Lecture 5: (S)VAR models: part 1

Geert Mesters (Universitat Pompeu Fabra, Barcelona GSE and VU Amsterdam)

Motivation

Introduction

- In this set of slides we introduce (structural) vector autoregressive models: (S)VAR
- Since Sims (1980) these models have been one of the main tools for empirical research in macroeconomics
- We do two things
    - We discuss inference for reduced-form VAR models
    - We discuss the usage of structural VAR models
- The next lecture discusses different identification methods that allow to go from the reduced form to the structural form

Running example

Uncertainty boom

- In recent years policy makers have become increasingly worried about the influence of uncertainty
- *Uncertainty is not just an important feature of the monetary policy landscape; it is the defining characteristic of that landscape*, Alan Greenspan

Uncertainty boom

- Since uncertainty is not observable there are different ways of measuring uncertainty: model based (e.g. volatility), newspaper scoring, disagreement among forecasters, VIX index, and more ...

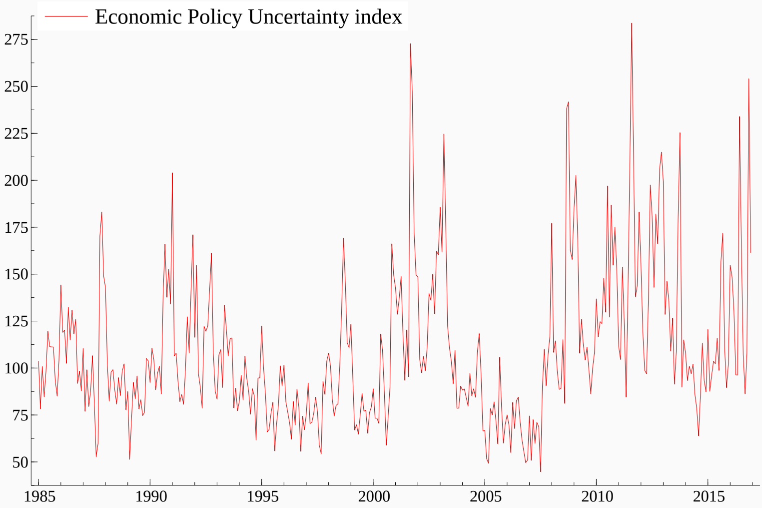
- This has stimulated a large literature that investigates the impact of differrent measures of uncertainty on macroeconomic variables

  see Bloom et al 2009, Baker et al 2015, Jurado et al 2015, Caldara et al 2016, Carriero et al 2017, Gorodnichenko & Ng 2018, Rossi & Sekhposyan 2015, Dibiasi et al 2018, Bloom et al 2018

Uncertainty boom

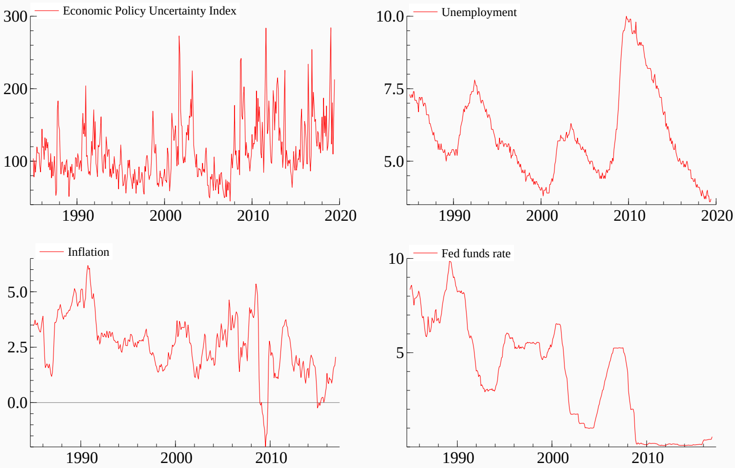
- A key question is how different measures of uncertainty interact with the macroeconomic variables of interest
- For instance: does policy uncertainty reduce the level of economic activity?
- In our running example
  - Measure uncertainty: Economic policy uncertainty index of Baker, Bloom, and Davis (2015)
  - Macro variables: inflation, Fed funds rate and unemployment

Economic policy uncertainty index



see https://www.policyuncertainty.com

Data series



Economic Policy Uncertainty Index
300
200
100
1990
2000
2010
2020
Unemployment
10.0
7.5
5.0
1990
2000
2010
2020
Inflation
5.0
2.5
0.0
1990
2000
2010
Fed funds rate
10
5
1990
2000
2010

The reduced form

Reduced form VAR

The reduced form VAR(p) model is defined as

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{A}_p \mathbf{y}_{t-p} + \boldsymbol{u}_t \qquad \boldsymbol{u}_t \sim WN(0, \boldsymbol{\Sigma}_u)$$

where

$$\mathbf{y}_t = \begin{bmatrix} y_{1,t} \\ \vdots \\ y_{K,t} \end{bmatrix} \qquad \boldsymbol{u}_t = \begin{bmatrix} u_{1,t} \\ \vdots \\ u_{K,t} \end{bmatrix}$$

$$\mathbf{A}_j = \begin{bmatrix} a_{11,j} & \ldots & a_{1K,j} \\ \vdots & \ddots & \vdots \\ a_{K1,j} & \ldots & a_{KK,j} \end{bmatrix} \qquad \boldsymbol{\Sigma}_u = \begin{bmatrix} \sigma_{11} & \ldots & \sigma_{1K} \\ \vdots & \ddots & \vdots \\ \sigma_{K1} & \ldots & \sigma_{KK} \end{bmatrix}$$

Simple example

To illustrate: take $K = 2$ and $p = 2$, to obtain

$$
\begin{aligned}
y_{1,t} &= a_{11,1} y_{1,t-1} + a_{12,1} y_{2,t-1} + a_{11,2} y_{1,t-2} + a_{12,2} y_{2,t-2} + u_{1,t} \\
y_{2,t} &= a_{21,1} y_{1,t-1} + a_{22,1} y_{2,t-1} + a_{21,2} y_{1,t-2} + a_{22,2} y_{2,t-2} + u_{2,t}
\end{aligned}
$$

- The VAR model stacks linear regressions
- Inference should be standard

Uses of VARs

The reduced form VAR model is useful for

- Describing dynamic relationships among multiple variables
- Out-of-sample forecasting (especially with appropriate parameter restrictions or shrinkage)
- Testing predictive relationships among the variables

Polynomial notation

The VAR(p) model can be compactly written as

$$\mathbf{A}(L)\mathbf{y}_t = \boldsymbol{u}_t \qquad \boldsymbol{u}_t \sim WN(0, \boldsymbol{\Sigma}_u)$$

where

$$\mathbf{A}(L) = \mathbf{I}_m - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L^p$$

The VAR is stable (stationary and causal) if

$$\det(\mathbf{I}_K - \mathbf{A}_1 z - \ldots - \mathbf{A}_p z^p) \neq 0 \qquad \forall \quad z \in \mathbb{C} \quad |z| \leq 1$$

Companion form

For later reference, we can always write the companion form

$$\mathbf{Y}_t = \mathbf{A}^c \mathbf{Y}_{t-1} + \boldsymbol{U}_t$$

where

$$\mathbf{Y}_t = \underbrace{\begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}}_{Kp \times 1} \qquad \mathbf{A}^c = \underbrace{\begin{bmatrix} \mathbf{A}_1 & \dots & & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_K & 0 & & 0 & 0 \\ 0 & \mathbf{I}_K & & & \vdots \\ & & \ddots & 0 & \\ 0 & & 0 & \mathbf{I}_K & 0 \end{bmatrix}}_{Kp \times Kp}$$

- The VAR is stable if eigenvalues of $\mathbf{A}^c$ are strictly less then 1

Moving average representation

The VAR model

$$\mathbf{A}(L)\mathbf{y}_t = \boldsymbol{u}_t$$

can be rewritten in moving average form as

$$\begin{aligned} \mathbf{y}_t &= \mathbf{A}(L)^{-1}\boldsymbol{u}_t \\ &= \textstyle\sum_{i=0}^{\infty} \mathbf{\Phi}_i \boldsymbol{u}_{t-i} \end{aligned}$$

where the matrices $\mathbf{\Phi}_i$ can be found recursively from

$$\mathbf{\Phi}_0 = \mathbf{I}_K, \qquad \mathbf{\Phi}_i = \sum_{j=1}^{i} \mathbf{\Phi}_{i-j}\mathbf{A}_j, \quad i = 1, 2, \ldots$$

Estimation

Consider the VAR(p) model

$$\mathbf{y}_t = [\mathbf{A}_1, \ldots, \mathbf{A}_p]\mathbf{Z}_{t-1} + \boldsymbol{u}_t$$

where $\mathbf{Z}_t = (\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})'$.

The least squares estimator is given by

$$\begin{aligned}\hat{\mathbf{A}} = [\hat{\mathbf{A}}_1, \ldots, \hat{\mathbf{A}}_p] &= \left(\sum_{t=1}^{T} \mathbf{y}_t \mathbf{Z}'_{t-1}\right) \left(\sum_{t=1}^{T} \mathbf{Z}_{t-1}\mathbf{Z}'_{t-1}\right)^{-1} \\ &= \mathbf{Y}\mathbf{Z}'(\mathbf{Z}\mathbf{Z}')^{-1}\end{aligned}$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_T]$ and $\mathbf{Z} = [\mathbf{Z}_0, \ldots, \mathbf{Z}_{T-1}]$

A word on theory

- Assume that $\boldsymbol{u}_t \sim IID(0, \boldsymbol{\Sigma}_u)$ and $\boldsymbol{u}_t$ has finite fourth moments[1]

When we stack the columns of $\mathbf{A}$, e.g. $\boldsymbol{\alpha} = \text{vec}(\mathbf{A})$, we can show

$$\sqrt{T}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}})$$

where

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}} = \boldsymbol{\Sigma}_z^{-1} \otimes \boldsymbol{\Sigma}_u \qquad \text{with} \qquad \frac{1}{T}\sum_{t=1}^{T} \mathbf{Z}_{t-1}\mathbf{Z}'_{t-1} \xrightarrow{p} \boldsymbol{\Sigma}_z$$

[1] This assumption can be relaxed to allow for serial correlation, then we need to use HAC standard errors.

Estimation

The residuals are given by

$$\hat{\mathbf{U}} = \mathbf{Y} - \mathbf{Z}\hat{\mathbf{A}}$$

which allows to estimate the variance matrix

$$\hat{\mathbf{\Sigma}}_u = \hat{\mathbf{U}}\hat{\mathbf{U}}' / (T - Kp)$$

which allows to estimate the asymptotic variance

$$\hat{\mathbf{\Sigma}}_{\hat{\alpha}} = \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_{t-1}\mathbf{Z}'_{t-1} \right)^{-1} \otimes \hat{\mathbf{\Sigma}}_u$$

Estimates Uncertainty

- Consider a bi-variate VAR ($K = 2$) for policy uncertainty and unemployment with $p = 4$ lags

$\hat{\mathbf{A}}_1$	Epu	Une
Epu	0.62	0.04
Une	0.04	0.98

$\hat{\mathbf{A}}_2$	Epu	Une
Epu	-0.05	0.08
Une	0.02	0.16

$\hat{\mathbf{A}}_3$	Epu	Une
Epu	-0.03	-0.00
Une	-0.02	0.05

$\hat{\mathbf{A}}_4$	Epu	Une
Epu	0.14	-0.09
Une	0.01	-0.22

- Overall, there seems to be some interaction :-)

Estimates Uncertainty

$\hat{\mathbf{A}}_1$	Unc	Une	Inf	Fed
Unc	0.60	0.04	-0.01	-0.03
Une	0.04	0.95	-0.02	-0.10
Inf	-0.09	0.10	1.40	0.16
Fed	-0.13	-0.10	0.00	1.36

$\hat{\mathbf{A}}_2$	Unc	Une	Inf	Fed
Unc	-0.06	0.09	0.06	-0.04
Une	0.01	0.15	0.02	0.14
Inf	-0.04	-0.19	-0.66	-0.17
Fed	0.03	0.03	0.00	-0.35

The reduced form VAR coefficients are hard to interpret

Prediction from known VAR models

Define the MSE optimal prediction by

$$\mathbf{y}_{T+h|T} = \mathbb{E}(\mathbf{y}_{T+h}|\mathbf{y}_T, \ldots, \mathbf{y}_1)$$

If the errors are not serially correlated we can compute this quantity recursively

$$\mathbf{y}_{T+h|T} = \mathbf{A}_1\mathbf{y}_{T+h-1|T} + \ldots + \mathbf{A}_p\mathbf{y}_{T+h-p|T},$$

where $\mathbf{y}_{T+j|T} = \mathbf{y}_{T+j}$ for $j \leq 0$

Prediction error variance

The prediction errors are given by

$$\mathbf{y}_{T+h} - \mathbf{y}_{T+h|T} = \boldsymbol{u}_{T+h} + \boldsymbol{\Phi}_1 \boldsymbol{u}_{T+h-1} + \ldots + \boldsymbol{\Phi}_{h-1} \boldsymbol{u}_{T+1}$$

where $\boldsymbol{\Phi}_i$ are the coefficients from the moving average form: $\mathbf{y}_t = \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i \boldsymbol{u}_{t-i}$.

The prediction error covariance or mean-squared prediction error (MSPE) matrix is

$$\boldsymbol{\Sigma}_y(h) = \mathbb{E}(\mathbf{y}_{T+h} - \mathbf{y}_{T+h|T})(\mathbf{y}_{T+h} - \mathbf{y}_{T+h|T})' = \sum_{j=0}^{h-1} \boldsymbol{\Phi}_j \boldsymbol{\Sigma}_u \boldsymbol{\Phi}_j'$$

Granger causality

- Granger (1969) calls a variable $y_{2,t}$ causal for a variable $y_{1,t}$ if the information in past and present values of $y_{2,t}$ helps reduce in expectation the prediction error for $y_{1,t}$: Granger causality

Granger causality

Let $\mathbf{\Omega}_t$ denote all information relevant for predicting $y_{1,t}$ denote by $y_{1,t+h|\Omega_t}$ and the corresponding MSPE $\sigma^2_{y_1}(h|\Omega_t)$

The variable $y_{2,t}$ Granger causes $y_{1,t}$ if

$$\sigma^2_{y_1}(h|\Omega_t) < \sigma^2_{y_1}(h|\Omega_t \backslash \{y_{2,s}, s \leq t\}) \qquad \text{for at least one } h$$

where $\Omega_t \backslash \{y_{2,s}, s \leq t\}$ denotes all information except the current and past values of $y_{2,t}$

Granger causality

When we have two variables it is easy to test for Granger causality

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \sum_{j=1}^{p} \begin{bmatrix} a_{11,j} & a_{12,j} \\ a_{21,j} & a_{22,j} \end{bmatrix} \begin{bmatrix} y_{1,t-j} \\ y_{2,t-j} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$

- Then $y_{2,t}$ is not Granger causal for $y_{1,t}$ if and only if $a_{12,j} = 0$ for all $i = 1, 2, \ldots, p$

Some comments

- Testing Granger causality is more difficult with multiple variables, see Al Sadoon (2014, 2019) for an excellent treatment
- Granger causality does not speak to instantaneous relationships
- It is often more convenient to think about Granger causality as predictive ability

Granger causality for Uncertainty

- Consider a bi-variate VAR ($K = 2$) for policy uncertainty and unemployment with $p = 4$ lags

	F-stat	p-value
Epu	6.20	0.18
Une	11.40	0.02

- Granger causality tests indicate that economic policy uncertainty Granger causes unemployment

Lag length selection

The final important practical issue is lag length selection, we discuss

- Sequential testing procedures
- Information criteria
- Recursive Mean-Squared Prediction Error rankings

Sequential testing procedures

Sequential testing procedures come in two forms

- Top down

  Sequentially test $H_0 : \mathbf{A}_{p^{max}} = 0$, $H_0 : \mathbf{A}_{p^{max}-1} = 0$, ...

  Test can be implemented using standard likelihood ratio statistics

- Bottom up

  Sequentially test $H_0 : \mathbb{E}(\boldsymbol{u}_t \boldsymbol{u}_{t-i}) = 0$ for all i, increase lag length until you do not reject

  Test can be implemented using standard Portmanteau statistic or LM statistic

Sequential testing procedures

Sequential testing procedures come in two forms

- Top down

  Sequentially test $H_0 : \mathbf{A}_{p^{max}} = 0$, $H_0 : \mathbf{A}_{p^{max}-1} = 0$, ...

  Test can be implemented using standard likelihood ratio statistics

- Bottom up

  Sequentially test $H_0 : \mathbb{E}(\boldsymbol{u}_t \boldsymbol{u}_{t-i}) = 0$ for all i, increase lag length until you do not reject

  Test can be implemented using standard Portmanteau statistic or LM statistic

The finite sample properties of these procedures are not great

Gonzalo and Pitarakis (2002) and Kilian and Ivanov (2005)

Information criteria

Information criteria take the general form

$$C(m) = \log\det(\tilde{\boldsymbol{\Sigma}}_u(m)) + c_T\psi(m)$$

where

- $\tilde{\boldsymbol{\Sigma}}_u(m) = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{u}}_t\hat{\boldsymbol{u}}_t'$ for the model with m lags
- c_T some function of the sample size
- $\psi(m)$ penalizes large lag choices, often
  $\psi(m) =$ number of parameters

Optimal lag length corresponds to m that minimizes $C(m)$

Information criteria

Examples of Information criteria

- Akaike Information Criterion (AIC)

$$AIC(m) = \log\det(\tilde{\boldsymbol{\Sigma}}_u(m)) + \frac{2}{T} mK^2$$

- Hannan-Quinn Criterion (HQC)

$$HQC(m) = \log\det(\tilde{\boldsymbol{\Sigma}}_u(m)) + \frac{2\log\log T}{T} mK^2$$

- Schwarz Information Criterion (SIC) (also known as BIC)

$$SIC(m) = \log\det(\tilde{\boldsymbol{\Sigma}}_u(m)) + \frac{\log T}{T} mK^2$$

Information criteria

Examples of Information criteria

- Akaike Information Criterion (AIC)

$$AIC(m) = \log\det(\tilde{\mathbf{\Sigma}}_u(m)) + \frac{2}{T} mK^2$$

- Hannan-Quinn Criterion (HQC)

$$HQC(m) = \log\det(\tilde{\mathbf{\Sigma}}_u(m)) + \frac{2\log\log T}{T} mK^2$$

- Schwarz Information Criterion (SIC) (also known as BIC)

$$SIC(m) = \log\det(\tilde{\mathbf{\Sigma}}_u(m)) + \frac{\log T}{T} mK^2$$

Finite sample evidence suggests AIC performs reasonably well for recovering the true lag order

Recursive Mean-Squared Prediction Error rankings

Recursive Mean-Squared Prediction Error rankings are based on out-of-sample predictive ability

- Split the sample in two parts and perform an out-of-sample forecasting study (see Lecture 4)
- Do this for all possible models (e.g. different lag choices) and select the model with the lowest prediction error

Lag length selection for Uncertainty

- Consider a bi-variate VAR ($K = 2$) for policy uncertainty and unemployment

- AIC selects $p = 10$
- $HQIC$ selects $p = 9$
- SIC selects $p = 1$

Clearly the SIC penalizes heavily!!!

Some comments

- Lag length selection is difficult and different choices can lead to vastly different conclusions, see Hamilton & Herrera (2004)
- Model selection issue, see Leep & Potscher (2005)
- Best choice depends on the end goal of the analyses, e.g. forecasting or structural analysis

Summary

We have reviewed inference for reduced form VAR models

- Estimation
- Forecasting
- Granger Causality
- Lag length selection

Summary

We have reviewed inference for reduced form VAR models

- Estimation
- Forecasting
- Granger Causality
- Lag length selection

Important: All parameters in the reduced form VAR model are identified and can be estimated given the observed time series!!!

Structural VARs

Structural VAR

The structural VAR(p) model is defined as

$$\mathbf{B}_0 \mathbf{y}_t = \mathbf{B}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{B}_p \mathbf{y}_{t-p} + \boldsymbol{w}_t$$

where

- $\mathbf{B}_0, \ldots, \mathbf{B}_p$ are $K \times K$ coefficient matrices
- $\boldsymbol{w}_t$ is the $K \times 1$ vector of structural shocks
  - in the sense that the elements of $\boldsymbol{w}_t$ are mutually uncorrelated and have clear interpretations in terms of an underlying economic model, e.g. $w_{1,t}$ is a monetary policy shock; see Ramey (2016) for a nice discussion about what is a shock
  - without loss of generality we normalize $\boldsymbol{w}_t \sim IID(0, \mathbf{I}_K)$

Identification problem

Not all parameters of the structural VAR are identified, to see this consider the bi-variate example

$$
\begin{aligned}
b_{11,0}y_{1,t} &= -b_{12,0}y_{2,t} + b_{11,1}y_{1,t-1} + b_{12,1}y_{2,t-1} + w_{1,t} \\
b_{22,0}y_{2,t} &= -b_{21,0}y_{1,t} + b_{21,1}y_{1,t-1} + b_{22,1}y_{2,t-1} + w_{2,t}
\end{aligned}
$$

- This is the same simultaneous equations problem as in Lecture 2

Identification problem

In general,

$$\mathbf{y}_t = \underbrace{\mathbf{B}_0^{-1}\mathbf{B}_1}_{\mathbf{A}_1}\mathbf{y}_{t-1} + \ldots + \underbrace{\mathbf{B}_0^{-1}\mathbf{B}_p}_{\mathbf{A}_p}\mathbf{y}_{t-p} + \underbrace{\mathbf{B}_0^{-1}\boldsymbol{w}_t}_{\boldsymbol{u}_t}$$

Which implies that

$$\mathbf{B}_0\boldsymbol{u}_t = \boldsymbol{w}_t$$

- $\mathbf{B}_0$ is required to go back and forth between the reduced form and the structural form
- But $\mathbf{B}_0$ cannot be consistency estimated without further information, e.g. restrictions or instruments

Structural VAR

- In this set of slides we assume that $\mathbf{B}_0$ is known[2]
- Given $\mathbf{B}_0$ we highlight the uses of Structural VARs for
  - Computing impulse responses
  - Computing forecast error variance decompositions
  - Historical decompositions

[2] The next lecture discusses several methods for identifying $\mathbf{B}_0$

Impulse responses

Often we like to answer questions like

- What is the effect of an unexpected change in the interest rate on aggregate macroeconomic variables?
- What is the effect of an unexpected oil price change on output growth?

We are interested in the dynamic causal effect or impulse response

$$\frac{\partial \mathbf{y}_{t+i}}{\partial \boldsymbol{w}_t'} = \boldsymbol{\Theta}_i \qquad i = 0, 1, 2, 3, \ldots$$

where $\boldsymbol{w}_t$ is the vector of structural shocks and $\boldsymbol{\Theta}_i$ is a $K \times K$ matrix

Impulse responses

The impulse response matrix

$$\frac{\partial \mathbf{y}_{t+i}}{\partial w_t'} = \mathbf{\Theta}_i \qquad i = 0, 1, 2, 3, \dots$$

has elements

$$\theta_{lk,i} = \frac{\partial y_{l,t+i}}{\partial w_{k,t}}$$

- We typically plot $\theta_{lk,i}$ for $i = 0, 1, 2, \dots, H$
- Within each SVAR we may construct K^2 impulse responses

Impulse responses

- The impulse responses can be easily computed from the moving average representation

Recall that for the reduced form VAR

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \mathbf{\Phi}_i \boldsymbol{u}_{t-i}$$

Since, $\boldsymbol{u}_t = \mathbf{B}_0^{-1} \boldsymbol{w}_t$ we have that

$$\begin{aligned} \mathbf{y}_t &= \textstyle\sum_{i=0}^{\infty} \mathbf{\Phi}_i \boldsymbol{u}_{t-i} \\ &= \textstyle\sum_{i=0}^{\infty} \mathbf{\Phi}_i \mathbf{B}_0^{-1} \boldsymbol{w}_{t-i} \\ &= \textstyle\sum_{i=0}^{\infty} \mathbf{\Theta}_i \boldsymbol{w}_{t-i} \end{aligned}$$

where $\mathbf{\Theta}_i = \mathbf{\Phi}_i \mathbf{B}_0^{-1}$. It is easy to see that

$$\frac{\partial \mathbf{y}_{t+i}}{\partial w_t'} = \mathbf{\Theta}_i$$

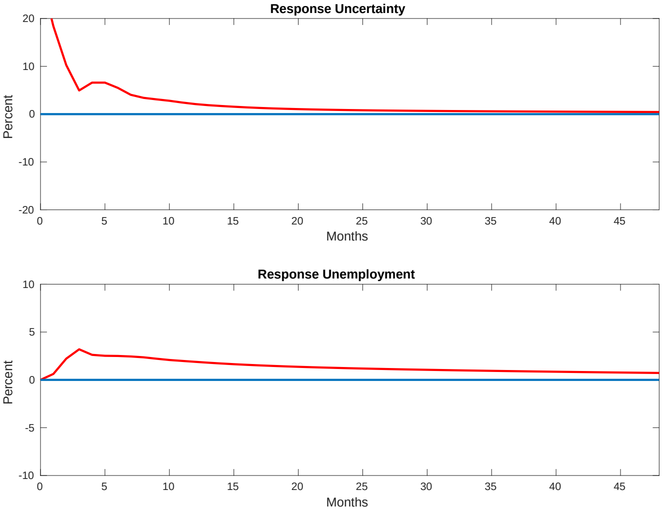
Impulse responses for Uncertainty

To illustrate impulse responses for the bi-variate SVAR with uncertainty and unemployment

- We need to solve the identification problem, here a very simplistic solution:
  - Assume that uncertainty does not contemporaneously affect unemployment, e.g. $b_{21,0} = 0$
  - Note that the other elements of $\mathbf{B}_0$ can be obtained from $\mathbf{\Sigma}_u$, e.g. $\mathbf{\Sigma}_u = \mathbf{B}_0^{-1}\mathbf{B}_0^{-1\prime}$ [3]
- Take lag length $p = 12$

[3] In the next lecture we will label this procedure as imposing short run restrictions and discuss many alternative approaches.

Impulse responses for 1SD deviation shock to Uncertainty

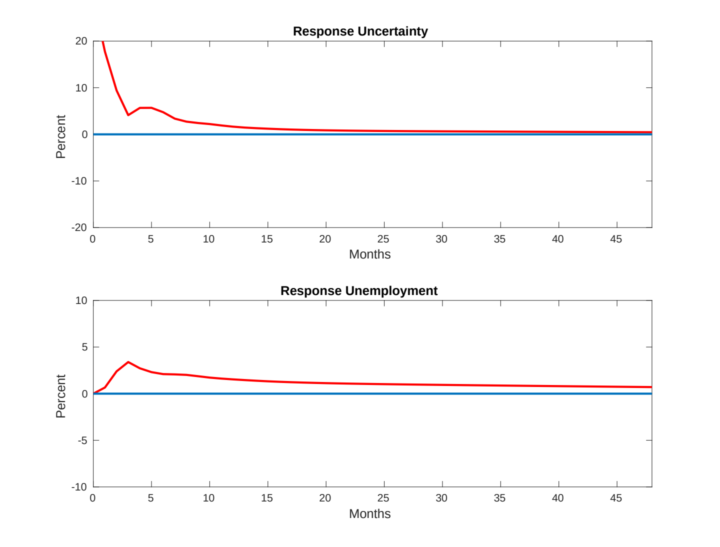


Some words of caution

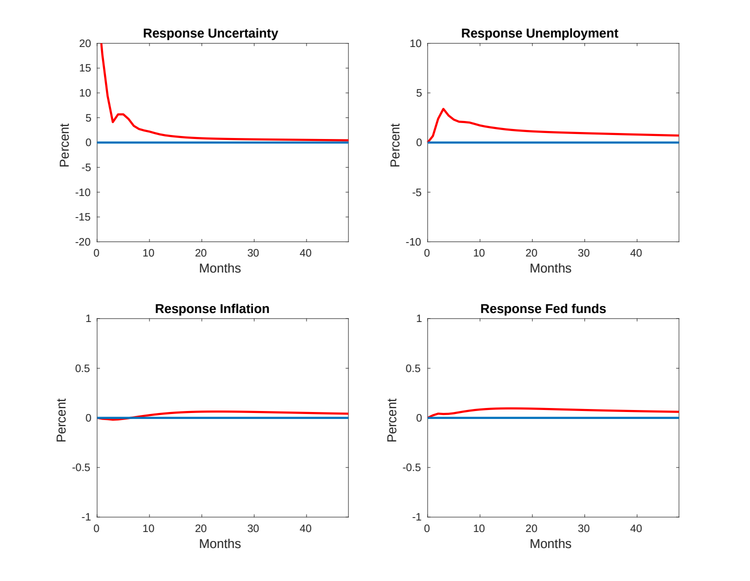
Issues

- My identification strategy boils down to controlling for some lags of uncertainty and inflation: this is not very strong
  - This explains why the point effects seem so large !!!
  - A more careful identification strategy will yield a smaller effect
- Next, I repeat the exercise using the 4 variable SVAR (same identification)

Impulse responses for 1SD deviation shock to Uncertainty



Impulse responses for 1SD deviation shock to Uncertainty



Forecast error variance decompositions

- A second practically important question that a structural VAR model can answer is how much of the forecast error variance or prediction mean squared error (MSPE) of y_{t+h} at horizon $h = 0, 1, \ldots, H$ is accounted for by each structural shock $w_{k,t}$, $k = 1, \ldots, K$.

Forecast error variance decompositions

Recall from the reduced form that

$$
\begin{aligned}
\mathbf{y}_{t+h} - \mathbf{y}_{t+h|t} &= \boldsymbol{u}_{t+h} + \boldsymbol{\Phi}_1 \boldsymbol{u}_{t+h-1} + \ldots + \boldsymbol{\Phi}_{h-1} \boldsymbol{u}_{t+1} \\
&= \textstyle\sum_{i=0}^{h-1} \boldsymbol{\Phi}_i \boldsymbol{u}_{t+h-i} \\
&= \textstyle\sum_{i=0}^{h-1} \boldsymbol{\Phi}_i \mathbf{B}_0^{-1} \boldsymbol{w}_{t+h-i} \\
&= \textstyle\sum_{i=0}^{h-1} \boldsymbol{\Theta}_i \boldsymbol{w}_{t+h-i}
\end{aligned}
$$

From this is follows that

$$
\begin{aligned}
MSPE(h) &= \mathbb{E}(\mathbf{y}_{t+h} - \mathbf{y}_{t+h|t})(\mathbf{y}_{t+h} - \mathbf{y}_{t+h|t})' \\
&= \textstyle\sum_{i=0}^{h-1} \boldsymbol{\Theta}_i \boldsymbol{\Theta}_i'
\end{aligned}
$$

This decomposes the forecast error variance into components that can be attributed to different shocks

Forecast error variance decompositions

Let $\theta_{kj,h}$ be k, j element of $\mathbf{\Theta}_h$, we have that

$$MSPE_j^k(h) = \theta_{kj,0} + \ldots + \theta_{kj,h-1}$$

Summing over j gives total contribution of variable k

$$MSPE^k(h) = \sum_{j=1}^{K} MSPE_j^k(h)$$

and the relative contribution is given by

$$\frac{MSPE_j^k(h)}{MSPE^k(h)} \qquad j, k = 1, \ldots, K$$

In words: contribution of variable j to forecast error variance of variable k at horizon h

Forecast error variance decomposition uncertainty

	Epu	Unc	Inf	Fed
Epu	99.67	0.21	0.06	0.03
Unc	5.89	93.62	0.08	0.39
Inf	0.09	0.11	99.62	0.17
Fed	7.21	0.40	0.58	91.79

Historical decompositions

- Structural forecast error variance decompositions and structural impulse response functions describe the average movements in the data
- They represent unconditional expectations
- Sometimes we are interested instead in quantifying how much a given structural shock explains of the historically observed fluctuations in the VAR variables
- In other words, we would like to know the cumulative effect of a given structural shock on each variable at every given point in time

Historical decompositions

As an example

- We may not be interested in the average contribution of monetary policy shocks to the variability of real GDP growth over the last decades
- but rather in the question of whether monetary policy shocks caused the 1982 recession

Historical decompositions

At any point in time we may decompose

$$y_t = \sum_{s=0}^{t-1} \boldsymbol{\Theta}_s \boldsymbol{w}_{t-s} + \sum_{s=t}^{\infty} \boldsymbol{\Theta}_s \boldsymbol{w}_{t-s}$$

The shocks from before the sample starts are unknown, so

$$y_t \approx \sum_{s=0}^{t-1} \boldsymbol{\Theta}_s \boldsymbol{w}_{t-s} \qquad \rightarrow \qquad \hat{y}_t = \sum_{s=0}^{t-1} \boldsymbol{\Theta}_s \boldsymbol{w}_{t-s}$$

Now the idea is to simply decompose each $\hat{y}_{i,t}$ into contributions from different structural shocks

Historical decompositions

We may decompose

$$\hat{y}_{i,t}^{(1)} = \sum_{s=0}^{t-1} \theta_{i1,s} w_{1,t-s}$$

$$\hat{y}_{i,t}^{(2)} = \sum_{s=0}^{t-1} \theta_{i2,s} w_{2,t-s}$$

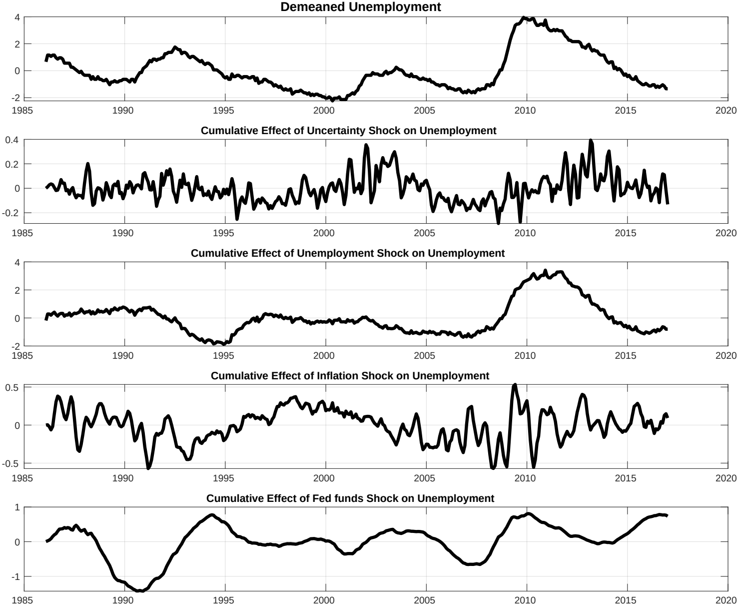
$$\vdots$$

$$\hat{y}_{i,t}^{(K)} = \sum_{s=0}^{t-1} \theta_{iK,t-s} w_{K,t-s}$$

where we note that $\hat{y}_{i,t} = \sum_{j=1}^{K} \hat{y}_{i,t}^{(j)}$, by construction.

- The historical decomposition plots $\hat{y}_{i,t}^{(j)}$ for different variables i and structural shocks j

Historical decomposition unemployment



Some comments

We have reviewed three methods for summarizing SVAR results

- Impulse responses
- Forecast error variance decompositions
- Historical decompositions

References & Material

Reference & Material

- References:
  Structural Vector Autoregressive Analysis' by Lutz Kilian and Helmut Lutkepohl, Cambridge University Press, 2017.