AEA Continuing Education Course

Time Series Econometrics

Lecture 7

James H. Stock

January 7, 2010

Recent Developments in Structural Vector Autoregression (SVAR) Modeling

Outline

1) VARs, SVARs, and the Identification Problem
2) Identification by Short Run Restrictions
3) Identification by Long Run Restrictions
4) The Quasi-Experiment Approach
5) Direct Estimation of Shocks from High Frequency Data
6) Identification by Sign Restrictions
7) Identification from Heteroskedasticity
8) DSGE Priors
9) Identification from Regional/Multicountry Restrictions
10) Inference: Challenges and Recently Developed Tools

1) VARs, SVARs, and the Identification Problem

A classic question in empirical macroeconomics: what is the effect of a policy intervention (interest rate increase, fiscal stimulus) on macroeconomic aggregates of interest – output, inflation, etc?

Let Y_t be a vector of macro time series, and let ε_t^r denote an unanticipated monetary policy intervention. We want to know the ***dynamic causal effect*** of ε_t^r on Y_t:

$$\frac{\partial Y_{t+h}}{\partial \varepsilon_t^r}, \; h = 1, 2, 3, \ldots.$$

where the partial derivative holds all other interventions constant. In macro, this dynamic causal effect is called the ***impulse response function*** (***IRF***) of Y_t to the "shock" (unexpected intervention) ε_t^r.

The challenge is to estimate $\left\{\frac{\partial Y_{t+h}}{\partial \varepsilon_t^r}\right\}$ from observational macro data.

Two conceptual approaches to estimating dynamic causal effects (*IRF*)

1) Structural model (Cowles Commission)
   a) tightly parameterized (many restrictions): FMP,…, DSGE
   b) structural vector autoregressions (SVARs)
2) Quasi-Experiments

The identification problem

Consider a 2*-variable system of linear simultaneous equations:* Let ε_{1t} and ε_{2t} be uncorrelated structural shocks, where $E(\varepsilon_t|Y_{t-1}, Y_{t-2},\ldots) = 0$:

$$Y_{1t} = \boldsymbol{B_{0,12}Y_{2t}} + B_{1,12}Y_{2t-1} + \ldots + B_{p,12}Y_{2t-p} + B_{1,11}Y_{1t-1} + \ldots + B_{p,11}Y_{1t-p} + \varepsilon_{1t}$$

$$Y_{2t} = \boldsymbol{B_{0,21}Y_{1t}} + B_{1,21}Y_{1t-1} + \ldots + B_{p,21}Y_{1t-p} + B_{1,22}Y_{2t-1} + \ldots + B_{p,22}Y_{2t-p} + \varepsilon_{2t}$$

Given the B's, we could compute structural impulse responses from this system (formulas below). But the coefficients of this system are not identified. To identify them, we either need an instrument Z_t, or a restriction on the parameters.

VAR background and notation:

$$Y_{1t} = B_{0,12}Y_{2t} + B_{1,12}Y_{2t-1} + \ldots + B_{p,12}Y_{2t-p} + B_{1,11}Y_{1t-1} + \ldots + B_{p,11}Y_{1t-p} + \varepsilon_{1t}$$

$$Y_{2t} = B_{0,21}Y_{1t} + B_{1,21}Y_{1t-1} + \ldots + B_{p,21}Y_{1t-p} + B_{1,22}Y_{2t-1} + \ldots + B_{p,22}Y_{2t-p} + \varepsilon_{2t}$$

This simultaneous equations system can be written, $B(L)Y_t = \varepsilon_t$, where

$$B(L) = B_0 - B_1L - B_2L^2 - \ldots - B_pL^p$$

and in general B_0 is not diagonal. ε_t are the ***structural shocks***.

The system $B(L)Y_t = \varepsilon_t$ is called a ***structural VAR*** (***SVAR***).

This SVAR has a reduced form (Sims (1980)), which is identified:

Reduced form VAR(p): $Y_t = A_1Y_{t-1} + \ldots + A_pY_{t-p} + u_t$

or $A(L)Y_t = u_t$,

where $A(L) = I - A_1L - A_2L^2 - \ldots - A_pL^p$

innovations: $u_t = Y_t - \text{Proj}(Y_t|Y_{t-1},\ldots, Y_{t-p})$

$Eu_tu_t' = \Sigma_u$

Reduced form to structure:

Suppose: (i) A(L) is finite order p (known or knowable)

(ii) u_t spans the space of structural shocks ε_t, that is, $\varepsilon_t = Ru_t$, where R is square (equivalently, Y_t is linear in the structural shocks & the model is invertible)

(iii) A(L), Σ_u , and R are time-invariant, e.g. A(L) is invariant to policy changes over the relevant period

Then IRFs can be obtained from the SVAR, $R\text{A(L)}Y_t = Ru_t$ or $\text{B(L)}Y_t = \varepsilon_t$:

SVAR: $\text{B(L)}Y_t = \varepsilon_t$,

where $\text{B(L)} = R\text{A(L)}$ and $Ru_t = \varepsilon_t$

Reduced form VAR: $\text{A(L)}Y_t = u_t$

MA representation: $Y_t = D(\text{L})\varepsilon_t$, $D(\text{L}) = \text{B(L)}^{-1} = \text{A(L)}^{-1}R^{-1}$

Impulse response: $\dfrac{\partial Y_{t+h}}{\partial \varepsilon_t} = D_h$

Summary of VAR and SVAR notation

Reduced form VAR	Structural VAR
$\mathrm{A}(\mathrm{L})Y_t = u_t$	$\mathrm{B}(\mathrm{L})Y_t = \varepsilon_t$
$Y_t = \mathrm{A}(\mathrm{L})^{-1}u_t = \mathrm{C}(\mathrm{L})u_t$	$Y_t = \mathrm{B}(\mathrm{L})^{-1}\varepsilon_t = \mathrm{D}(\mathrm{L})\varepsilon_t$
$\mathrm{A}(\mathrm{L}) = I - A_1\mathrm{L} - A_2\mathrm{L}^2 - \ldots - A_p\mathrm{L}^p$	$\mathrm{B}(\mathrm{L}) = B_0 - B_1\mathrm{L} - B_2\mathrm{L}^2 - \ldots - B_p\mathrm{L}^p$
$Eu_tu_t' = \Sigma_u$ (unrestricted)	$E\varepsilon_t\varepsilon_t' = \Sigma_\varepsilon = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_k^2 \end{pmatrix}$

$$Ru_t = \varepsilon_t$$

$$B(\mathrm{L}) = RA(\mathrm{L}) \quad (B_0 = R)$$

$$\mathrm{D}(\mathrm{L}) = \mathrm{C}(\mathrm{L})R^{-1}$$

- Note the assumption that the structural shocks are uncorrelated
- D(L) is the structural IRF of Y_t w.r.t. ε_t.
- structural forecast error variance decompositions are computed from D(L) and Σ_ε

Identification of *R* and identification of shocks

Two takes on identification:

1. ***Identification of R.*** In population, we can know A(L). If we can identify *R*, we can obtain the SVAR coefficients, B(L) = *R*A(L).
2. ***Identification of shocks.*** If you knew (or could estimate) one of the shocks, you could estimate the structural IRF of *Y* w.r.t. that shock. Partition Y_t into a policy variable r_t and all other variables:

$$Y_t = \begin{pmatrix} {}^{(k-1\times 1)} \\ X_t \\ {}^{(1\times 1)} \\ r_t \end{pmatrix}, \; u_t = \begin{pmatrix} u_t^X \\ u_t^r \end{pmatrix}, \; \varepsilon_t = \begin{pmatrix} \varepsilon_t^X \\ \varepsilon_t^r \end{pmatrix},$$

The IRF/MA form is $Y_t = \mathrm{D}(\mathrm{L})\varepsilon_t$, or

$$Y_t = \begin{pmatrix} D_{YX}(L) & D_{Yr}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_t^X \\ \varepsilon_t^r \end{pmatrix} = D_{Yr}(\mathrm{L})\varepsilon_t^r + v_t,$$

where $v_t = D_{YX}(\mathrm{L})\varepsilon_t^X$. Because $E\varepsilon_t^r v_t = 0$, the IRF of Y_t w.r.t. ε_t^r, $D_{Yr}(\mathrm{L})$ is identified by the population OLS regression of Y_t onto ε_t^r.

A word on "invertibility":

Recall the SVAR assumption:

(ii) u_t spans the space of structural shocks ε_t, that is, $\varepsilon_t = Ru_t$, where R is square

- This is often called the assumption of invertibility: the VAR can be inverted to span the space of structural shocks. If there are more structural shocks than u_t's, then condition (ii) will not hold.
- One response is to add more variables so that u_t spans ε_t. This response is an important motivation of the FAVAR approach, which will be discussed in Lecture 12.
- If agents see future shocks, invertibility fails. Or, does the definition of shock just become more subtle (an expectations shock)?
- See Lippi and Reichlin (1993, 1994), Sims and Zha (2006b), Fernandez-Villaverde, Rubio-Ramirez, Sargent, and Watson (2007), and Hansen and Sargent (2007).

This talk

- Early promise of SVARs and the critiques of the 1990s
- Survey of new ideas about how to tackle the identification problem
- Issues of inference, new and old, including some tools

Comments and references

- This is a mature literature – I will survey the developments in the past 10 years or so on identification and inference, focusing on the econometric issues
- Some general background references:

  Christiano, Eichenbaum, and Evans (1999)

  Lütkepohl (2005)

  Stock and Watson (2001)

  Watson (1994)

Outline

1) VARs, SVARs, and the Identification Problem
2) **Identification by Short Run Restrictions**
3) Identification by Long Run Restrictions
4) The Quasi-Experiment Approach
5) Direct Estimation of Shocks from High Frequency Data
6) Identification by Sign Restrictions
7) Identification from Heteroskedasticity
8) DSGE Priors
9) Identification from Regional/Multicountry Restrictions
10) Inference: Challenges and Recently Developed Tools

2) Identification by Short Run Restrictions

Overview: the traditional SVAR identification approach

Bernanke (1986), Blanchard and Watson (1986), Sims (1986)

(a) System identification. In general, the SVAR is fully identified if

$$R\Sigma_u R' = \Sigma_\varepsilon \qquad (1)$$

can be solved for the unknown elements of R and Σ_ε..

- There are $k(k+1)/2$ distinct equations in (1), so the order condition says that you can estimate (at most) $k(k+1)/2$ parameters. If we set $\Sigma_\varepsilon = I$ (just a normalization), it is clear that we need $k^2 - k(k+1)/2 = k(k-1)/2$ restrictions on R.
- If $k = 2$, then $k(k-1)/2 = 1$, which is delivered by imposing a single restriction (commonly, that R is lower or upper triangular).
- This ignores rank conditions, which matter!

- This description of identification is via method of moments (equation (1)), however identification can equally be described via IV, e.g. see Blanchard and Watson (1986).

(b) Partial identification. Many applications now take a limited information approach, in which only a row of R is identified. Partition $\varepsilon_t = Ru_t$, and partition Y_t so that:

$$\begin{pmatrix} \varepsilon_t^X \\ \varepsilon_t^r \end{pmatrix} = \begin{pmatrix} R_{XX} & R_{Xr} \\ R_{rX} & R_{rr} \end{pmatrix} \begin{pmatrix} u_t^X \\ u_t^r \end{pmatrix} \qquad (2)$$

If R_{rX} and R_{rr} are identified, then (in population) ε_t^r can be computed using just the final row of (2), and D_{Yr}(L) can be computed by the regression of Y_t on $\varepsilon_t^r, \varepsilon_{t-1}^r, \ldots$ as discussed above.

The "fast-r-slow" scheme. Almost all short-run restriction applications can be written as "fast-r-slow." Following CEE (1999), partition Y_t as,

$$Y_t = \begin{pmatrix} X_{St} \\ r_t \\ X_{ft} \end{pmatrix}$$

The benchmark timing identification assumption is

$$\begin{pmatrix} \varepsilon_t^S \\ \varepsilon_t^r \\ \varepsilon_t^f \end{pmatrix} = \begin{pmatrix} R_{SS} & 0 & 0 \\ R_{rS} & R_{rr} & 0 \\ R_{fS} & R_{fr} & R_{ff} \end{pmatrix} \begin{pmatrix} u_t^S \\ u_t^r \\ u_t^f \end{pmatrix}$$

or

slow-moving variables: $u_t^S = R_{SS}^{-1} \varepsilon_t^S$

policy instrument: $u_t^r = -R_{rr}^{-1} R_{rS} u_t^S + R_{rr}^{-1} \varepsilon_t^r$

fast-moving variable: $u_t^f = -R_{ff}^{-1} R_{fS} u_t^S + -R_{ff}^{-1} R_{fr} u_t^r + R_{ff}^{-1} \varepsilon_t^f$

Identification by Short Run Restrictions, ctd.

slow-moving variables: $u_t^S = R_{SS}^{-1}\varepsilon_t^S$

policy instrument: $u_t^r = -R_{rr}^{-1}R_{rS}u_t^S + R_{rr}^{-1}\varepsilon_t^r$

fast-moving variable: $u_t^f = -R_{ff}^{-1}R_{fS}u_t^S + -R_{ff}^{-1}R_{fr}u_t^r + R_{ff}^{-1}\varepsilon_t^f$

The space spanned by ε_t^S is spanned by (is identified as) the residual from regressing u_t^r on u_t^S.

Selected criticisms of timing restrictions (Rudebusch (1998), others)

- The implicit policy reaction function doesn't accord with theory or practical experience (does Fed ignore the stock market?)
- Implementations often ignore changes in policy reaction functions
- questionable credibility of lack of in-period response of X_{st} to r_t
- VAR information is typically far less than standard information sets
- Estimated monetary policy shocks don't match futures market data

Outline

1) VARs, SVARs, and the Identification Problem
2) Identification by Short Run Restrictions
3) **Identification by Long Run Restrictions**
4) The Quasi-Experiment Approach
5) Direct Estimation of Shocks from High Frequency Data
6) Identification by Sign Restrictions
7) Identification from Heteroskedasticity
8) DSGE Priors
9) Identification from Regional/Multicountry Restrictions
10) Inference: Challenges and Recently Developed Tools

3) Identification by Long Run Restrictions

Reduced form VAR: $A(L)Y_t = u_t$

Structural VAR: $B(L)Y_t = \varepsilon_t, \quad Ru_t = \varepsilon_t, \; B(L) = RA(L)$

This approach identifies R by imposing restrictions on the long run effect of one or more ε's on one or more Y's.

Long run variance matrix from VAR: $\Omega = A(1)^{-1}\Sigma_u A(1)^{-1\prime}$

Long run variance matrix from SVAR: $\Omega = B(1)^{-1}\Sigma_\varepsilon B(1)^{-1\prime}$

Digression: $B(1)^{-1} = D(1)$ is the long-run effect on Y_t of ε_t; this can be seen using the Beveridge-Nelson decomposition,

$$\sum_{s=1}^{t} Y_s = D(1)\sum_{s=1}^{t}\varepsilon_s + D^*(L)\varepsilon_t, \text{ where } D_i^* = -\sum_{j=i+1}^{\infty} D_j$$

Notation: think of Y_t as being growth rates, e.g. if Y_t is employment growth, $\Delta \ln N_t$, then $\sum_{s=1}^{t} Y_s$ is log employment, $\ln N_t$

Long run restrictions, ctd.

From VAR: $\Omega = A(1)^{-1}\Sigma_u A(1)^{-1\prime}$

From SVAR: $\Omega = B(1)^{-1}\Sigma_\varepsilon B(1)^{-1\prime} = RA(1)^{-1}\Sigma_\varepsilon A(1)^{-1\prime}R'$

System identification by long run restrictions. The SVAR is identified if

$$RA(1)^{-1}\Sigma_\varepsilon A(1)^{-1\prime}R' = \Omega \qquad (3)$$

can be solved for the unknown elements of R and Σ_ε..

- There are $k(k+1)/2$ distinct equations in (3), so the order condition says that you can estimate (at most) $k(k+1)/2$ parameters. If we set $\Sigma_\varepsilon = I$ (just a normalization), it is clear that we need $k^2 - k(k+1)/2 = k(k-1)/2$ restrictions on R.
- If $k = 2$, then $k(k-1)/2 = 1$, which is delivered by imposing a single exclusion restriction (that is, R is lower or upper triangular).
- This ignores rank conditions, which matter
- This is a moment matching approach; an IV interpretation comes later

Long run restrictions, ctd.

The long run neutrality restriction. The main way long restrictions are implemented in practice is by setting $\Sigma_\varepsilon = I$ and imposing zero restrictions on D(1). Imposing $D_{ij}(1) = 0$ says that the effect the long-run effect on the i^{th} element of Y_t, of the j^{th} element of ε_t is zero

If $\Sigma_\varepsilon = I$, the moment equation (3) can be rewritten,

$$\Omega = \text{D}(1)\text{D}(1)' \qquad (4)$$

where $\text{D}(1) = \text{B}(1)^{-1}$. Because $R\text{A}(1) = \text{B}(1)$, R is obtained from D(1) as $R = \text{A}(1)^{-1}\text{B}(1)$, and B(L) = RA(L) as above.

Comments:

- If the zero restrictions on D(1) make D(1) lower triangular, then D(1) is the Cholesky factorization of Ω.

Long run restrictions, ctd.

- Blanchard-Quah (1989) had 2 variables (unemployment and output), with the restriction that the demand shock has no long-run effect on the unemployment rate. This imposed a single zero restriction, which is all that is needed for system identification when $k = 2$.
- King, Plosser, Stock, and Watson (1991) work through system and partial identification (identifying the effect of only some shocks), things are analogous to the partial identification using short-run timing.
- This approach has been at the center of a "spirited" debate about whether technology shocks lead to a short-run decline in hours, based on long-run restrictions (Gali (1999), Christiano, Eichenbaum, and Vigfusson (2004, 2006), Erceg, Guerrieri, and Gust (2005), Chari, Kehoe, and McGrattan (2007), Francis and Ramey (2005), Kehoe (2006), and Fernald (2007))

Long run restrictions, ctd.

In this literature, Ω is estimated using the so-called VAR-HAC estimator,

VAR-HAC estimator of Ω: $\hat{\Omega} = \hat{A}(1)^{-1}\hat{\Sigma}_u\hat{A}(1)^{-1\prime}$

D(1) and R are estimated as: $\hat{D}(1) = \text{Chol}(\hat{\Omega}), \hat{R} = \left[\hat{D}(1)\hat{A}(1)\right]^{-1}$

Comments:

- A recurring theme is the sensitivity of the results to apparently minor specification changes, in Chari, Kehoe, and McGrattan's (2007) example results are sensitive to the lag length. It is unlikely that $\hat{\Sigma}_u$ is sensitive to specification changes, but $\hat{A}(1)$ is much more difficult to estimate.
- These observations are closely linked to the critiques by Faust and Leeper (1997), Pagan and Robertson (1998), Sarte (1997), Cooley and Dwyer (1998), Watson (2006), and Gospodinov (2008); we return to this below.
- An alternative is to use medium-run restrictions, see Uhlig (2004)

Outline

1) VARs, SVARs, and the Identification Problem
2) Identification by Short Run Restrictions
3) Identification by Long Run Restrictions
4) **The Quasi-Experiment Approach**
5) Direct Estimation of Shocks from High Frequency Data
6) Identification by Sign Restrictions
7) Identification from Heteroskedasticity
8) DSGE Priors
9) Identification from Regional/Multicountry Restrictions
10) Inference: Challenges and Recently Developed Tools

4) The Quasi-Experiment Approach

The quasi-experiment approach entails finding some external information (outside the model) that is relevant (correlated with the shock of interest) and exogenous (uncorrelated with the other shocks)

Selected references. Romer and Romer (1989, 2004, 2008); Ramey and Shapiro (1998); Ramey (2009).

Terminological digression. "Quasi-experiments" are also known as "Natural experiments." In the macro SVAR literature typically this approach is called the "narrative approach" because it is based on turning text-based information into quantitative information. The "quasi-experiment" term is used here to connect with the huge microeconometric literature on quasi-experiments that macro can/should draw on.

The Quasi-Experiment approach, ctd.

Suppose you have an instrumental variable Z_t (not in Y_t) such that

(i) $EZ_t u_t^r \neq 0$ (relevance)

(ii) $EZ_t \varepsilon_t^X = 0$ (exogeneity)

Then you can identify (estimate) ε_t^r. To show this, again, partition Y_t:

$$Y_t = \begin{pmatrix} {}^{(k-1\times 1)} X_t \\ {}^{(1\times 1)} r_t \end{pmatrix}, \; u_t = \begin{pmatrix} u_t^X \\ u_t^r \end{pmatrix}, \; \varepsilon_t = \begin{pmatrix} \varepsilon_t^X \\ \varepsilon_t^r \end{pmatrix}, \text{ and } R = \begin{pmatrix} R_{XX} & R_{Xr} \\ R_{rX} & R_{rr} \end{pmatrix}$$

so $Ru_t = \varepsilon_t$ becomes: $R_{XX} u_t^X = -R_{Xr} u_t^r + \varepsilon_t^X$

$$R_{rr} u_t^r = -R_{rX} u_t^X + \varepsilon_t^r$$

or

$$u_t^X = -R_{XX}^{-1} R_{Xr} u_t^r + R_{XX}^{-1} \varepsilon_t^X$$

$$u_t^r = -R_{rr}^{-1} R_{rX} u_t^X + R_{rr}^{-1} \varepsilon_t^r \qquad \text{where } R_{rr} \text{ is a scalar}$$

The Quasi-Experiment approach, ctd.

$$u_t^X = -R_{XX}^{-1} R_{Xr} u_t^r + R_{XX}^{-1} \varepsilon_t^X \quad (5)$$

$$u_t^r = -R_{rr}^{-1} R_{rX} u_t^X + R_{rr}^{-1} \varepsilon_t^r \quad (6)$$

Suppose Z_t (not in Y_t) is such that

(i) $EZ_t u_t^r \neq 0$ (relevance)

(ii) $EZ_t \varepsilon_t^X = 0$ (exogeneity)

Then Z_t can be used as an instrument for u_t^r in (5):

(i) Estimate $-R_{XX}^{-1} R_{Xr}$ by IV estimation of (5)

(ii) Estimate $\tilde{\varepsilon}_t^X = R_{XX}^{-1} \varepsilon_t^X$ as $\hat{\tilde{\varepsilon}}_t^X = u_t^X + \widehat{R_{XX}^{-1} R_{Xr}}\, u_t^r$

(iii) Use $\hat{\tilde{\varepsilon}}_t^X$ as an instrument for u_t^X in (6) to estimate $-R_{rr}^{-1} R_{rX}$

(iv) Estimate $\tilde{\varepsilon}_t^r = R_{rr}^{-1} \varepsilon_t^r$ as $u_t^r + \widehat{R_{rr}^{-1} R_{rX}}\, u_t^X$: this delivers ε_t^r up to scale.

(v) Impulse responses can be computed by regressing Y_t on $\tilde{\varepsilon}_t^r, \tilde{\varepsilon}_{t-1}^r, \ldots$

The Quasi-Experiment approach, ctd.

Comments

1. The IV approach outlined above needs one IV to identify one shock. More than one Z_t per shock yields overidentification.

2. Most papers that use this approach don't actually do IV, they report reduced-form regressions of variables of interest onto Z_t. In general the reduced form regressions don't give you the structural coefficients of interest. On the other hand one reason for choosing reduced form over IV is that there might be heterogeneity in "treatment effects" of different types of shocks. In this case the IV estimator using Z_t would have the additional complication that the

The Quasi-Experiment approach, ctd.

local average treatment effect (LATE) for the Z_t used in the study would differ from the average treatment effect, or from the LATE using a different instrument. This set of issues has received a lot of attention in the "program evaluation" literature but almost no attention in empirical macro.

Outline

1) VARs, SVARs, and the Identification Problem
2) Identification by Short Run Restrictions
3) Identification by Long Run Restrictions
4) The Quasi-Experiment Approach
5) **Direct Estimation of Shocks from High Frequency Data**
6) Identification by Sign Restrictions
7) Identification from Heteroskedasticity
8) DSGE Priors
9) Identification from Regional/Multicountry Restrictions
10) Inference: Challenges and Recently Developed Tools

5) Direct Estimation of Shocks from High Frequency Data

Monetary shock application: Estimate ε_t^r directly from daily data on monetary announcements or policy-induced FF rate changes:

Recall,

$$Y_t = \begin{pmatrix} D_{YX}(L) & D_{Yr}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_t^X \\ \varepsilon_t^r \end{pmatrix} = D_{Yr}(\mathrm{L})\varepsilon_t^r + v_t,$$

where $v_t = D_{YX}(\mathrm{L})\varepsilon_t^X$, so if you observed ε_t^r you could estimate $D_{Yr}(\mathrm{L})$.

- Cochrane and Piazessi (2002)

  aggregates daily ε_t^r (Eurodollar rate changes after FOMC announcements) to a monthly ε_t^r series

- Faust, Swanson, and Wright (2003. 2004)

  estimates IRF of r_t wrt ε_t^r from futures market, then matches this to a monthly VAR IRF (results in set identification – discuss later)

- Bernanke and Kuttner (2005)

Outline

1) VARs, SVARs, and the Identification Problem
2) Identification by Short Run Restrictions
3) Identification by Long Run Restrictions
4) The Quasi-Experiment Approach
5) Direct Estimation of Shocks from High Frequency Data
6) **Identification by Sign Restrictions**
7) Identification from Heteroskedasticity
8) DSGE Priors
9) Identification from Regional/Multicountry Restrictions
10) Inference: Challenges and Recently Developed Tools

6) Identification by Sign Restrictions

Consider restrictions of the form: a monetary policy shock…

- does not decrease the FF rate for months 1,…,6
- does not increase inflation for months 6,..,12

These are restrictions on the sign of elements of D(L).

Sign restrictions can be used to set-identify D(L). Let $\mathcal{D}$ denote the set of D(L)'s that satisfy the restriction. There are currently three ways to handle sign restrictions:

1. Faust's (1998) quadratic programming method
2. Uhlig's (2005) Bayesian method
3. Uhlig's (2005) penalty function method

I will describe #2 (the first steps are the same as #3)

Sign restrictions, ctd.

SVAR identification: $R\Sigma_u R' = \Sigma_\varepsilon$

Normalize $\Sigma_\varepsilon = I$; then $\Sigma_u = R^{-1}R^{-1\prime}$

One statement of the SVAR identification problem is that, without additional restrictions, R is identified only up to a rotation. Let $R_c^{-1} = Chol(\Sigma_u)$ so $R_c^{-1}R_c^{-1\prime} = \Sigma_u$. Then it is also true that

$$\Sigma_u = R_c^{-1}HH'R_c^{-1\prime}$$

for any $n \times n$ orthonormal matrix H. Let $R^{-1} = R_c^{-1}H$, or

$$R = H^{-1}R_c.$$

Then R is also a solution to $R\Sigma_u R' = \Sigma_\varepsilon$.

Sign restrictions, ctd.

Uhlig's algorithm (slight modification):

(i) Draw $\tilde{H}$ randomly from the space of orthonormal matrices

(ii) Compute $\tilde{R} = \tilde{H}^{-1} R_c$

(iii) Compute the IRF $\tilde{D}(L) = \mathrm{C(L)}\tilde{R}^{-1} = \mathrm{A(L)}^{-1} R_c^{-1} \tilde{H}$

(iv) If $\tilde{D}(L) \notin \mathcal{D}$, discard this trial $\tilde{R}$ and go to (i). Otherwise, if $\tilde{D}(L) \in \mathcal{D}$, retain $\tilde{R}$ then go to (i)

(v) Compute the posterior (using a prior on A(L) and Σ_u, plus the retained $\tilde{R}$'s) and conduct Bayesian inference, e.g. compute posterior mean (integrate over A(L), Σ_u, and the retained $\tilde{R}$'s), compute credible sets (Bayesian confidence sets), etc.

This algorithm implements Bayes inference using a prior proportional to

$$\pi(A(\mathrm{L}), \Sigma_u) \times \mathbf{1}(\tilde{D}(L) \in \mathcal{D}) \mu(H)$$

where $\mu(H)$ is the distribution from which H is drawn.

Sign restrictions, ctd.

Sign restriction prior: $\pi(A(\mathrm{L}), \Sigma_u) \times \mathbf{1}(\tilde{D}(L) \in \mathcal{D}) \mu(H)$

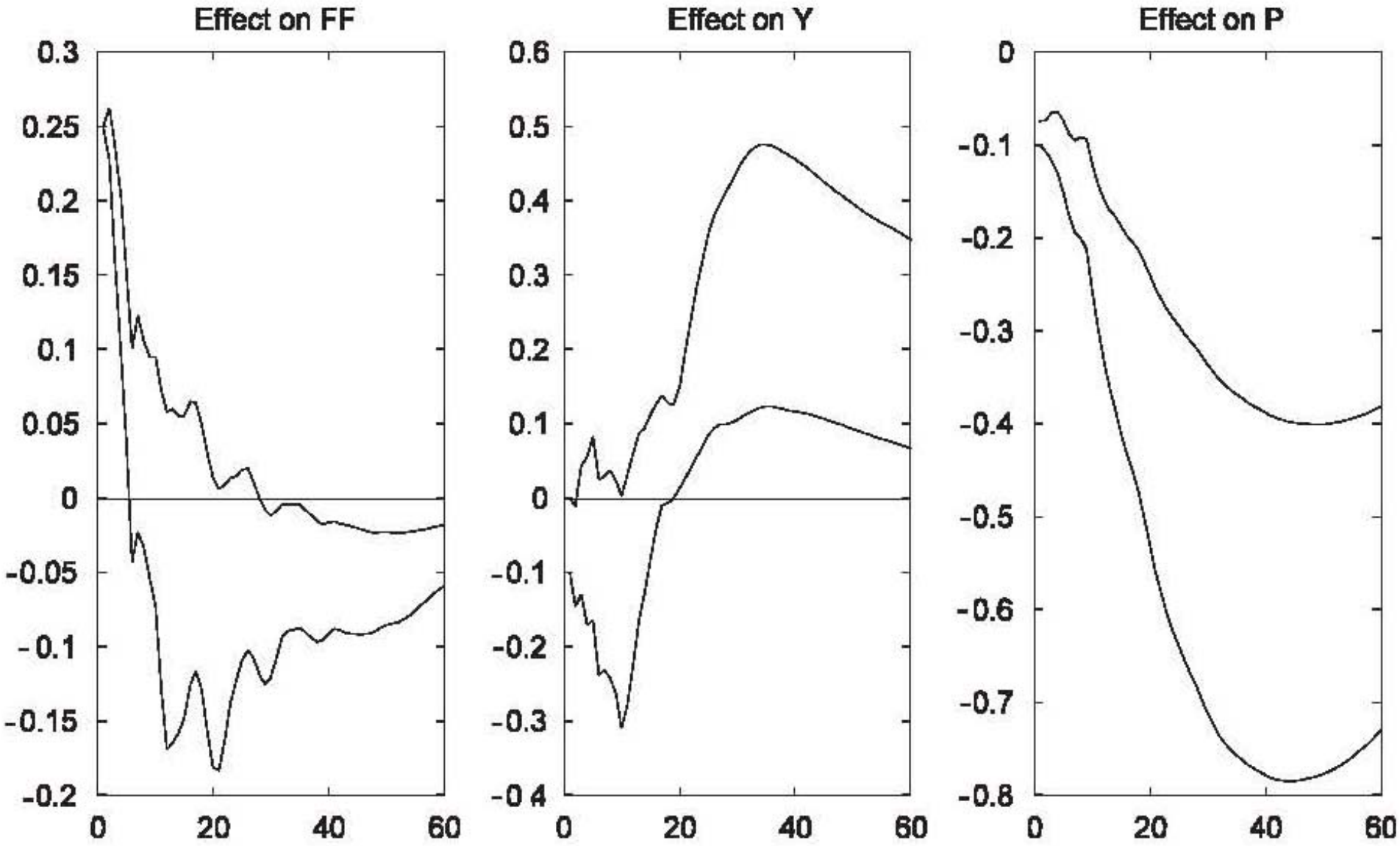
Comments:

- This procedure results in set identification. This raises difficult issues for inference and is an active area of research in econometric theory.
- From a frequentist perspective, the identified set is estimated by the collection of nonrejected impulse responses,
  - $\{\hat{D}(L) = \hat{C}(L) R_c^{-1} \tilde{H} : \hat{D}(L) \in \mathcal{D}\}$
- This set is readily computed by saving the $\tilde{H}$'s that result in nonrejected $\hat{D}(L)$'s.
- The nonrejected $\hat{D}(L)$'s are ***not*** draws that can be used to compute frequentist confidence bands – they ***all*** fall in the set of point estimates!
- It appears (here) that the choice of μ matters but it doesn't in Uhlig's implementation.

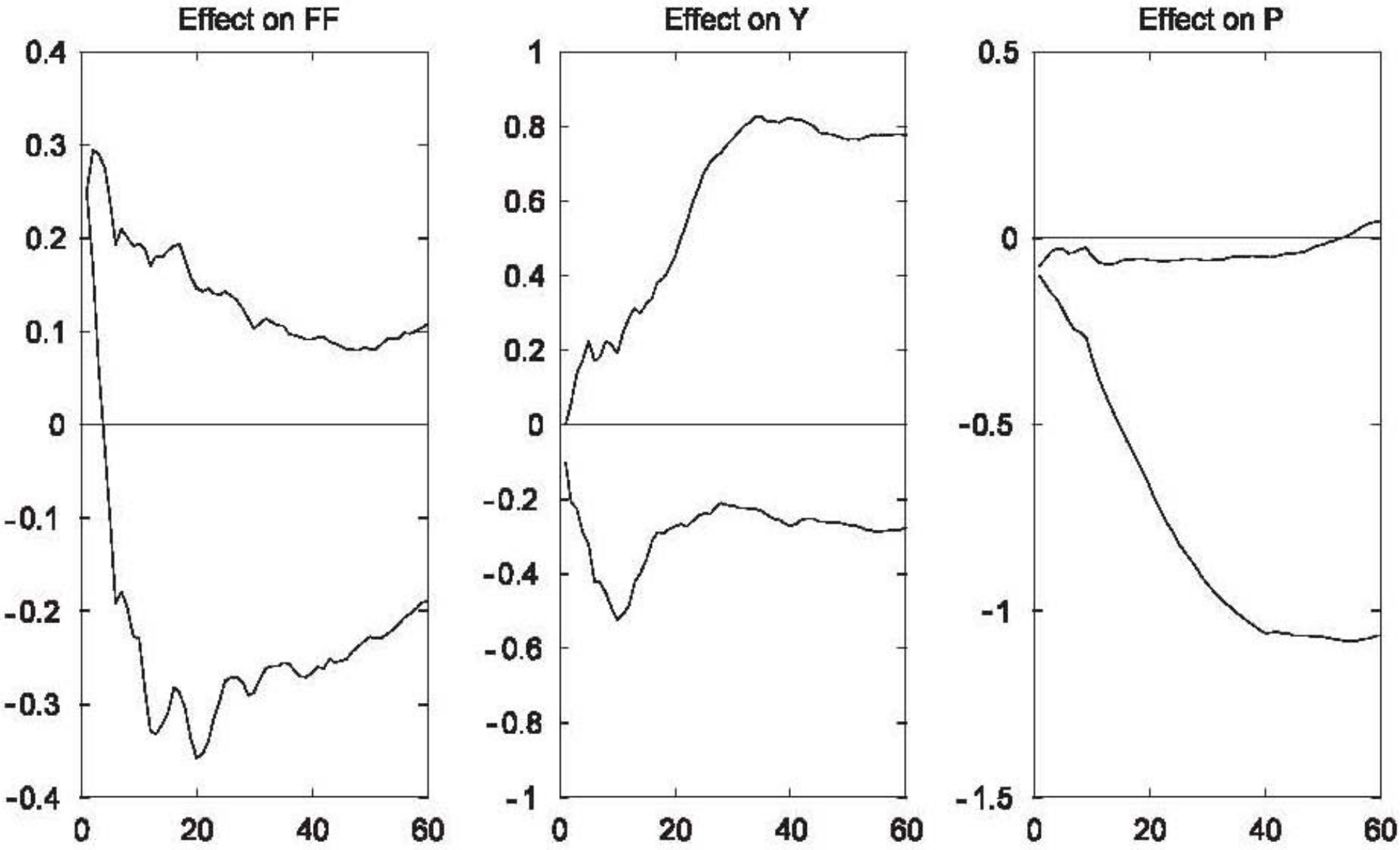
Sign restrictions, ctd.

- Frequentist inference about estimation of the set is difficult – it requires imagining what the set would have been, had different ($\hat{A}(L)$, $\hat{\Sigma}_u$)'s been realized. See Moon, Schorfheide, Granziera & Lee (2009).
- A confidence set for the identified set is a set-valued function of the data that contains the true identified set in 95% of all realizations.
- ***Methodologically related example – set identification.*** Faust, Swanson, and Wright (2004) have a related identification scheme – ***not sign restrictions –*** that also gives set identification (they require selected SVAR IRFs to match IRFs based on high frequency asset data). They take a stab at performing inference about the identified set itself. As an illustration look first at their estimates of the identified set (figure 3), then their confidence intervals for the identified set:

Estimates of identified sets (Faust, Swanson, Wright (2003), Fig 3)



Confidence sets for identified sets (Faust, Swanson, Wright (2003) , Fig 4)



Outline

1) VARs, SVARs, and the Identification Problem
2) Identification by Short Run Restrictions
3) Identification by Long Run Restrictions
4) The Quasi-Experiment Approach
5) Direct Estimation of Shocks from High Frequency Data
6) Identification by Sign Restrictions
7) **Identification from Heteroskedasticity**
8) DSGE Priors
9) Identification from Regional/Multicountry Restrictions
10) Inference: Challenges and Recently Developed Tools

7) Identification from Heteroskedasticity

Suppose:

(a) The structural shock variance breaks at date s: $\Sigma_{\varepsilon,1}$ before, $\Sigma_{\varepsilon,2}$ after

(b) R doesn't change between variance regimes

(c) normalize R to have 1's on the diagonal, but no other restrictions; thus the unknowns are: R (k^2–k); $\Sigma_{\varepsilon,1}$ (k), and $\Sigma_{\varepsilon,2}$(k).

First period: $R\Sigma_{u,1}R' = \Sigma_{\varepsilon,1}$ $k(k+1)/2$ equations, k^2 unknowns

Second period: $R\Sigma_{u,2}R' = \Sigma_{\varepsilon,2}$ $k(k+1)/2$ equations, k more unknowns

Number of equations = $k(k+1)/2 + k(k+1)/2 = k(k+1)$

Number of unknowns = $k^2 - k + k + k = k(k+1)$

Rigobon (2003), Rigobon and Sack (2003, 2004)

ARCH version by Sentana and Fiorentini (2001)

Identification from Heteroskedasticity,ctd.

Comments:

1. There is a rank condition here too – for example, identification will not be achieved if $\Sigma_{\varepsilon,1}$ and $\Sigma_{\varepsilon,2}$ are proportional.
2. The break date need not be known as long as it can be estimated consistently
3. Different intuition: suppose only one structural shock is homoskedastic. Then find the linear combination without any heteroskedasticity!
4. This idea also can be implemented exploiting conditional heteroskedasticity (Sentana and Fiorentini (2001))
5. But, some cautionary notes:
   a. R must remain constant despite change in Σ_{ε} (think about it…)
   b. Strong identification will come from large differences in variances

Outline

1) VARs, SVARs, and the Identification Problem
2) Identification by Short Run Restrictions
3) Identification by Long Run Restrictions
4) Identification by Sign Restrictions
5) The Quasi-Experiment Approach
6) Direct Estimation of Shocks from High Frequency Data
7) Identification from Heteroskedasticity
8) **DSGE Priors**
9) Identification from Regional/Multicountry Restrictions
10) Inference: Challenges and Recently Developed Tools

8) DSGE Priors

Use priors based on a DSGE towards which to shrink the VAR parameters. For example compute the approximate VAR(p) implied by the DSGE, use these as point estimates for B(L), and center conjugate prior at those. More on this in Lecture 8.

Selected references

Ingram and Whiteman (1994)

Del Negro and Schorfheide (2004)

Del Negro, Schorfheide, Smets, and Wouters (2004)

Outline

1) VARs, SVARs, and the Identification Problem
2) Identification by Short Run Restrictions
3) Identification by Long Run Restrictions
4) The Quasi-Experiment Approach
5) Direct Estimation of Shocks from High Frequency Data
6) Identification by Sign Restrictions
7) Identification from Heteroskedasticity
8) DSGE Priors
9) **Identification from Regional/Multicountry Restrictions**
10) Inference: Challenges and Recently Developed Tools

9) Identification from Regional/Multicountry Restrictions

The idea is to impose restrictions arising from country boundaries:

(a) distinguish between country-specific and common shocks, e.g. using a factor structure;

(b) impose additional restrictions arising from transmission via trade shares

Selected references

Canova and Ciccarelli (2008)

Dees, di Mauro, Pesaran, and Smith (2007) (many other Pesaran/Smith)

Elliott and Fatás (1996)

Norrbin and Schlagenhauf (1996)

Stock and Watson (2005)

Outline

1) VARs, SVARs, and the Identification Problem
2) Identification by Short Run Restrictions
3) Identification by Long Run Restrictions
4) The Quasi-Experiment Approach
5) Direct Estimation of Shocks from High Frequency Data
6) Identification by Sign Restrictions
7) Identification from Heteroskedasticity
8) DSGE Priors
9) Identification from Regional/Multicountry Restrictions
10) **Inference: Challenges and Recently Developed Tools**

10) Inference: Challenges and Recently Developed Tools

Two topics:

(a) Inference using long-run restrictions

(b) Inference about IRFs

Inference using long-run restrictions

Recall the estimator of R under the long-run neutrality condition with lower triangular restrictions on D(1):

$$\hat{R} = \left[Chol\left(\hat{\Omega}\right)\hat{A}(1) \right]^{-1} = \left[Chol\left(\hat{A}(1)^{-1}\hat{\Sigma}_u\hat{A}(1)^{-1\prime}\right)\hat{A}(1) \right]^{-1}$$

Conventional inference requires that $\hat{R}$ be consistent with a sampling distribution that is well-approximated by a normal. However, inference about Ω is difficult. There are two ways to think about these inference issues: as a HAC estimator (Lecture 5) and as an IV estimator (now).

IV interpretation of LR restrictions

Shapiro and Watson (1988); Pagan and Robertson (1998), Sarte (1997), Cooley and Dwyer (1998); Watson (2006), Gospodinov (2008)

Preliminaries

(i) *Restrictions on D(1) mean restrictions on B(1)*:

SVAR: $B(L)Y_t = \varepsilon_t$

$D(1) = B(1)^{-1}$

Consider 2-variable VAR:

$$\begin{pmatrix} D_{11}(1) & D_{12}(1) \\ D_{21}(1) & D_{22}(1) \end{pmatrix} = \begin{pmatrix} B_{11}(1) & B_{12}(1) \\ B_{21}(1) & B_{22}(1) \end{pmatrix}^{-1} = \begin{pmatrix} B_{22}(1) & -B_{12}(1) \\ -B_{21}(1) & B_{11}(1) \end{pmatrix} \Big/ \det(B(1))$$

so $D_{12}(1) = 0$ is equivalent to $B_{12}(1) = 0$. Estimation of D(1) with $D_{12}(1) = 0$ is equivalent to estimation of B(1) with $B_{12}(1) = 0$.

(ii) *Lag manipulation*. Recall the Beveridge-Nelson decomposition for a lag polynomial of degree p:

$$c(\mathrm{L}) = c(1) + c^*(\mathrm{L})\Delta, \text{ where } c_j^* = -\sum_{i=j+1}^{p} c_i$$

This is not unique; you can load $c(1)$ on any lag, in particular, lag p:

$$c(\mathrm{L}) = c(1)\mathrm{L}^p + c^+(\mathrm{L})\Delta, \text{ where } c_j^+ = \sum_{i=1}^{j} c_i$$

Call this the "reverse BN decomposition."

IV interpretation of LR restrictions for a 2-variable SVAR

Let $\quad B(L) = \begin{pmatrix} 1-b_{11}(L) & -b_{12}(L) \\ -b_{21}(L) & 1-b_{22}(L) \end{pmatrix}$

so $B(L)Y_t = \varepsilon_t$ becomes,

$$Y_{1t} = b_{12}(L)Y_{2t} + b_{11}(L)Y_{1t-1} + \varepsilon_{1t}$$
$$Y_{2t} = b_{21}(L)Y_{1t} + b_{22}(L)Y_{2t-1} + \varepsilon_{2t}$$

Apply the "reverse BN decomposition":

$$Y_{1t} = b_{12}(1)Y_{2t-p} + b_{12}^{+}(L)\Delta Y_{2t} + b_{11}(L)Y_{1t-1} + \varepsilon_{1t}$$
$$Y_{2t} = b_{21}(1)Y_{1t-p} + b_{21}^{+}(L)\Delta Y_{1t} + b_{22}(L)Y_{2t-1} + \varepsilon_{2t}$$

Impose the long-run neutrality restriction $D_{12}(1) = 0$, i.e. $b_{12}(1) = 0$:

$$Y_{1t} = b_{12}^{+}(L)\Delta Y_{2t} + b_{11}(L)Y_{1t-1} + \varepsilon_{1t}$$
$$Y_{2t} = b_{21}(1)Y_{1t-p} + b_{21}^{+}(L)\Delta Y_{1t} + b_{22}(L)Y_{2t-1} + \varepsilon_{2t}$$

IV interpretation of LR restrictions for a 2-variable SVAR, ctd

$$Y_{1t} = b_{12}(1)Y_{2t-p} + b_{12}^{+}(L)\Delta Y_{2t} + b_{11}(\mathrm{L})Y_{1t-1} + \varepsilon_{1t} \qquad (7)$$

$$Y_{2t} = b_{21}(1)Y_{1t-p} + b_{21}^{+}(L)\Delta Y_{1t} + b_{22}(\mathrm{L})Y_{2t-1} + \varepsilon_{2t} \qquad (8)$$

The long-run restriction $b_{12}(1)=0$ implies an exclusion restriction: Y_{2t-p} doesn't appear in (7), but it does appear in (8). Thus:

- the coefficient $b_{12,0}^{+}$ on ΔY_{2t} in (7) can be estimated by IV, using Y_{2t-p} as an instrument for ΔY_{2t}.

Because $\Delta Y_{t-1},\ldots, \Delta Y_{t-p+1}$ appear as regressors in (7), this is equivalent to:

- the coefficient $b_{12,0}^{+}$ on ΔY_{2t} in (7) can be estimated by IV, using Y_{2t-1} as an instrument for ΔY_{2t}.

IV interpretation of LR restrictions for a 2-variable SVAR, ctd

Weak instrument interpretation

Is Y_{2t-1} a weak or strong instrument?

First-stage regression:

regress ΔY_{2t} on Y_{2t-1}, ΔY_{2t-1}, ΔY_{2t-2},…, Y_{1t-1}, Y_{1t-2},…

Back of the envelope calculation: approximate Y_{2t} as the AR(1),

Approximation: $Y_{2t} = \alpha Y_{2t-1} + \varepsilon_{1t}$

or $\Delta Y_{2t} = (\alpha - 1) Y_{2t-1} + \varepsilon_{1t}$ (9)

In IV notation: $\mathbf{Y} = \mathbf{Z}\Pi + \mathbf{v}$

Concentration parameter: $\mu^2 = \Pi' \mathbf{Z}' \mathbf{Z} \Pi / \sigma_v^2$

Translated to the first-stage regression (9):

$$E\mu^2 = (\alpha - 1)^2 \times \text{var}(Y_{2t-1}) \times T / \sigma_{\varepsilon_1}^2 = \frac{(\alpha - 1)^2}{1 - \alpha^2} T$$

IV interpretation of LR restrictions for a 2-variable SVAR, ctd

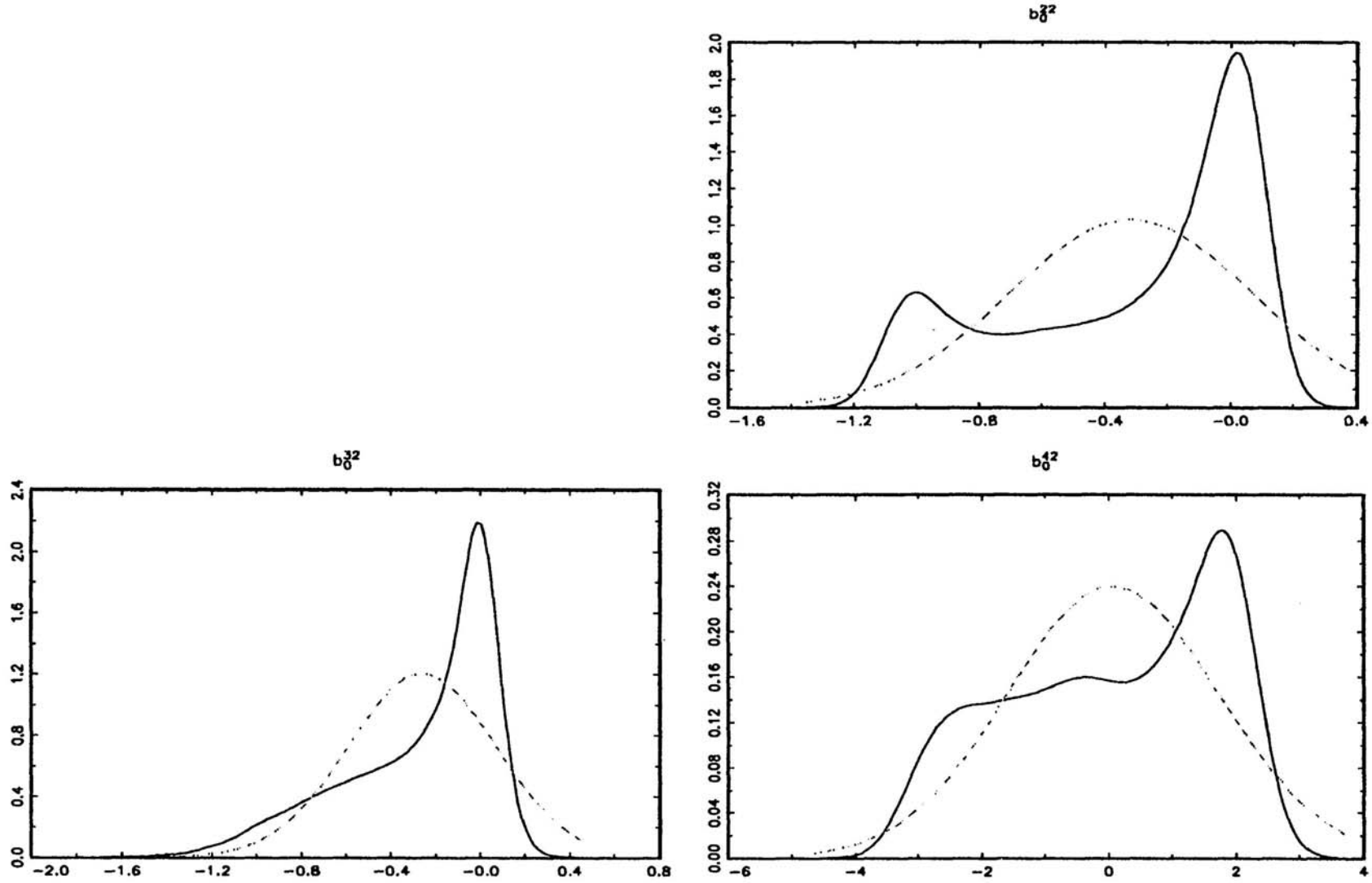
Values of $\mu^2 = \frac{(\alpha - 1)^2}{1 - \alpha^2} T$ for $T = 100$:

α	μ^2
0.5	33
0.9	5.3
0.95	2.6

- These are probably best-case numbers, in higher order ARs and in VARs the marginal contribution of Y_{2t-1} given additional lags would be less
- In the local to unit case (very persistent), $\mu^2 = O_p(1)$ random variable (Gospodinov (2008)) (*but note, Y_t is supposed to be stationary*)
- Some simulations from Pagan and Robertson (1998) of estimated long-run effects:

MC results of long run effects estimated by imposing LR neutrality restrictions, from Pagan and Robertson (1998)

FIGURE 3.—SIMULATED DENSITIES OF $\hat{b}_0^{j2}$ IN LASTRAPES AND SELGIN MODEL



IV interpretation of LR restrictions ctd

Comments

- The IV interpretation and the Ω estimation interpretation both suggest that (in some applications) there can be considerable sensitivity to sample period and especially lag length – which we would expect if identification is weak.
- Whether this is an issue depends on the amount of persistence. If persistence is small, Y_{2t} will be a stronger instrument (and Ω will be easier to estimate)
- Some practical advice: perform a MC simulation – and don't trust boostrap SEs without checking in a MC
- Francis, Owyang, and Roush (2005) change the infinite-run restriction to a finite long-run restriction using Faust's algorithm – a sensible approach worth following up.
- More work is still needed (especially tools for handling weak IVs)

(b) Confidence Intervals for IRFs

The goal is to provide confidence intervals for IRFs that provide the stated coverage rate and are as tight as possible. The natural starting point is first order asymptotic theory. Remember the "delta method"?

If $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \text{N}(0, \Sigma_{\hat{\theta}})$ and if $g(\bullet)$ has continuous derivatives, then

$$\sqrt{T}\,[g(\hat{\theta}) - g(\theta_0)] \approx \sqrt{T}\,\frac{\partial g}{\partial \theta'}\bigg|_{\theta_0}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, \frac{\partial g}{\partial \theta'}\bigg|_{\theta_0} \Sigma_{\hat{\theta}} \frac{\partial g}{\partial \theta'}\bigg|'_{\theta_0}\right)$$

For this to provide a good approximation, $\hat{\theta}$ should be nearly normal to start with and g must be essentially linear over most of the mass of the sampling distribution of $\hat{\theta}$.

For SVAR IRFs, $\hat{\theta} = (\hat{A}(L), \hat{R})$, and $\text{g}(\hat{\theta}) = \hat{D}(L) = \hat{A}(L)^{-1}\hat{R}^{-1}$.

Confidence Intervals for IRFs, ctd.

$$\hat{\theta} = (\hat{A}(L),\ \hat{R}), \text{ and } g(\hat{\theta}) = \hat{D}(L) = \hat{A}(L)^{-1}\hat{R}^{-1}$$

In SVAR applications there are two main problems with the delta method:

1. The function g is very nonlinear so that even if $\hat{A}(L)$ were exactly normally distributed, the impulse response functions might not be. Let $\hat{\alpha} \sim$ N(.25,.25), what is the distribution of $\hat{\alpha}^4$? $\hat{\alpha}^5$?

2. Moreover, $\hat{A}(L)$ is not well approximated by a normal. It is well known that if the roots of A(L) are large, then $\hat{A}(L)$ will exhibit substantial bias towards zero. In fact, in the limit that the roots are local to unity, $\hat{A}(L)$ will not have a normal asymptotic distribution.

The problem of persistent roots complicating inference on D(L) is particularly important for medium and longer horizons…

Confidence Intervals for IRFs, ctd.

An example: the distribution of estimated IRF of an AR(1) at long horizons (Stock (1996, 1997), Phillips (1998)).

Suppose

$$Y_t = \alpha Y_{t-1} + \varepsilon_t$$

and model α as local to unity: $\alpha = 1+c/T$, where c is a constant. The estimated IRF at horizon h is,

$$\hat{\alpha}^h = (1+\hat{c}/\mathrm{T})^h \approx e^{\frac{h}{T}\hat{c}}$$

Suppose that $h/T \to \kappa$ (the horizon is a fraction of the sample size). Then a direct application of local-to-unity asymptotic theory yields,

$$\hat{\alpha}^h \Rightarrow e^{\kappa c} \exp\left[\kappa \frac{\int J_c(s)dW(s)}{\int J_c^2(s)ds}\right]$$

where J_c is an Ornstein-Uhlenbeck process. The true IRF is $e^{\kappa c}$; the estimated IRF is that, times a nonnormal random factor.

Confidence Intervals for IRFs, ctd.

The problem posed by large roots worsens as the horizon increases. The fraction of the sample at which problems for the normal approximation arise is surprisingly short, say 10%.

Pesavento and Rossi (2005, 2006) provide one method for handling this (and provide an application to the LR technology shock debate). Their procedure is theoretically justified – controls coverage rates – but is cumbersome and can produce wide intervals.

The other methods in the literature (the methods we now turn to) do not handle large roots in theory, but some seem to work OK in Monte Carlo studies (and most certainly some are better than others).

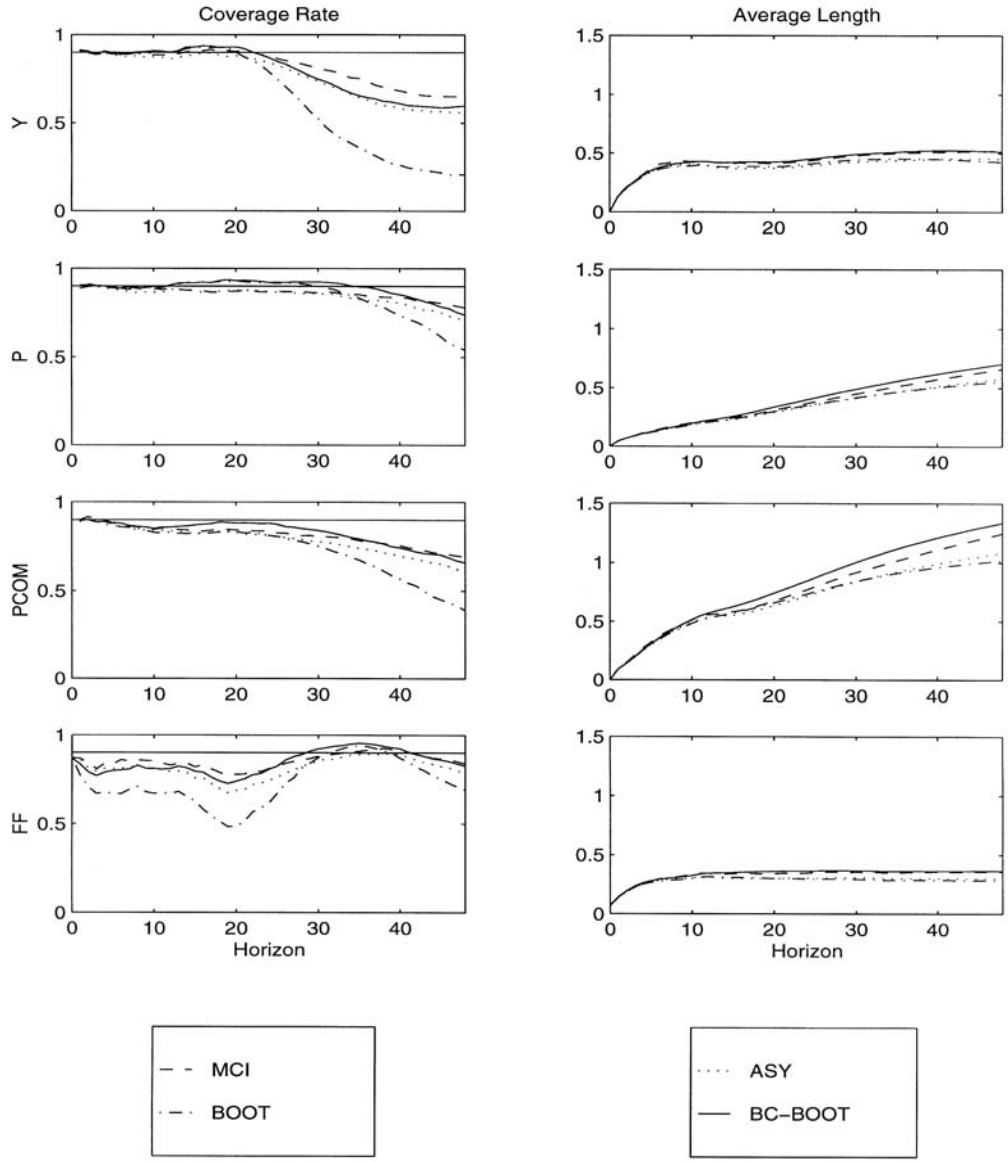
Confidence Intervals for IRFs, ctd.

Standard methods for constructing confidence intervals for D(L):

1. Delta method (see Lütkepohl (2005))
2. Bootstrap methods (Runkle (1987), Kilian (1998a, 1998b, 2001))
3. Bayesian methods (Sims and Zha (1999))

Kilian and Chang (2000) results

- MC simulation evidence comparing the delta method, the Sims-Zha (1999) Bayesian method, the Runkle (1987) unadjusted bootstrap, Kilian (1998) adjusted bootstrap.
- Results presented here are for the Bernanke-Gertler (1995) VAR(12), 4 variables, 348 monthly observations



NOTES: Simulation results based on 1,000 trials. DGP described in text and data appendix. MCI = Monte Carlo integration method of Sims and Zha (1999), ASY = asymptotic delta method of Lütkepohl (1990), BC-BOOT = bias-corrected bootstrap method of Kilian (1998a), BOOT = bootstrap method of Runkle (1987).

Fig. 1. Coverage accuracy and average length of impulse response confidence intervals. Bernanke–Gertler model (VAR(12) for monthly data).

Confidence Intervals for IRFs, ctd.

Comments:

- The coverage rates of IRFs depend on the VAR (the MC design).
- The method that seems to work best (even though it technically isn't valid when roots are local to unity) is Kilian's bias-adjusted bootstrap. Here is Kilian's algorithm:
  - (i) Compute VAR estimates $\hat{A}(L)$
  - (ii) Compute bias-adjusted VAR estimates (there are two ways to do this – using Pope's (1990) bias formulas for VARs or by bootstrap simulation; see Kilian (1998b, 2001 appendix) for details)
  - (iii) Bootstrap the foregoing (that is, the bias-adjusted estimates)
  - (iv) Use percentiles of the bootstrap IRF draws, horizon by horizon, to compute confidence bands (i.e. use the percentile, not percentile-*t*, bootstrap method).

Concluding comments

- Identification remains at the core of successful SVAR modeling
- There have been some creative and promising developments (the DSGE, high frequency, IRF sign restrictions, and heteroskedasticity techniques all have plausibly valid applications)
- There are also some subtle issues of inference which are not fully resolved – you need to be aware of these (at least):
  - Set identification issues and interpretation of "confidence bands" computed using sign restriction methods
  - Inference with long run restrictions: can be viewed either as a (possibly) weak instruments problem or a difficult HAC estimation problem – the problem is most serious when the data have substantial low frequency components.
  - Deterioration of IRF confidence bands at long horizons (and sometimes at short horizons); limitations of standard IRF bootstrap.