Empirical Macroeconomics

Lecturer: Luca Benati

18 October, 2012

Identification based on short-run restrictions: sign restrictions

First, a recap of the motivation behind sign restrictions.

As we said, Canova and Pina (*book chapter*, 2005):

- **general equilibrium logic implies that impact of all shocks at zero should be, in general, non-zero**
- **indeed, this is exactly what you see in DSGE models: all of the elements of the A_0 matrix are typically non-zero ...**

But so, how can you identify anything??

Solution: impose a pattern of signs for the impact of the shocks on the variables such that they are sorted out from one another

Let's see a simple illustration of the basic logic ...

Consider a VAR(p) for inflation and output growth

$$Y_t = B_0 + B_1 Y_{t-1} + B_2 Y_{t-2} + \ldots + B_p Y_{t-p} + A_0 \epsilon_t$$

Assume that there are two shocks, demand and supply: simple way to separate them is to postulate these signs for their impact at $t = 0$ on the two variables ...

$$\begin{bmatrix} \pi_t \\ y_t \end{bmatrix} = \begin{bmatrix} + & - \\ + & + \end{bmatrix} \begin{bmatrix} \epsilon_t^D \\ \epsilon_t^S \end{bmatrix}$$

Signs come from simple aggregate demand-aggregate supply framework ...

- **a positive demand shock raises both inflation and output growth**
- **a positive supply shock raises output growth and decreases inflation**

For example, an A_0 matrix satisfying these sign restrictions is:

$$A_0 = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix}$$

and for given VAR residuals u_t, you can immediately recover the demand and supply structural shocks:

$$\begin{bmatrix} \epsilon_t^D \\ \epsilon_t^S \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix}^{-1} \begin{bmatrix} u_t^\pi \\ u_t^y \end{bmatrix}$$

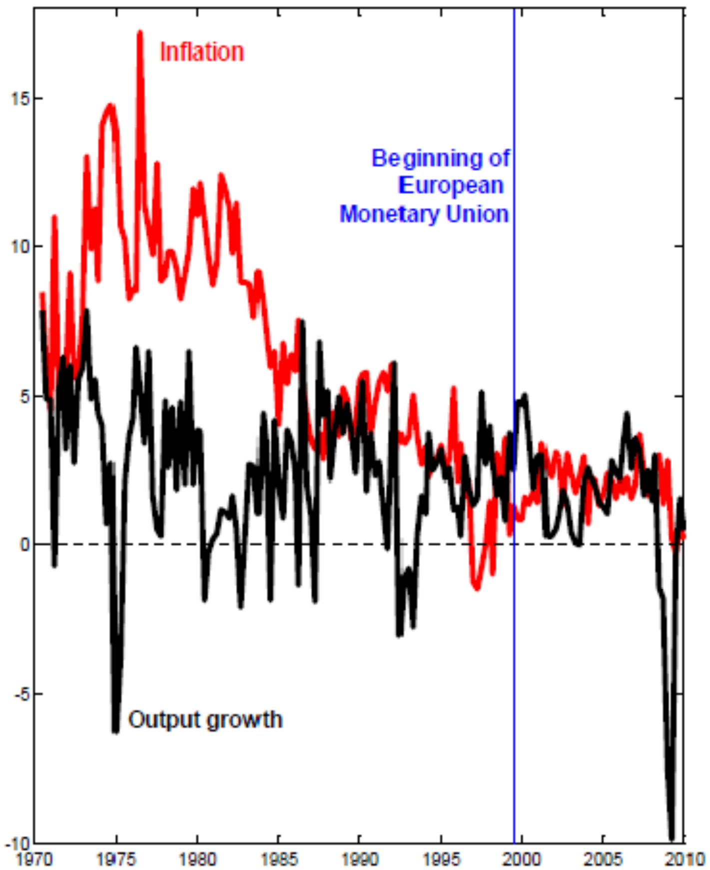
where u_t^π and u_t^y are VAR residuals for the inflation and output growth equations, respectively ...

Obvious issue with all this: the above matrix A_0 is not the only one satisfying these sign restrictions ...

➔ In fact, from a mathematical point of view, there is an infinite number of matrices satisfying these restrictions ...

How do we get them?

Let's see a practical example. We have inflation and output growth data for the Euro area, and we estimate a VAR(p).



Key thing we need to get the impact matrix A_0 is the VAR's covariance matrix, since we know that:

$$\mathrm{Var}(u_t) \equiv V = A_0 A_0'$$

In this case we have:

$$\hat{V} = \begin{bmatrix} 2.6214 & -0.0707 \\ -0.0707 & 5.4280 \end{bmatrix}$$

Each of the A_0 matrices must satisfy:

$$A_0 A_0' = \begin{bmatrix} 2.6214 & -0.0707 \\ -0.0707 & 5.4280 \end{bmatrix}$$

Several algorithms have been devised to get the distribution of the elements of A_0 satisfying 2 things:

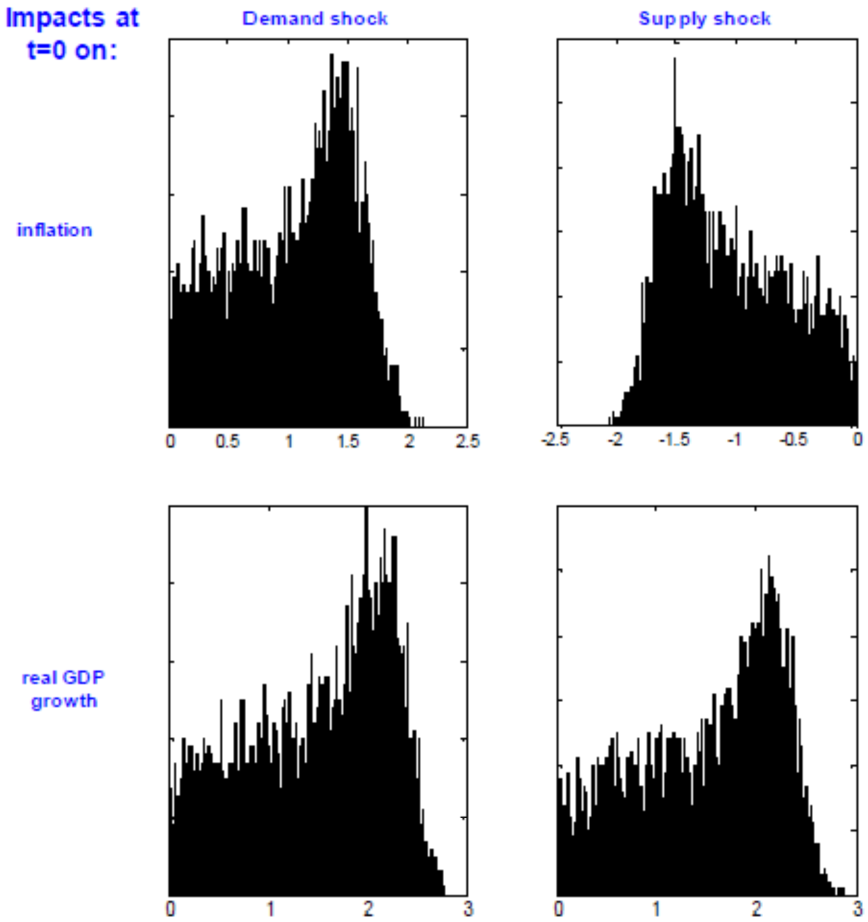
- **the equation $V = A_0 A_0'$, and**
- **a specific pattern of signs ...**

Today, the standard is the algorithm proposed by:

J. Rubio-Ramirez, D. Waggoner, and T. Zha (2010), 'Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference', *Review of Economic Studies*, 77(2), pp. 665-696

Let's see an application of this to the simple bivariate system for inflation and GDP growth in the Euro area ...

Distributions of impacts of demand and supply shocks at $t = 0$, based on RWZ algorithm ...



For each draw from the distribution, the four elements of the matrix A_0 satisfy the relationship in slide 47:

$$A_0 A_0' = \begin{bmatrix} 2.6214 & -0.0707 \\ -0.0707 & 5.4280 \end{bmatrix}$$

To be noticed: distributions of impacts are quite 'spread out', thus pointing towards a substantial imprecision in estimating the impact matrix ...

As we'll see, that's indeed one of key problems of this methodology, and it is pretty much 'intrinsic' to the method …

Why is that? Because sign restrictions are 'weak information', and it should therefore not be surprising that they produce comparatively imprecise results …

Intuition: suppose I ask you to guess two numbers I have thought of, and only piece of information I give you is:

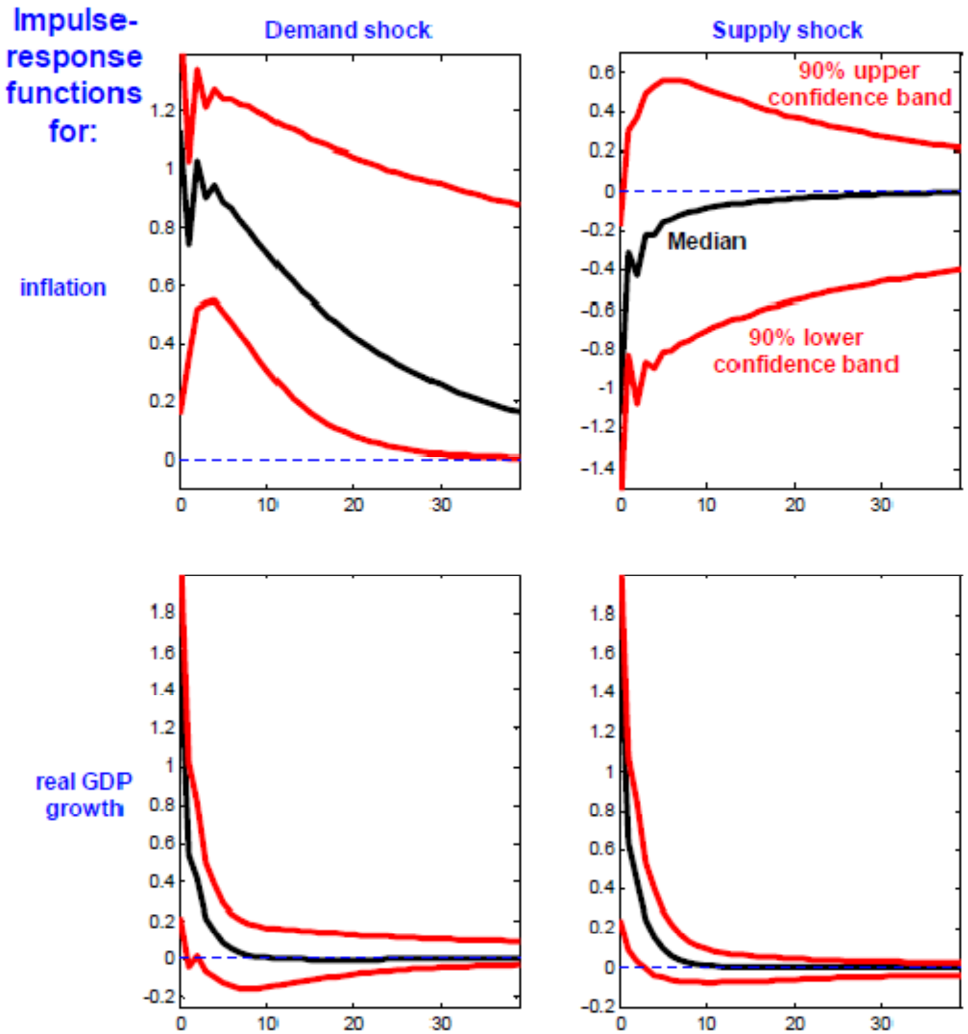
- **one number is positive, the other is negative**
- **their sum is equal to 7**

and that's it … Quite obviously, you're going to have a hard time in guessing the numbers correctly …

Logic is the same for Rubio-Waggoner-Zha algorithm, or for any other method you can think of: the information you have is, quite simply, pretty limited …

Next: the impulse-response functions …

Since restrictions have been imposed only on impact, during subsequent quarters IRFs are not constrained in any way ...



Indeed, in 3 cases out of 4, after impact a non-negligible number of IRFs crosses zero boundary ...

For inflation IRFs, imprecision is quite substantial ..

Why only imposing restrictions on impact? Several papers impose them for several quarters, but Canova and Paustian stress that beyond impact sign restrictions are often 'whimsical', which means, basically, elusive ...

Where do the sign restrictions come from?

In earlier applications of this methodology, signs were motivated *via* either very simple macroeconomic models (e.g., the AD-AS model), or intuitive reasoning …

Recently, profession moved towards the following approach:

- **consider a sufficiently vast class of DSGE models, comprising, as special cases, many alternatives of interest**
  ➔ e.g., New Keynesian models and Real Business Cycle models can be thought of as special cases of a single encompassing model
- **consider a wide range of parameters values for the structural coefficients**
- **focus on the signs of the IRFs on impact which are robust across all possible values of the parameters, and impose them on the data …**

Let's see an example: consider the standard New Keynesian model:

$$R_t = \rho R_{t-1} + (1-\rho)[\phi_\pi \pi_t + \phi_y y_t] + \epsilon_{R,t}$$

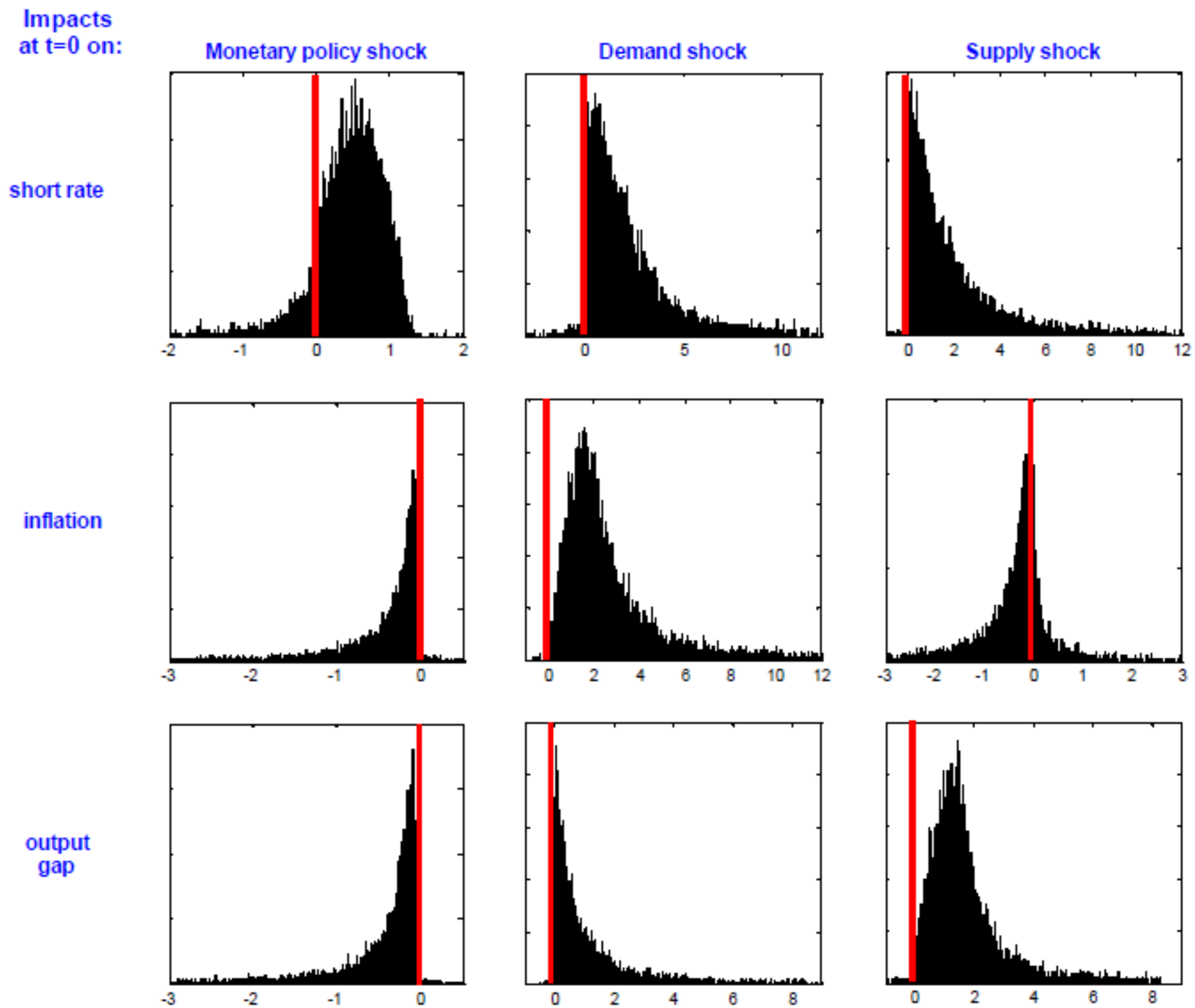
$$\pi_t = \frac{\beta}{1+\alpha\beta}\pi_{t+1|t} + \frac{\alpha}{1+\alpha\beta}\pi_{t-1} + \kappa y_t + \epsilon_{\pi,t}$$

$$y_t = \gamma y_{t+1|t} + (1-\gamma) y_{t-1} - \sigma^{-1}(R_t - \pi_{t+1|t}) + \epsilon_{y,t}$$

Then, following Canova and Paustian (*JME*, 2011) I consider uniform distributions for the parameters' values, defined over 'plausible' intervals …

For example, α, γ, and all of the ρ's are defined over [0, 1] …

Then I randomly draw 10,000 parameters vectors and I look at the signs of the impacts at $t = 0$ …



As you see, in vast majority of cases signs of shocks on impact are indeed robustly either positive or negative, which justifies the approach ...

Impacts for which sign is not sufficiently robust are left unconstrained ...

Limitations of sign restrictions

Limitations of the sign restrictions approach to identifying structural VARs have been mostly discussed by Fry and Pagan (*JEL*, forthcoming)

There are 2 key problems with this approach:

- **because of the infinite number of possible solutions, the true values of the elements of the impact matrix will be obtained only by chance …**

  ➔ mathematically, probability of getting true values is zero

  This means the estimates of IRFs obtained with sign restrictions are in general biased, and magnitudes are in general wrong—and you don't know by how much ...

  ➔ only thing you get right is the sign …

- **Fry and Pagan (2007):** ***'It is a powerful adage that weak information produces weak results, and sign restrictions are weak information.'***

  → you're just telling the algorithm: *'give me a positive (negative) number'*, and that's it ...

  You can't possibly expect to get very strong results based on such weak restrictions ...

The problem of weak information becomes more and more acute the less and less restrictions you impose.

Suppose you have a VAR with 5 variables: the structural impact matrix, A_0, is 5×5, which means it has 25 elements upon which you ought to impose sign restrictions ...

If you impose sign restrictions on all of the 25 elements, you'll get results with a certain extent of precision ...

But suppose you only identify one shock, which means you impose at most 5 restrictions ...

The results you'll get will be much less precise ...

In particular, Raf Wouters showed, based on a calibrated DSGE model, that the less restrictions you impose, the less precise the results you obtain are ...

This brings us to Uhlig's (*JME*, 2005) investigation of the impact of monetary policy on output ...

Uhlig's (*JME*, 2005)

In a nutshell: Uhlig performs an 'agnostic investigation' of the impact of monetary policy shocks on output ...

Why 'agnostic'? Because he imposes a truly minimal set of restrictions ...

- **he estimates a VAR for six variables: real GDP, GDP deflator, FED Funds rate, total reserves, non-borrowed reserves, and a commodity price index**
- **he only identifies one shock, the monetary policy shock ...**
- **this shock:**
  (*i*) is identified by its responses on the FED Funds rate, non-borrowed reserves, and prices,but
  (*ii*) its impact on output is left completely unrestricted ...

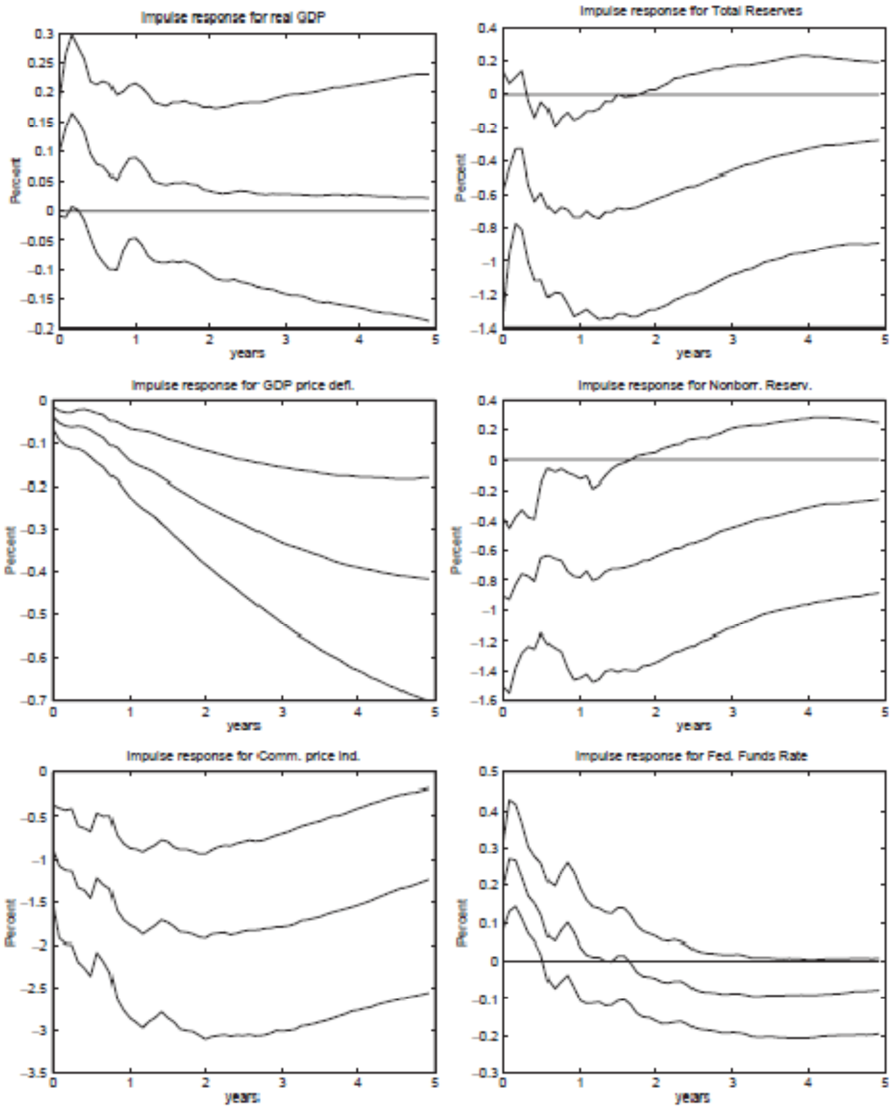
Impact of monetary policy shock on output is focus of Uhlig's paper ...

His key finding is that a contractionary monetary policy shock has an ambiguous effect on output ...
➔ it may either increase it or decrease it, with probabilities equal to roughly one-third and two-thirds ...

VAR is estimated based on monthly data ...

Sign restrictions are imposed both on impact, and for the subsequent *K* months ...

Let's see his benchmark results, based on $K = 5$ ...

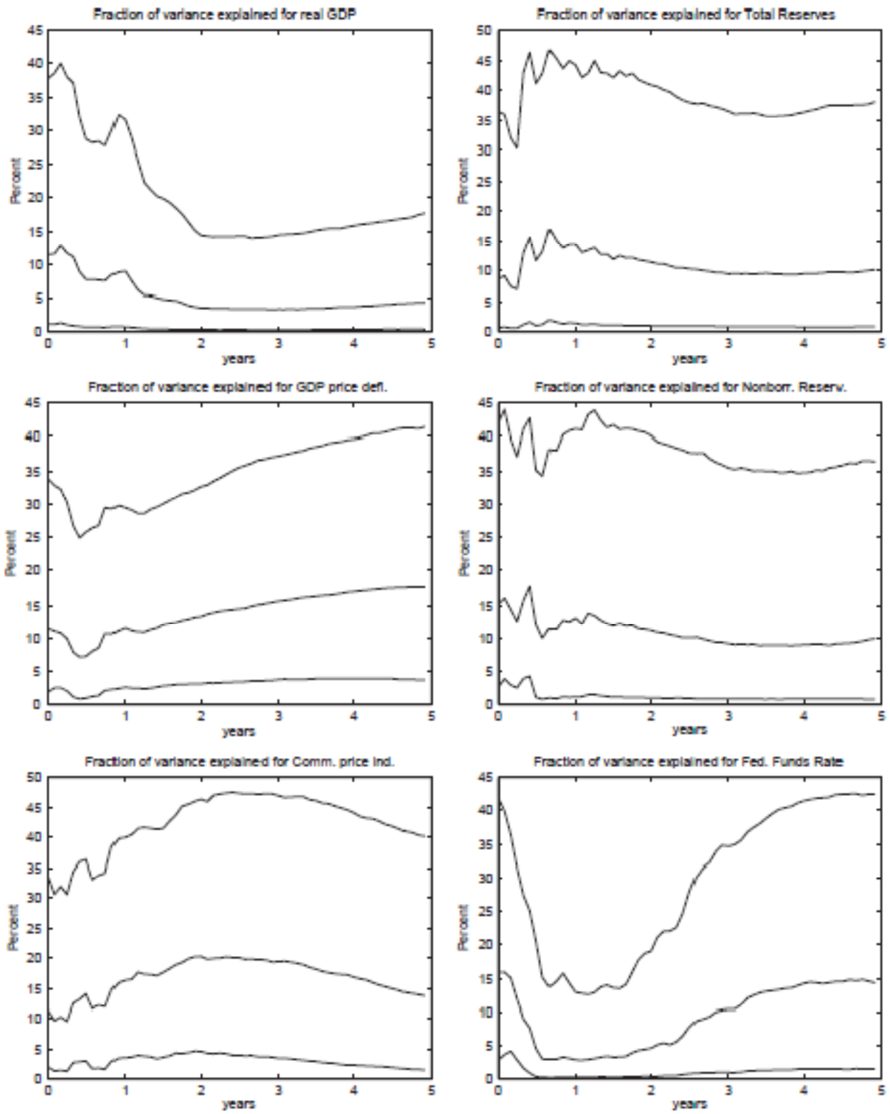


IRFs to monetary shock when restrictions are imposed at zero and for subsequent 5 months

FED Funds rate behaves as expected, it jumps up and then slowly gets back …

Notice the permanent negative impact on the price level …

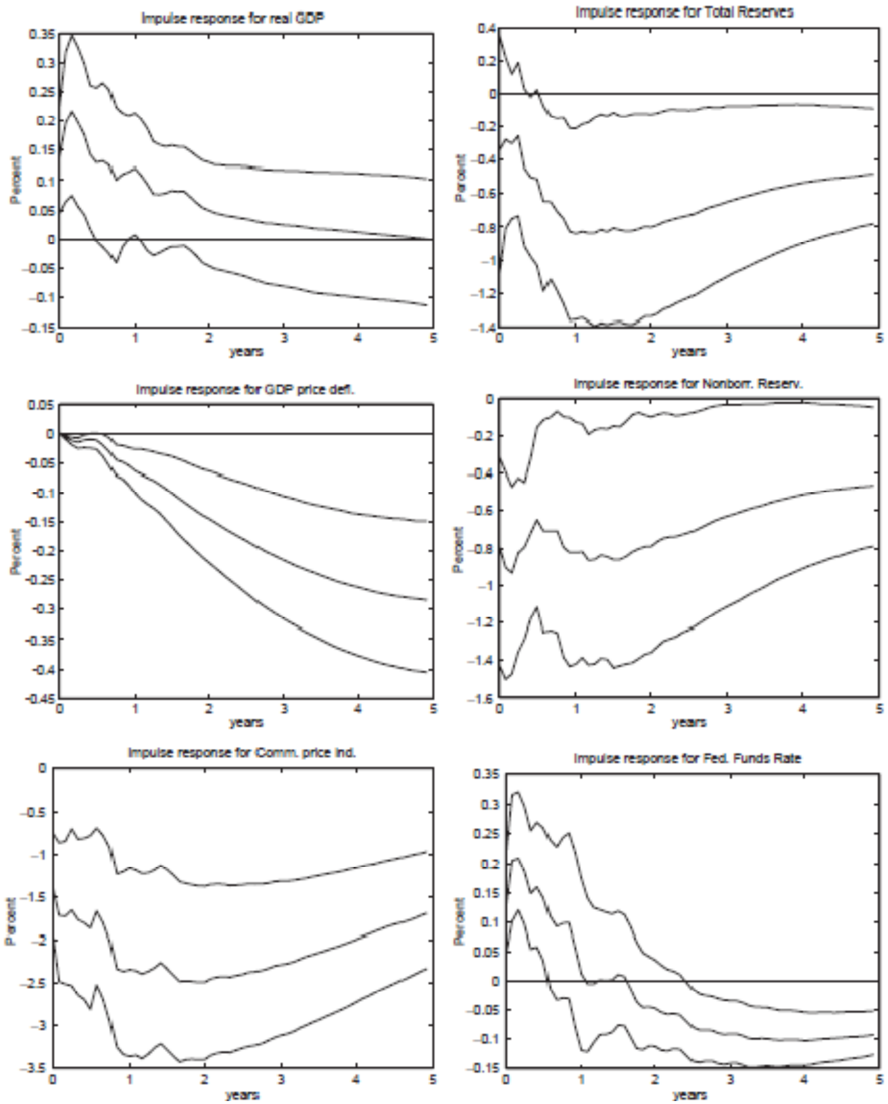
A quite stunning result is that for output impact is positive , rather than negative: a monetary contraction is estimated to lead to an economic expansion …



What about fractions of variance of different variables explained by monetary shock?

On the left, variance decomposition at various horizons: in line with most of literature, monetary shock explains, based on median estimates, about 5% of variance of real GDP at horizons beyond 2 years …

Fraction is greater for GDP deflator and FED funds rate, but it is not exceptionally high …



Result is robust to a number of 'manipulations': e.g., on the left, if you move one step towards inertial restrictions, and you impose a zero contemporaneous response of GDP deflator to the monetary policy shock …

But how can this be? This runs against what overwhelming majority of economists and central bankers believe ...

'A monetary contraction decreases both inflation and output'

And so, how can Uhlig's result be rationalised??

Raf Wouters, in discussing (back in 2005) a paper by Scholl and Uhlig based on exactly the same methodology as Uhlig (*JME*, 2005), showed that with this methodology:

- **only one shock identified, and**
- **very little restrictions**

you can get basically anything for the series you have left unrestricted ...

Wouters' discussion of Scholl and Uhlig (2005, mimeo)

Questions Wouters tackle:

- ***'Is minimal set of sign restrictions sufficient to recover IRFs and historical time series of monetary policy shock?'***
- ***'Is it possible that other shocks (or combinations of shocks) also fulfil the sign restrictions and thereby distort identification?'***

His approach: ***'Apply identification strategy to simulated data from an estimated DSGE model. Then check whether approach correctly identifies IRFs and historical time series of the monetary policy shocks.'***

His conclusion: ***'Simulation results indicate that there is no guarantee that a minimum set of sign restrictions will correctly identify the monetary policy shock.'***

In particular, he shows that

- **IRFs estimated based on minimal set of sign restrictions are sometimes significantly distorted ...**

  ➔ this is a very simple and elegant explanation for the 'strange' result obtained by Uhlig (*JME*, 2005)
- **estimated monetary policy shocks are often weakly correlated with the true shocks, generated by the DSGE model ...**

Let's see a simple example in the spirit of Wouters' point ...

I take the usual standard New Keynesian model:

$$R_t = \rho R_{t-1} + (1-\rho)[\phi_\pi \pi_t + \phi_y y_t] + \epsilon_{R,t}$$

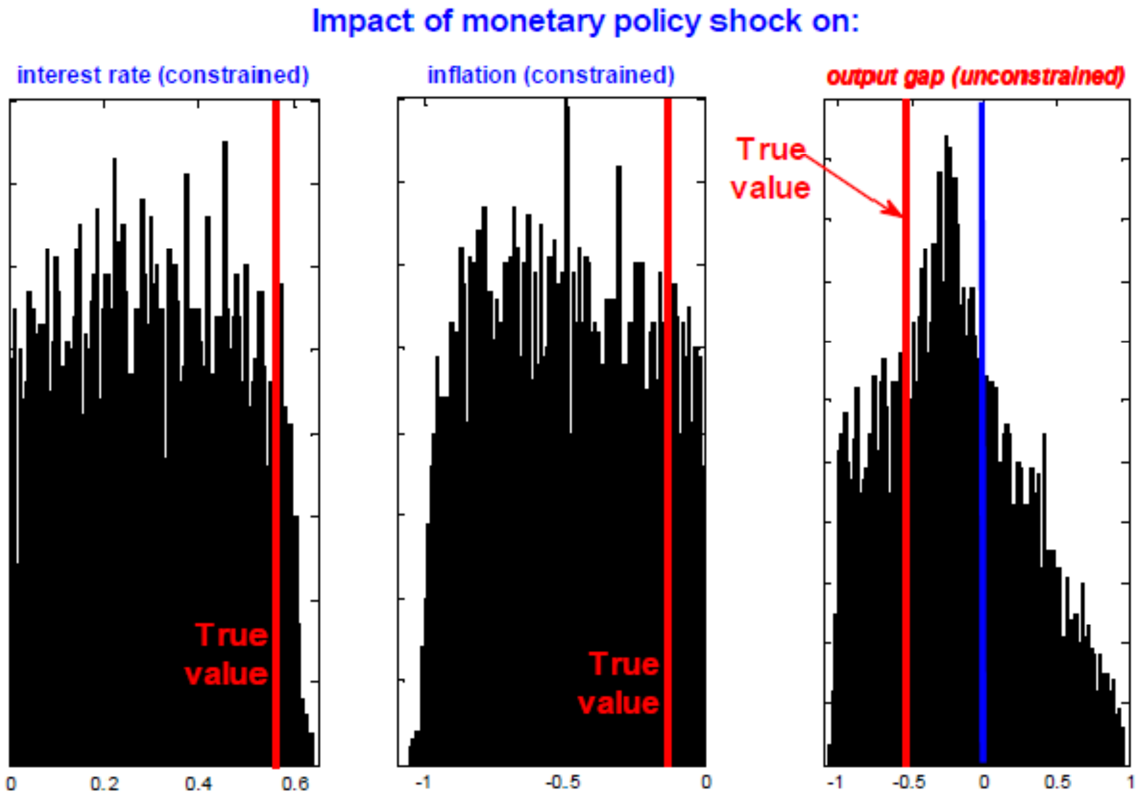
$$\pi_t = \frac{\beta}{1+\alpha\beta}\pi_{t+1|t} + \frac{\alpha}{1+\alpha\beta}\pi_{t-1} + \kappa y_t + \epsilon_{\pi,t}$$

$$y_t = \gamma y_{t+1|t} + (1-\gamma) y_{t-1} - \sigma^{-1}(R_t - \pi_{t+1|t}) + \epsilon_{y,t}$$

- **I calibrate it based on estimates from a paper of mine,**
- **I simulate it 5,000 times,**
- **Each time I estimate a VAR and, following Uhlig, I identify a monetary policy shock**
- **I only restrict its impact on the short rate and inflation**
- **As Uhlig did, I leave the impact on output unrestricted**

Let's see the results ...

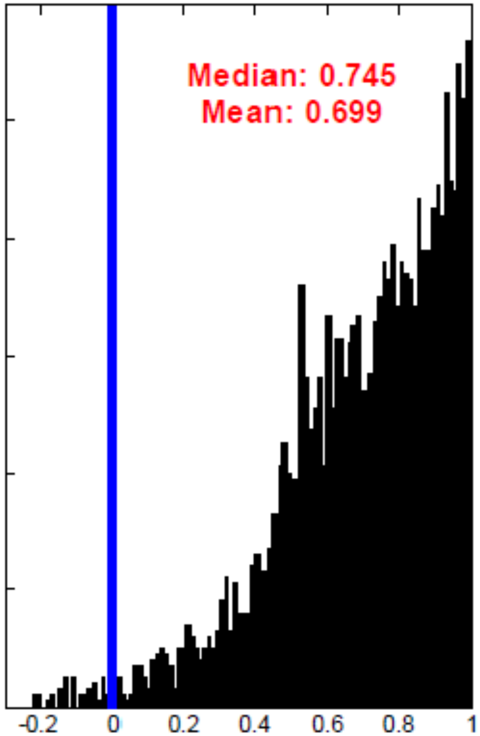
Distributions of the estimated impacts of policy shock:



Two things to notice:

(1) Even when you impose the restrictions, results you get are biased (see interest rate and inflation)
➔ Fry and Pagan's point: *'weak information, weak results'*

(2) For output gap—which is left unrestricted—impact may be positive or negative, even if true impact is negative ...



On the left, distribution of contemporaneous correlation between

- true monetary policy shocks, as generated by DSGE model, and
- monetary policy shocks identified based on sign restrictions, imposing 'minimal' set of restrictions (just 2)

Estimates of policy shocks are not very precise: mode of distribution is at 1—what we'd like to get—but

- distribution is quite spread out,
- median is 0.75

Also, keep in mind we are imposing 2/9 = 22.2% of possible restrictions: Uhlig imposed 3/25 = 12% of restrictions, so it is to be expected his results are even less precise ...

A further important point on this:

- **shock for which you leave one or more impacts unrestricted should explain a sufficiently large fraction of the variance of the data ...**

  ➔ that is: it ought to be a 'big shock' ...

  ➔ if not, the data simply do not contain enough information about what this shock does ...
- **monetary policy shocks—as opposed to the systematic component of monetary policy—are usually thought to explain very little of the variance of the data ...**
- **it is not surprising, therefore, that estimating the unrestricted effects of a policy shock on some variables may produce imprecise results ...**

Once again, Fry and Pagan's point ...

So what should we conclude?

Uhlig's result may be genuine, but it might also be the product of a methodology that uses very little information and imposes little restrictions ...

Indeed, the conclusion from Paustian (*B.E. Journal of Macroeconomics*, 2007) is that

'[t]wo conditions must be met for sign restrictions to unambiguously deliver correct sign of unconstrained IRFs. First, sufficiently large number of restrictions must be imposed—more than what is typically imposed in applied work. Second, variance of the shock under study must be sufficiently large.'

Finally, let's see an important point made by Fry and Pagan, concerning not sign restrictions *per se*, but rather a specific aspect of the traditional implementation of this method ...

As we pointed, researchers typically

- **estimate VAR, and get estimate of covariance matrix of innovations, V ...**
- **based on covariance matrix of innovations they impose sign restrictions, and via (say) Rubio-Waggoner-Zha algorithm they get distribution of A_0 matrices ...**
- **based on this, they compute the IRFs to structural shocks**
- **finally, they sort the IRFs and get the median and the confidence bands ...**

Fry and Pagan point out that the 'sorting' stage leads to an undesirable consequence ...

Two key features of identification are:

- **each A_0 matrix corresponds to a different economic model**
- **identified shocks associated to each A_0 matrix are orthogonal to one another ...**

This implies that IRFs computed based on each A_0 matrix show response to orthogonal shocks, which is what we want in the first place ...

However: when you sort the IRFs and you compute the percentiles—which you need to get the median and the confidence bands—you break down all this ...

For example, focusing on the median: there is no guarantee at all that all elements of the median IRF at all horizons have been produced by the same A_0 matrix ...

On the contrary, in general

- **median IRF on impact will come from *h*-th matrix A_0;**
- **median IRF at horizon 1 will come from *k*-th matrix A_0;**
- **median IRF at horizon 2 will come from *j*-th matrix A_0;**
- **and so on ...**

These means that in general these IRFs do not show the reaction of the economy to orthogonal shocks, but rather to shocks which are correlated ...

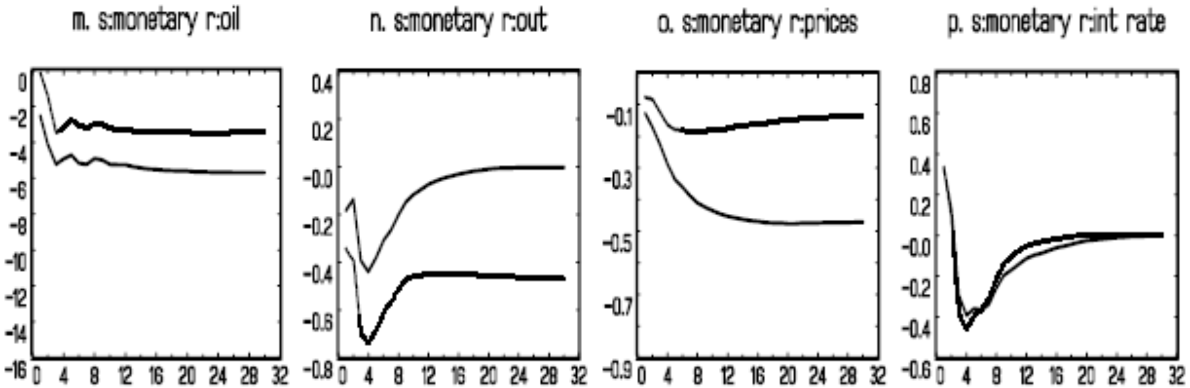
How to solve this problem? Fry and Pagan suggest a very simple solution:

- **compute median IRFs as people typically do, and then**
- **focus on the IRF produced by a single A_0 matrix which is closest to the median of the distribution ...**

In this way you still capture the notion of 'median' of the distribution of the IRFs, which people are typically interested in, but:

- IRF you consider has been generated by a single A_0 matrix (that is: a single model), and therefore
- the associated structural shocks are orthogonal

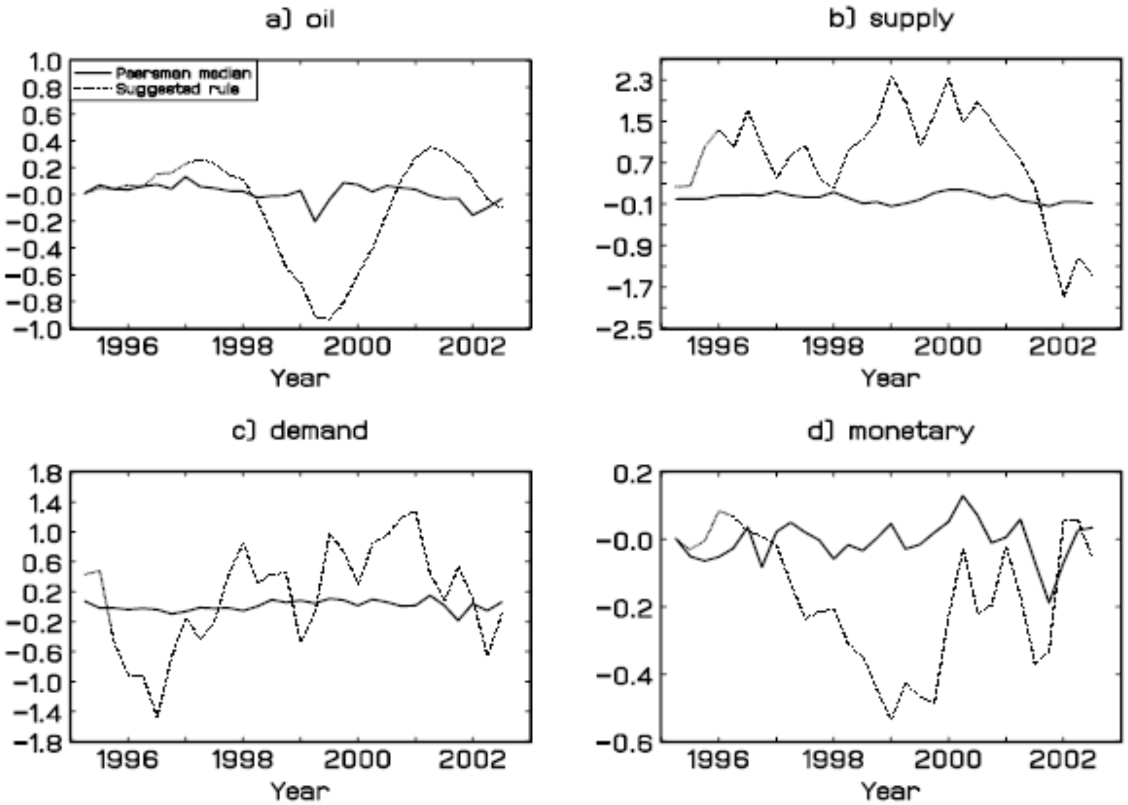
Fry and Pagan show application to Peersman's (2005) model:



Thin lines are from Fry-Pagan method: most of times quite different from median IRFs ...

Fry and Pagan's point appears to be pretty much relevant ...

Results are much worse for the historical decomposition of output growth into the components due to the various shocks ...



Decomposition based on the median is significantly different from the one based on Fry and Pagan's alternative method …

Bottom line: keep this in mind in applied research, it is potentially very serious problem …

Imposing sign restrictions: how it is done in practice

Let's now see how all this is done in practice ...

As a matter of fact, imposing sign restrictions means

'imposing on the data prior information the researcher has about the sign of the impulse-response functions of endogenous variables to structural shocks ...'

... and that's why sign restrictions are typically imposed by working within a Bayesian context ...

Why is that, exactly???

Because that's the entire point of Bayesian econometrics: combining the information contained in the data with prior beliefs the researcher has about the economy ...

So when you want to impose sign restrictions on the IRFs, a Bayesian approach is the natural way to go ...

Combining the prior with the information contained in the data

How is it done? In this way:

- **we start by specifiying a prior distribution for the parameters, reflecting our beliefs about 'how likely' different values for them actually are, and**
- **we combine this prior information with the information contained in the data …**

In practice, this means revising the prior in the light of the data, thus getting the posterior distribution ...

How do we do this? In the following way, *via* Bayes' Law ...

Given a sample of data Y_t, and parameters θ, we have

$$F(Y_t, \theta) \; = \; \phi(\theta) \times L(Y_t \mid \theta) = \psi(Y_t) \times P(\theta \mid Y_t)$$

where

$$\begin{aligned}
F(Y_t, \theta) &= \text{joint density of the data and the parameters} \\
\phi(\theta) &= \text{prior distribution of the parameters} \\
L(Y_t \mid \theta) &= \text{likelihood function of the data conditional on the parameters} \\
\psi(Y_t) &= \text{density of the data (key: it is independent of the parameters)} \\
P(\theta \mid Y_t) &= \text{posterior distribution of the parameters conditional on the data}
\end{aligned}$$

This means that we can rearrange the above equation in order to get the posterior distribution of the parameters, reflecting

- **the prior information we have, and**
- **the information contained in the data, as encoded in the likelihood function**

So we have the posterior distribution of the parameters conditional on the data:

$$P(\theta \mid Y_t) = \frac{\phi(\theta) \times L(Y_t \mid \theta)}{\psi(Y_t)}$$

Finally, since $\psi(Y_t)$ is just a number, we take it out, and we have that the posterior distribution is proportional to the product of the prior and the likelihood

$$P(\theta \mid Y_t) \propto \underbrace{\phi(\theta)}_{\text{Prior distribution}} \times \underbrace{L(Y_t \mid \theta)}_{\text{Likelihood of the data}}$$

Getting the posterior distribution

In practice, there are 2 ways to get the posterior distribution:

- **in very few cases—specifically: when the problem is sufficiently simple—it is possible to get an analytical expression for the posterior ...**

  Luckily enough, this is going to be our case here, in estimating the Bayesian VAR ...

- **In the vast majority of cases, however, it is not possible to get a simple, neat expression for the posterior ...**

  Then, only way to go is to simulate the posterior numerically, *via* algorithms such as Gibbs sampling, or Metropolis-Hastings—this is what we do in my other class on estimating DSGE models ...

But here we are lucky, and in estimating the Bayesian VAR we can use the mathematical results found in Uhlig (*Carnegie-Rochester*, 1998; *JME*, 2005) ...

Estimation of a Bayesian VAR

Given the reduced-form VAR

$$Y_t = B_{(1)} Y_{t-1} + B_{(2)} Y_{t-2} + \cdots + B_{(l)} Y_{t-l} + u_t, t = 1, \ldots, T,$$

which—remember the discussion at the very beginning of this course—you can rewrite by using matrix algebra as:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{u}$$

where

$$X_t = [Y'_{t-1}, Y'_{t-2}, \ldots, Y'_{t-l}]' \qquad \mathbf{Y} = [Y_1, \ldots, Y_T]'$$
$$\mathbf{X} = [X_1, \ldots, X_T]' \qquad \mathbf{u} = [u_1, \ldots, u_T]'$$
$$\mathbf{B} = [B_{(1)}, \ldots, B_{(l)}]'$$

and with the covariance matrix of the VAR's innovations being Σ, the maximum likelihood estimators for B and Σ are:

$$\hat{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$
$$\hat{\Sigma} = \frac{1}{T}(\mathbf{Y} - \mathbf{X}\hat{B})'(\mathbf{Y} - \mathbf{X}\hat{B})$$

These estimators will be a key input in getting the posterior distribution ...

The joint distribution for the B and Σ is postulated to belong to the so-called 'Normal-Wishart family', which means that

- **Σ is postulated to follow a Wishart distribution, and**
- **conditional on Σ, B is postulated to follow a Normal distribution ...**

A 'Normal–Wishart' distribution is parameterized by

- **a mean coefficient matrix $\bar{B}$,**
- **a positive definite mean covariance matrix *S*,**
- **a positive definite matrix *N*, and**
- **a degrees-of-freedom real number $v \geq 0$ to describe the uncertainty about (B, Σ) around $(\bar{B}, S)$**

The Normal–Wishart distribution specifies that

- **Σ^{-1} follows a Wishart distribution: $\mathscr{W}_m(S^{-1}/v, v)$, with $E[\Sigma^{-1}] = S^{-1}$, and that**
- **conditional on Σ, the coefficient matrix *B* in its columnwise vectorized form, vec(*B*); follows a Normal distribution ...**

Uhlig (*Econometric Theory*, 1994) shows that if the prior is characterized by the parameters

$$\bar{B}_0,\ N_0,\ S_0 \text{ and } v_0,$$

then the posterior is given by $\bar{B}_T,\ N_T,\ S_T$ and v_T**, where**

$$v_T = T + v_0,$$

$$N_T = N_0 + \mathbf{X}'\mathbf{X},$$

$$\bar{B}_T = N_T^{-1}(N_0\bar{B}_0 + \mathbf{X}'\mathbf{X}\hat{B}),$$

$$S_T = \frac{v_0}{v_T}S_0 + \frac{T}{v_T}\hat{\Sigma} + \frac{1}{v_T}(\hat{B} - \bar{B}_0)'N_0N_T^{-1}\mathbf{X}'\mathbf{X}(\hat{B} - \bar{B}_0).$$

Following Uhlig (1998, 2005), we use a weak prior (that is: we impose as little prior information on the data as possible):

We therefore set:

$$N_0 = 0;\ v_0 = 0;\ S_0 \text{ and } \bar{B}_0 \text{ arbitrary}$$

Then, the posterior is given by

$$\bar{B}_T = \hat{B}$$
$$S_T = \hat{\Sigma}$$
$$v_T = T$$
$$N_T = \mathbf{X}'\mathbf{X}$$

Notice how nice this posterior is:

- $\bar{B}_T$ **and S_T are nothing but the *MLE*—that is: the *OLS*—estimates, and**
- **v_T and N_T are just the sample length and the transposed matrix of the data times itself ...**

This means that in order to estimate a Bayesian VAR

- **we start by getting the OLS estimates;**
- **we then get the draws for the posterior of the covariance matrix, drawing from the Wishart distribution;**
- **finally, conditional on each single draw from the Wishart distribution for the covariance matrix, we draw from the Normal distribution for vec(B) ...**

This allows us to get the entire posterior distribution for the estimates of the VAR ...

Now let's do it in MATLAB ....

Here run the code `BayesianVAREstimation.m`

Imposing sign restrictions

Suppose we have estimated the Bayesian VAR for a vector of series Y_t, which means that we have (say) 10,000 draws from the posterior distribution for

- **the VAR's coefficients' matrices, B, and**
- **the VAR's covariance matrix, Σ ...**

Then, the next step is to impose the sign restrictions on the VAR, that is, for each draw of the VAR's covariance matrix, Σ, finding a structural impact matrix A_0 such that

(*i*) $A_0 A_0' = \Sigma$
(*ii*) A_0 **satisfies the pattern of signs we want to impose**

In general, we may also want to impose the signs on the IRFs, that is: at horizons greater than just on impact ...

Run the code `ImposingSignRestrictions.m` ..., and show how it works ...