	0.69
12	UR	3.22	0.85	1.12	0.79	0.86	0.47	0.85	0.40	0.51
	IR	10.28	1.08	1.15	0.82	0.91	1.03	1.32	0.79	0.86
	Avg.		1.07	1.17	0.85	0.92	1.14	1.26	0.82	0.72

The table reports the results relative to the forecasting accuracy using point forecasts. The variable we forecast are inflation (π_t), the unemployment rate (UR_t) and the interest rate (IR_t). The forecasting models are: RW - random walk; AR-REC - AR estimated recursively; AR-ROL - AR estimated with a rolling window; TV-VAR - time-varying VAR; VAR-REC - VAR estimated recursively; VAR-ROL - VAR estimated with a rolling window. For the random walk model we report the mean square forecast error (MSFE). For the other models we report the relative mean square forecast error (RMSFE), i.e. the ratio of the MSFE of a particular model to the MSFE of the naïve model. For each horizon it is also reported the average of the RMSFE across variables (Avg.).

Table 2: Forecasting Accuracy over the sample 1985-2007: Mean square forecast errors.

Horizon (quarters)	Variable	RW (MSFE)	AR-REC (RMSFE)	AR-ROL (RMSFE)	SV-AR (RMSFE)	TV-AR (RMSFE)	VAR-REC (RMSFE)	VAR-ROL (RMSFE)	SV-VAR (RMSFE)	TV-VAR (RMSFE)
	π	0.93	2.61	1.19	1.23	1.21	1.29	1.35	1.28	0.98
1	UR	0.05	2.80	1.16	1.05	1.07	1.09	1.17	0.99	1.03
	IR	0.27	3.64	1.08	0.85	0.83	0.87	1.02	0.77	0.83
	Avg.		3.02	1.14	1.05	1.04	1.08	1.18	1.01	0.94
	π	0.45	5.76	1.54	1.19	1.16	2.22	2.64	1.42	0.94
4	UR	0.37	3.00	1.15	0.82	0.82	0.97	1.23	0.77	0.89
	IR	2.09	1.74	1.17	0.78	0.81	0.78	1.20	0.74	0.81
	Avg.		3.50	1.29	0.93	0.93	1.32	1.69	0.97	0.87
	π	0.57	6.39	2.09	1.10	1.08	3.03	3.11	1.55	0.72
8	UR	1.33	1.72	0.86	0.61	0.56	0.42	0.72	0.38	0.57
	IR	5.16	1.53	1.05	0.68	0.74	0.67	1.20	0.63	0.75
	Avg.		3.21	1.33	0.80	0.79	1.37	1.68	0.85	0.69
	π	0.92	4.61	2.10	0.91	0.86	3.47	2.51	1.34	0.46
12	UR	2.25	1.22	0.72	0.48	0.43	0.35	0.73	0.27	0.48
	IR	7.69	1.44	0.89	0.55	0.63	0.70	1.13	0.51	0.61
	Avg.		2.42	1.24	0.65	0.64	1.51	1.46	0.71	0.51

The table reports the results relative to the forecasting accuracy using point forecasts. The variable we forecast are inflation (π_t), the unemployment rate (UR_t) and the interest rate (IR_t). The forecasting models are: RW - random walk; AR-REC - AR estimated recursively; AR-ROL - AR estimated with a rolling window; TV-VAR - time-varying VAR; VAR-REC - VAR estimated recursively; VAR-ROL - VAR estimated with a rolling window. For the random walk model we report the mean square forecast error (MSFE). For the other models we report the relative mean square forecast error (RMSFE), i.e. the ratio of the MSFE of a particular model to the MSFE of the naïve model. For each horizon it is also reported the average of the RMSFE across variables (Avg.).

TVAR and STVAR