AEA Continuing Education Course

Time Series Econometrics

Lecture 7

Structural Vector Autoregressions: Recent Developments

James H. Stock

Harvard University

January 6 & 7, 2015

Outline

1) VARs, SVARs, and the Identification Problem
2) Classical approaches to identification
    2a) Identification by Short Run Restrictions
    2b) [Identification by Long Run Restrictions]
3) New approaches to identification (post-2000)
    3a) Identification from Heteroskedasticity
    3b) Direct Estimation of Shocks from High Frequency Data
    3c) External instruments
    3d) Identification by Sign Restrictions

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): DSGE or SVAR
2) Quasi-Experiments

The identification problem. *Consider the Reduced form VAR(p)*:

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + u_t$$

or $\quad \mathrm{A(L)}Y_t = u_t$, where $\mathrm{A(L)} = I - A_1\mathrm{L} - A_2\mathrm{L}^2 - \ldots - A_p\mathrm{L}^p$

where A_i are the coefficients from the (population) regression of Y_t on $Y_{t-1},\ldots,Y_{t-p}$.

- $u_t = Y_t - \mathrm{Proj}(Y_t|Y_{t-1},\ldots, Y_{t-p})$ are the innovations, and are identified.
- If u_t were the shocks, then we could compute the structural IRF using the MA representation of the VAR, $Y_t = \mathrm{A(L)}^{-1}u_t$.
- But in general u_t is affected by multiple shocks: in any given quarter, GDP changes unexpectedly for a variety of reasons.
- For example, if $n = 2$,

$$u_{1t} = \mathbf{R_{12}}\boldsymbol{u_{2t}} + \varepsilon_{1t}$$

$$u_{2t} = \mathbf{R_{21}}\boldsymbol{u_{1t}} + \varepsilon_{2t}$$

  - To identify R we need an instrument Z_t or a restriction on the parameters.
  - For example, $\mathrm{R}_{12} = 0$ identifies R (Cholesky decomposition)

<u>Reduced form to structure:</u>

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

(ii) u_t spans the space of structural shocks ε_t, that is, $\varepsilon_t = \mathrm{R}u_t$, 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

Because $\varepsilon_t = \mathrm{R}u_t$,

$$\mathrm{RA(L)}Y_t = \mathrm{R}u_t = \varepsilon_t.$$

Letting RA(L) = B(L), this delivers the structural VAR,

$$\mathrm{B(L)}Y_t = \varepsilon_t,$$

The MA representation of the SVAR delivers the structural IRFs:

$$Y_t = D(\mathrm{L})\varepsilon_t,\ D(\mathrm{L}) = \mathrm{B(L)}^{-1} = \mathrm{A(L)}^{-1}\mathrm{R}^{-1}$$

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

Summary of VAR and SVAR notation

<table>
<tr><td>

Reduced form VAR

$$A(L)Y_t = u_t$$
$$Y_t = A(L)^{-1}u_t = C(L)u_t$$
$$A(L) = I - A_1L - A_2L^2 - \ldots - A_pL^p$$
$$Eu_tu_t' = \Sigma_u \text{ (unrestricted)}$$

</td><td>

Structural VAR

$$B(L)Y_t = \varepsilon_t$$
$$Y_t = B(L)^{-1}\varepsilon_t = D(L)\varepsilon_t$$
$$B(L) = B_0 - B_1L - B_2L^2 - \ldots - B_pL^p$$
$$E\varepsilon_t\varepsilon_t' = \Sigma_\varepsilon = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_k^2 \end{pmatrix}$$

</td></tr>
<tr><td colspan="2">

$$Ru_t = \varepsilon_t$$
$$B(L) = RA(L) \quad (B_0 = R)$$
$$D(L) = C(L)R^{-1}$$

</td></tr>
</table>

- 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 equivalent views

1. ***Identification of R.*** In population, we can know A(L). If we can identify R, we can obtain the SVAR coefficients, B(L) = RA(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:

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

The IRF/MA form is $Y_t = \mathrm{D(L)}\varepsilon_t$, 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 (references below)
- 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), Hansen and Sargent (2007), E. Sims (2012), Blanchard, L'Huillier, and Lorenzoni (2012), Forni, Gambetti, and Sala (2012), and Gourieroux and Monfort (2014)

This talk

- Early promise of SVARs

  Surveys of classical methods: Christiano, Eichenbaum, and Evans (1999), Lütkepohl (2005), Stock and Watson (2001), Watson (1994)

  Survey of new ideas about how to tackle the identification problem

- Critiques of the 1990s

- This talk focuses on the interesting new work on identification – much of it quite recent – in response to those critiques

Outline

1) VARs, SVARs, and the Identification Problem

2) Classical approaches to identification

 2a) Identification by Short Run Restrictions

 2b) [Identification by Long Run Restrictions]

3) New approaches to identification (post-2000)

 3a) Identification from Heteroskedasticity

 3b) Direct Estimation of Shocks from High Frequency Data

 3c) External instruments

 3d) Identification by Sign Restrictions

2a) Identification by Short Run Restrictions

<u>Overview: the traditional SVAR identification approach</u>

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

(a) 2-variable example.

$$u_{1t} = \mathbf{R_{12}} \boldsymbol{u_{2t}} + \varepsilon_{1t}$$

$$u_{2t} = \mathbf{R_{21}} \boldsymbol{u_{1t}} + \varepsilon_{2t}$$

- Suppose $R_{12} = 0$. E.g. Blanchard and Galí (2007) for oil price shocks.
- Then $\varepsilon_{1t} = u_{1t}$ so R_{21} can be estimated by OLS (u_{1t} is uncorrelated with ε_{2t}).
- How credible is the Blanchard-Galí assumption?

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

$$R\Sigma_u R' = \Sigma_\varepsilon$$

can be solved for the unknown elements of R and Σ_ε.. Recall that Σ_u is identified.

- There are $k(k+1)/2$ distinct equations in the matrix equation above, 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), there are k^2 parameters
- So 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 can matter.
- This description of identification is via method of moments, however identification can equally be described via IV, e.g. see Blanchard and Watson (1986).

(c) Identification of only one shock or IRF. 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:

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

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

(d) The "fast-r-slow" scheme. Almost all short-run restriction applications can be written as "fast-r-slow." Following CEE (1999), the benchmark timing identification assumption is

$$\begin{pmatrix} \varepsilon_t^S \\ \varepsilon_t^r \\ \varepsilon_t^f \end{pmatrix} = \begin{pmatrix} R_{SS} & 0 & 0 \\ R_{rS} & R_{rr} & 0 \\ R_{fS} & R_{fr} & R_{ff} \end{pmatrix} \begin{pmatrix} u_t^S \\ u_t^r \\ u_t^f \end{pmatrix} \text{ where } Y_t \text{ is partitioned } \begin{pmatrix} X_{St} \\ r_t \\ X_{ft} \end{pmatrix}$$

which identifies ε_t^r 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) Classical approaches to identification

  2a) Identification by Short Run Restrictions

  2b) [Identification by Long Run Restrictions]

3) New approaches to identification (post-2000)

  3a) Identification from Heteroskedasticity

  3b) Direct Estimation of Shocks from High Frequency Data

  3c) External instruments

  3d) Identification by Sign Restrictions

2b) [Identification by Long Run Restrictions]

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

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

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

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

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

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

- There are $k(k+1)/2$ distinct equations in (*), 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 above can be rewritten,

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

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 $\text{B(L)} = \text{RA(L)}$ as above.

Comments:

- If the zero restrictions on D(1) make D(1) lower triangular, then D(1) is the Cholesky factorization of Ω.
- 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 was at the center of a debate about whether technology shocks lead to a short-run decline in hours, based on long-run restrictions (Galí (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))
- More generally, the theoretical grounding of long-run restrictions is often questionable; for a case in favor of this approach, see Giannone, Lenza, and Primiceri (2014)

Long run restrictions, ctd.

In this literature, Ω is estimated using the 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), which are essentially weak instrument concerns.
- One alternative is to use medium-run restrictions, see Uhlig (2004)

Outline

1) VARs, SVARs, and the Identification Problem

2) Classical approaches to identification

    2a) Identification by Short Run Restrictions

    2b) [Identification by Long Run Restrictions]

3) New approaches to identification (post-2000)

    3a) Identification from Heteroskedasticity

    3b) Direct Estimation of Shocks from High Frequency Data

    3c) External instruments

    3d) Identification by Sign Restrictions

3a) 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

Example: Wright (2012), Monetary Policy at ZLB

Outline

1) VARs, SVARs, and the Identification Problem
2) Classical approaches to identification
    2a) Identification by Short Run Restrictions
    2b) [Identification by Long Run Restrictions]
3) New approaches to identification (post-2000)
    3a) Identification from Heteroskedasticity
    3b) Direct Estimation of Shocks from High Frequency Data
    3c) External instruments
    3d) Identification by Sign Restrictions

3b) 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,

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

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) Classical approaches to identification
    2a) Identification by Short Run Restrictions
    2b) [Identification by Long Run Restrictions]
3) New approaches to identification (post-2000)
    3a) Identification from Heteroskedasticity
    3b) Direct Estimation of Shocks from High Frequency Data
    3c) External Instruments
    3d) Identification by Sign Restrictions

3c) External Instruments

The external instrument 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).

***Example* 1**: The Cochrane- Piazessi (2002) shock (Z^{CP}) measures the part of the monetary policy shock revealed around a FOMC announcement – but not the shock revealed at other times. If CP's identification is sound, $Z^{CP} \neq \varepsilon_t^r$ but

(i) corr(ε_t^r,Z^{CP}) $\neq 0$ (relevance)

(ii) corr(other shocks, Z^{CP}) = 0 (exogeneity)

***Example* 2**: Romer and Romer (1989, 2004, 2008); Ramey and Shapiro (1998); Ramey (2009) use the narrative approach to identify moments at which fiscal/monetary shocks occur. If identification is sound, $Z^{RR} \neq \varepsilon_t^r$ but

(i) corr(ε_t^r,Z^{RR}) $\neq 0$ (relevance)

(ii) corr(other shocks, Z^{RR}) = 0 (exogeneity)

<u>Selected empirical papers that can be reinterpreted as external instruments</u>

- **Monetary shock**: Cochrane and Piazzesi (2002), Faust, Swanson, and Wright (2003. 2004), Romer and Romer (2004), Bernanke and Kuttner (2005), Gürkaynak, Sack, and Swanson (2005)

- **Fiscal shock**: Romer and Romer (2010), Fisher and Peters (2010), Ramey (2011)

- **Uncertainty shock**: Bloom (2009), Baker, Bloom, and Davis (2011), Bekaert, Hoerova, and Lo Duca (2010), Bachman, Elstner, and Sims (2010)

- **Liquidity shocks**: Gilchrist and Zakrajšek's (2011), Bassett, Chosak, Driscoll, and Zakrajšek's (2011)

- **Oil shock**: Hamilton (1996, 2003), Kilian (2008a), Ramey and Vine (2010)

The method of External Instruments

Stock (2007), Stock and Watson (2012); Mertens and Ravn (2013);Gertler and P. Karadi (2014); for IV in VAR (not full method) see Hamilton (2003), Kilian (2009).

Additional notation: *focus on shock* 1

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

Structural errors ε_t: $\mathrm{R}u_t = \varepsilon_t$ or $u_t = \mathrm{R}^{-1}\varepsilon_t$, or $u_t = \boldsymbol{H}\varepsilon_t$

Structural MAR: $Y_t = \mathrm{A(L)}^{-1}u_t = \mathrm{C(L)}u_t = \mathrm{C(L)}H\varepsilon_t$

Partitioning notation: $u_t = H\varepsilon_t = \begin{bmatrix} H_1 & \cdots & H_r \end{bmatrix}\begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{rt} \end{pmatrix} = \begin{bmatrix} H_1 & H_\bullet \end{bmatrix}\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{\bullet t} \end{pmatrix}$

Structural MAR: $Y_t = C(\mathrm{L})H\varepsilon_t = C(\mathrm{L})H_1\varepsilon_{1t} + C(\mathrm{L})H_\bullet\varepsilon_{\bullet t}$

Structural MAR for j^{th} variable: $Y_{jt} = \sum_{k=0}^{\infty} C_{k,j}H_1\varepsilon_{1t-k} + \sum_{k=0}^{\infty} \overset{1\times r}{C}_{k,j}\, H_\bullet\varepsilon_{\bullet t-k}$

Identification of H_1

$$A(L)Y_t = u_t, \quad u_t = H\varepsilon_t = \begin{bmatrix} H_1 & \cdots & H_r \end{bmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{rt} \end{pmatrix}$$

Suppose you have k instrumental variables Z_t (not in Y_t) such that

(i) $E\left(\varepsilon_{1t} Z_t'\right) = \alpha' \neq 0$ (relevance)

(ii) $E\left(\varepsilon_{jt} Z_t'\right) = 0$, $j = 2,\ldots, r$ (exogeneity)

(iii) $E\left(\varepsilon_t \varepsilon_t'\right) = \Sigma_{\varepsilon\varepsilon} = D = diag(\sigma^2_{\varepsilon_1},...,\sigma^2_{\varepsilon_r})$

Under (i) and (ii), you can identify H_1 up to sign & scale

$$E(u_t Z_t') = E(H\varepsilon_t Z_t') = \begin{bmatrix} H_1 & \cdots & H_r \end{bmatrix} \begin{pmatrix} E(\varepsilon_{1t} Z_t') \\ \vdots \\ E(\varepsilon_{rt} Z_t') \end{pmatrix} = \begin{bmatrix} H_1 & \cdots & H_r \end{bmatrix} \begin{pmatrix} \alpha' \\ 0 \\ 0 \end{pmatrix} = H_1 \alpha'$$

Identification of H_1, ctd.

$$E(u_t Z_t') = E(H\varepsilon_t Z_t') = \begin{bmatrix} H_1 & H_\bullet \end{bmatrix} \begin{pmatrix} E(\varepsilon_{1t} Z_t') \\ E(\varepsilon_{\bullet t} Z_t') \end{pmatrix} = H_1 \alpha'$$

<u>Normalization</u>

- The scale of H_1 and $\sigma^2_{\varepsilon_1}$ is set by a normalization subject to

$$\Sigma_{uu} = HDH' \quad \text{where } D = diag(\sigma^2_{\varepsilon_1}, ..., \sigma^2_{\varepsilon_r})$$

- Normalization used here: a unit positive value of shock 1 is defined to have a unit positive effect on the innovation to variable 1, which is u_{1t}. This corresponds to:

(iv) $H_{11} = 1$ (unit shock normalization)

where H_{11} is the first element of H_1

Identification of H_1, ctd.

Impose normalization (iv):

$$E(u_t Z_t') = \begin{pmatrix} Eu_{1t}Z_t' \\ Eu_{\bullet t}Z_t' \end{pmatrix} = H_1\alpha' = \begin{pmatrix} H_{11} \\ H_{1\bullet} \end{pmatrix}\alpha' = \begin{pmatrix} 1 \\ H_{1\bullet} \end{pmatrix}\alpha'$$

So

$$\begin{pmatrix} H_{1\bullet}Eu_{1t}Z_t' \\ Eu_{\bullet t}Z_t' \end{pmatrix} = \begin{pmatrix} H_{1\bullet}\alpha' \\ H_{1\bullet}\alpha' \end{pmatrix}$$

or

$$H_{1\bullet}Eu_{1t}Z_t' = Eu_{\bullet t}Z_t'$$

If Z_t is a scalar (k = 1): $$H_{1\bullet} = \frac{Eu_{\bullet t}Z_t}{Eu_{1t}Z_t}$$

Identification of ε_{1t}

$$\varepsilon_t = H^{-1}u_t = \begin{bmatrix} H^{1\prime} \\ \vdots \\ H^{r\prime} \end{bmatrix} u_t$$

- Identification of first column of H and $\Sigma_{\varepsilon\varepsilon} = D$ identifies first row of H^{-1} up to scale (can show via partitioned matrix inverse formula).

- Alternatively, let Φ be the coefficient matrix of the population regression of Z_t onto u_t:

$$\Phi = E(Z_t u_t{}')\Sigma_u^{-1} = \alpha H_1{}'(HDH')^{-1} = \alpha H_1{}' H'^{-1} D^{-1} H^{-1} = (\alpha/\sigma_{\varepsilon_1}^2)H^{1\prime}$$

because $H^{-1}H_1 = (1\ 0\ \ldots\ 0)'$. Thus ε_{1t} is identified up to scale by

$$\Phi u_t = \frac{\alpha}{\sigma_{\varepsilon_1}^2} H^{1\prime} u_t = \frac{\alpha}{\sigma_{\varepsilon_1}^2}\varepsilon_{1t}$$

Identification of ε_{1t}, ctd

Φu_t is the predicted value from the population projection of Z_t on η_t:

$$\tilde{\varepsilon}_{1t} = \Phi u_t = E(Z_t u_t')\Sigma_u^{-1} u_t = \frac{\alpha}{\sigma_{\varepsilon_1}^2}\varepsilon_{1t}$$

- Φ has rank 1 (in population), so this is a (population) reduced rank regression
- 2 instruments identify 2 shocks. Suppose they are shocks 1 and 2, identified by Z_{1t} and Z_{2t}. Then

$$E(\tilde{\varepsilon}_{1t}\tilde{\varepsilon}_{2t}) = E(Z_{1t}u_t')\Sigma_u^{-1}E(u_t Z_{2t})$$

which = 0 if both instruments satisfy (i) – (iii)

Estimation

Recall notation: $H_1 = \begin{bmatrix} H_{11} \\ H_{1\bullet} \end{bmatrix}$, $u_t = \begin{bmatrix} u_{1t} \\ u_{\bullet t} \end{bmatrix}$

Impose the normalization condition (iv) $H_{11} = 1$, so

$$E(u_t Z_t') = H_1 \alpha' = \begin{pmatrix} 1 \\ H_{1\bullet} \end{pmatrix} \alpha \text{ or } E(u_t \otimes Z_t) = \begin{pmatrix} 1 \\ H_{1\bullet} \end{pmatrix} \otimes \alpha$$

High level assumption (assume throughout)

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left([u_t \otimes Z_t] - [H_1 \otimes \alpha] \right) \xrightarrow{d} \mathrm{N}(0,\Omega)$$

Estimation of H_1

Efficient GMM objective function:

$S(H_{1\bullet},\alpha;\hat{\Omega})$

$$= \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left((\hat{u}_t \otimes Z_t) - (\begin{bmatrix} 1 \\ H_{1\bullet} \end{bmatrix} \otimes \alpha)\right)\right)' \hat{\Omega}^{-1} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left((\hat{u}_t \otimes Z_t) - (\begin{bmatrix} 1 \\ H_{1\bullet} \end{bmatrix} \otimes \alpha)\right)$$

k = 1 (exact identification): $E(u_t Z_t') = H_1\alpha' = \begin{pmatrix} \alpha \\ \alpha H_{1\bullet} \end{pmatrix}$

so GMM estimator solves, $T^{-1}\sum_{t=1}^{T}\hat{u}_t Z_t = \begin{pmatrix} \hat{\alpha} \\ \hat{\alpha}\hat{H}_{1\bullet} \end{pmatrix}$

GMM estimator: $$\hat{H}_{1\bullet} = \frac{T^{-1}\sum_{t=1}^{T}\hat{u}_{\bullet t} Z_t}{T^{-1}\sum_{t=1}^{T}\hat{u}_{1t} Z_t}$$

IV interpretation: $\hat{u}_{jt} = H_{1j}\hat{u}_{1t} + u_{jt}$,

$\hat{u}_{1t} = \Pi_j' Z_t + v_{jt}$

GMM estimation of $H^{1\prime}$ and ε_{1t}

Recall $$\tilde{\varepsilon}_{1t} = E(Z_t u_t{}')\Sigma_u^{-1} u_t = \Phi u_t$$

<u>Estimator:</u>

- $k = 1$:

  $\hat{\varepsilon}_{1t}$ is the predicted value (up to scale) in the regression of Z_t on $\hat{u}_t$

- $k > 1$(no-HAC):

  Absent serial correlation/no heteroskedasticity, the GMM estimator simplifies to reduced rank regression:

$$Z_t = \Phi \hat{u}_t + \nu_t \qquad \text{(RRR)}$$

- If Z_t is available only for a subset of time periods, estimate (RRR) using available data, compute predicted value over full period

Strong instrument asymptotics

- $k = 1$ case:

$$\sqrt{T}\left(\hat{H}_{1\bullet} - H_{1\bullet}\right) \xrightarrow{d} \text{N}(0, \Gamma'\Omega\Gamma), \text{ where } \Gamma = \begin{bmatrix} -H_{1\bullet}{}' \\ I_{r-1} \end{bmatrix}$$

- Overidentified case ($k > 1$):
  - usual GMM formula
  - J-statistics, etc. are standard textbook GMM

Weak instrument asymptotics: k = 1

(Stock and Watson (2012b)) Weak IV asymptotic setup – local drift (limit of experiments, etc.):

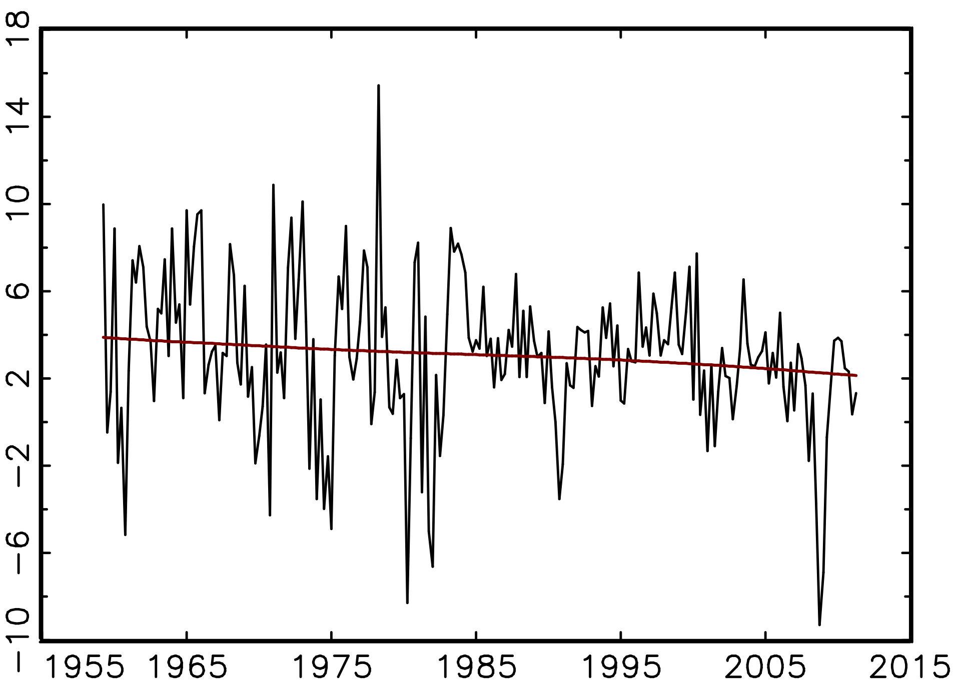
$$\alpha = \alpha_T = a/\sqrt{T}$$

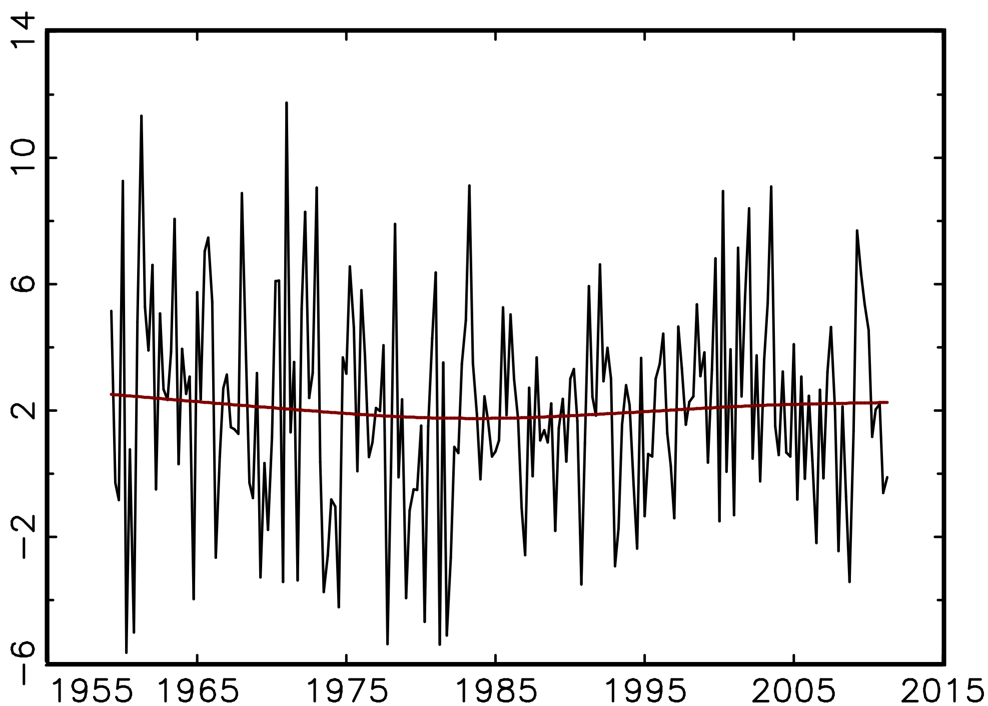
Obtain weak instrument distribution

Empirical Application: Stock-Watson (BPEA, 2012)

Dynamic factor model identified by external instruments:

- U.S., quarterly, 1959-2011Q2, 200 time series
- Almost all series analyzed in changes or growth rates
- All series detrended by local demeaning – approximately 15 year centered moving average:





Quarterly GDP growth (a.r.) — Quarterly productivity growth

Trend: 3.7% → 2.5% — 2.3% → 1.8% → 2.2%

Instruments

1. Oil Shocks
    a. Hamilton (2003) net oil price increases
    b. Killian (2008) OPEC supply shortfalls
    c. Ramey-Vine (2010) innovations in adjusted gasoline prices

2. Monetary Policy
    a. Romer and Romer (2004) policy
    b. Smets-Wouters (2007) monetary policy shock
    c. Sims-Zha (2007) MS-VAR-based shock
    d. Gürkaynak, Sack, and Swanson (2005), FF futures market

3. Productivity
    a. Fernald (2009) adjusted productivity
    b. Gali (200x) long-run shock to labor productivity
    c. Smets-Wouters (2007) productivity shock

Instruments, ctd.

4. Uncertainty
    a. VIX/Bloom (2009)
    b. Baker, Bloom, and Davis (2009) Policy Uncertainty

5. Liquidity/risk
    a. Spread: Gilchrist-Zakrajšek (2011) excess bond premium
    b. Bank loan supply: Bassett, Chosak, Driscoll, Zakrajšek (2011)
    c. TED Spread

6. Fiscal Policy
    a. Ramey (2011) spending news
    b. Fisher-Peters (2010) excess returns gov. defense contractors
    c. Romer and Romer (2010) “all exogenous” tax changes.

"First stage": F_1: regression of Z_t on u_t, F_2: regression of u_{1t} on Z_t

Structural Shock	F_1	F_2
1. Oil		
Hamilton	2.9	**15.7**
Killian	1.1	1.6
Ramey-Vine	1.8	0.6
2. Monetary policy		
Romer and Romer	4.5	**21.4**
Smets-Wouters	9.0	5.3
Sims-Zha	6.5	**32.5**
GSS	0.6	0.1
3. Productivity		
Fernald TFP	**14.5**	**59.6**
Smets-Wouters	7.0	**32.3**
Structural Shock	F_1	F_2
4. Uncertainty		
Fin Unc (VIX)	**43.2**	**239.6**
Pol Unc (BBD)	**12.5**	73.1

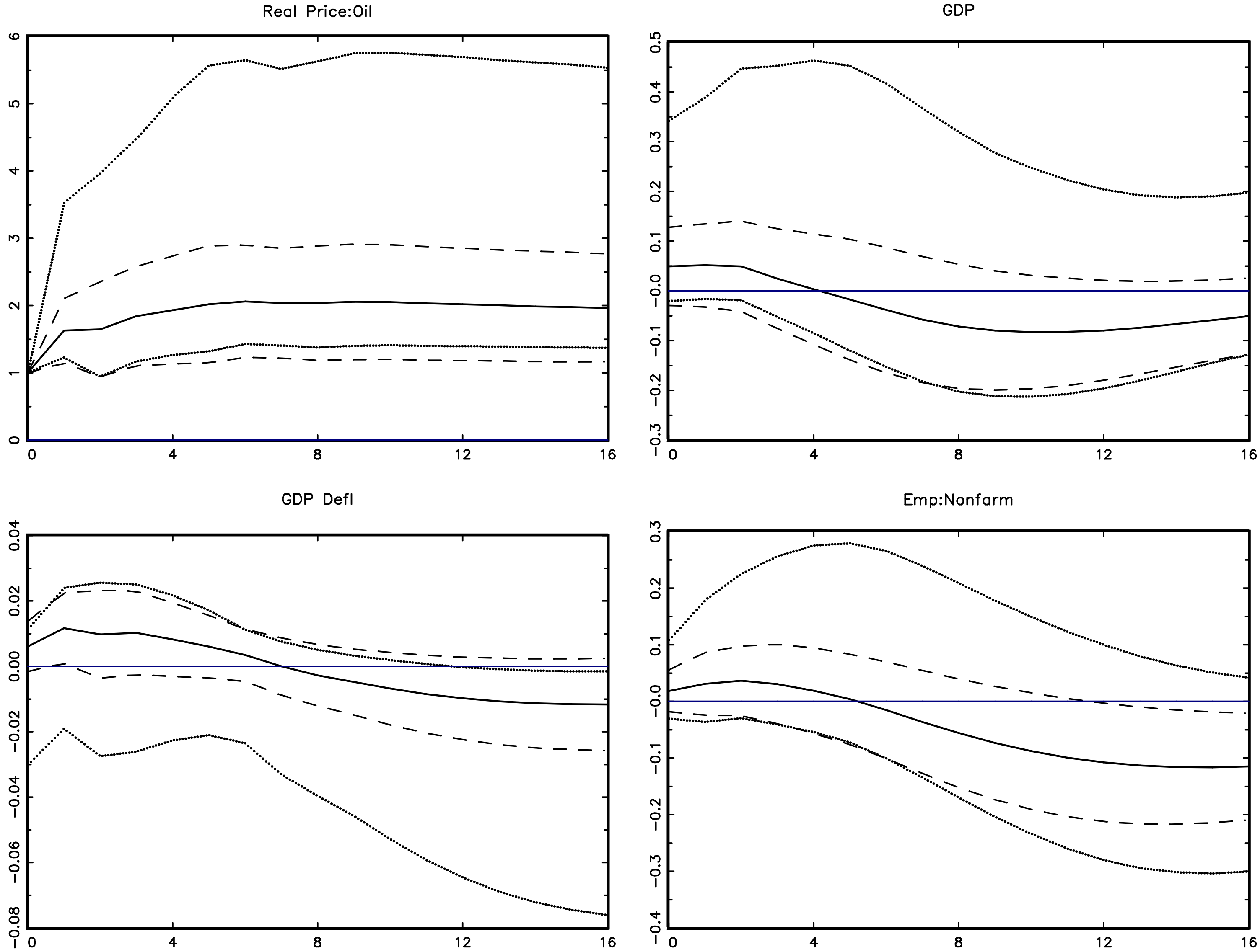
5. Liquidity/risk	F_1	F_2
GZ EBP Spread	4.5	**23.8**
TED Spread	**12.3**	**61.1**
BCDZ Bank Loan	4.4	4.2
6. Fiscal policy		
Ramey Spending	0.5	1.0
Fisher-Peters Spending	1.3	0.1
Romer-Romer Taxes	0.5	2.1

Correlations among selected structural shocks

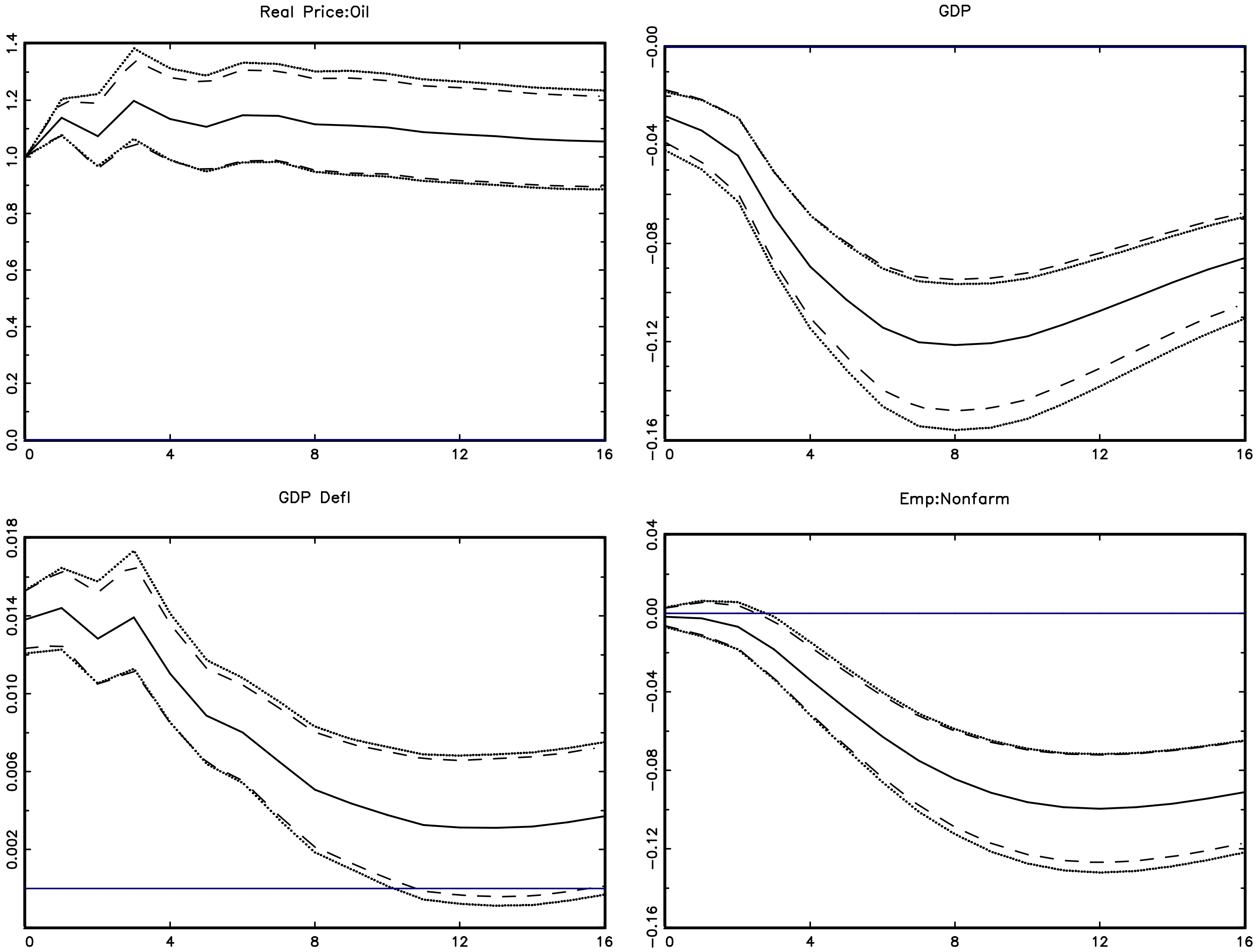
	O_K	M_{RR}	M_{SZ}	P_F	U_B	U_{BBD}	S_{GZ}	B_{BCDZ}	F_R	F_{RR}
O_K	1.00									
M_{RR}	**0.65**	1.00								
M_{SZ}	0.35	0.93	1.00							
P_F	0.30	0.20	0.06	1.00						
U_B	-0.37	-0.39	-0.29	0.19	1.00					
U_{BBD}	0.11	-0.17	-0.22	-0.06	0.78	1.00				
L_{GZ}	-0.42	-0.41	-0.24	0.07	**0.92**	**0.66**	1.00			
L_{BCDZ}	0.22	0.56	0.55	-0.09	**-0.69**	-0.54	-0.73	1.00		
F_R	**-0.64**	**-0.84**	**-0.72**	-0.17	0.26	-0.08	0.40	-0.13	1.00	
F_{RR}	0.15	**0.77**	**0.88**	0.18	0.01	-0.10	0.02	0.19	-0.45	1.00

Oil_{Kilian}	oil – Kilian (2009)
M_{RR}	monetary policy – Romer and Romer (2004)
M_{SZ}	monetary policy – Sims-Zha (2006)
P_F	productivity – Fernald (2009)
U_B	Uncertainty – VIX/Bloom (2009)
U_{BBD}	uncertainty (policy) – Baker, Bloom, and Davis (2012)
L_{GZ}	liquidity/risk – Gilchrist-Zakrajšek (2011) excess bond premium
L_{BCDZ}	liquidity/risk – BCDZ (2011) SLOOS shock
F_R	fiscal policy – Ramey (2011) federal spending
F_{RR}	fiscal policy – Romer-Romer (2010) federal tax

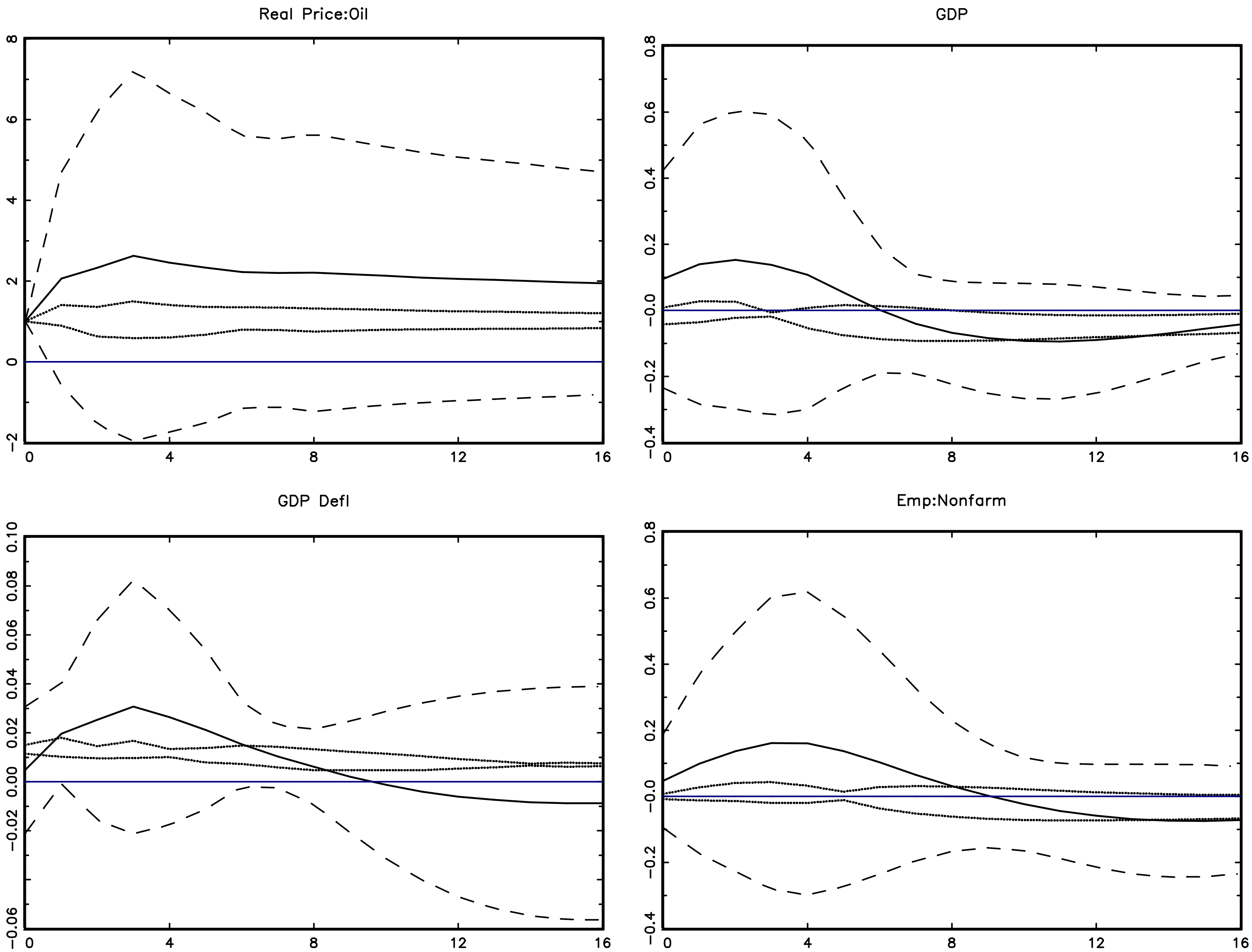
IRFs: strong-IV (dashed) and weak-IV robust (solid) pointwise bands



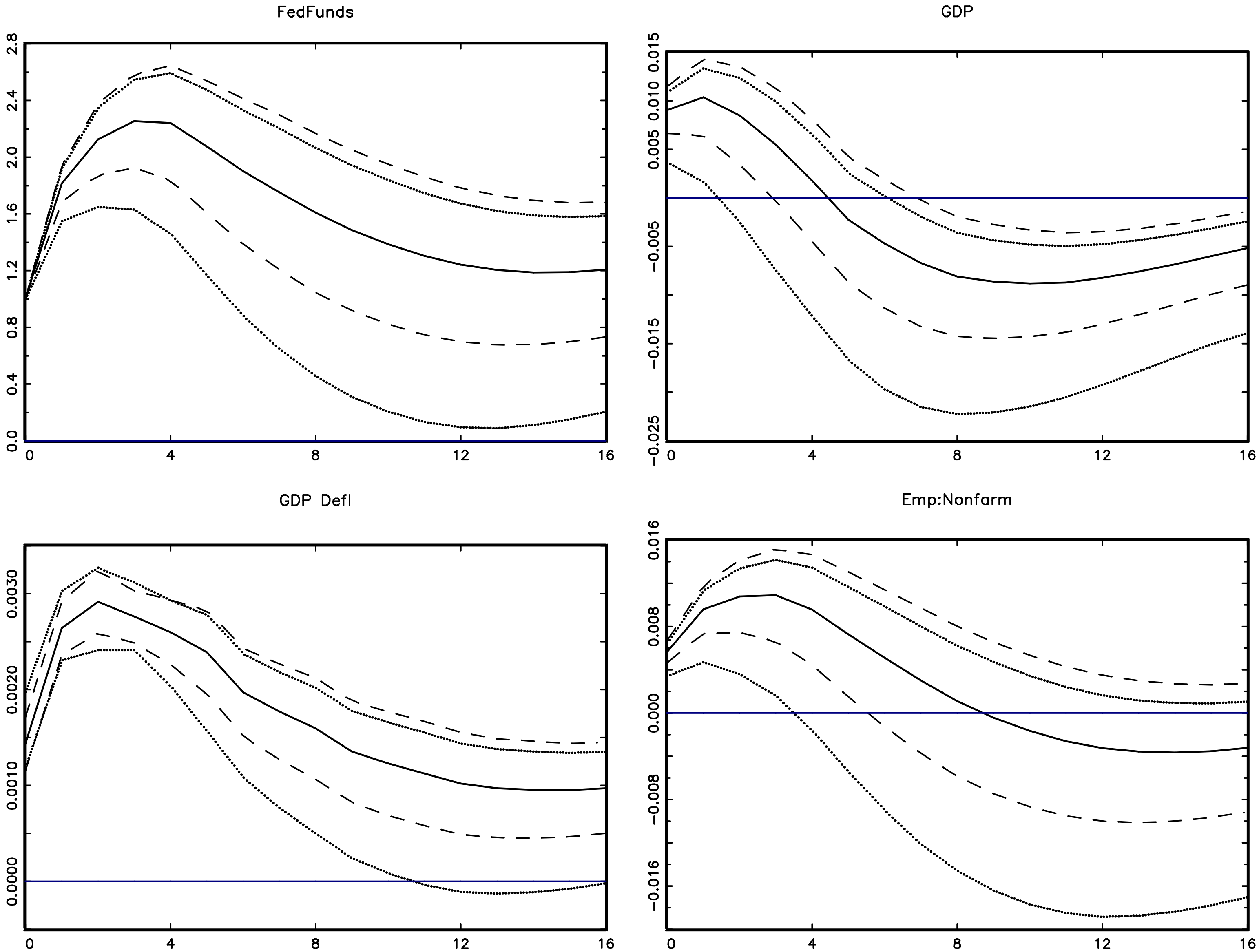
Kilian (2008) oil shock (F_2 = 1.6)



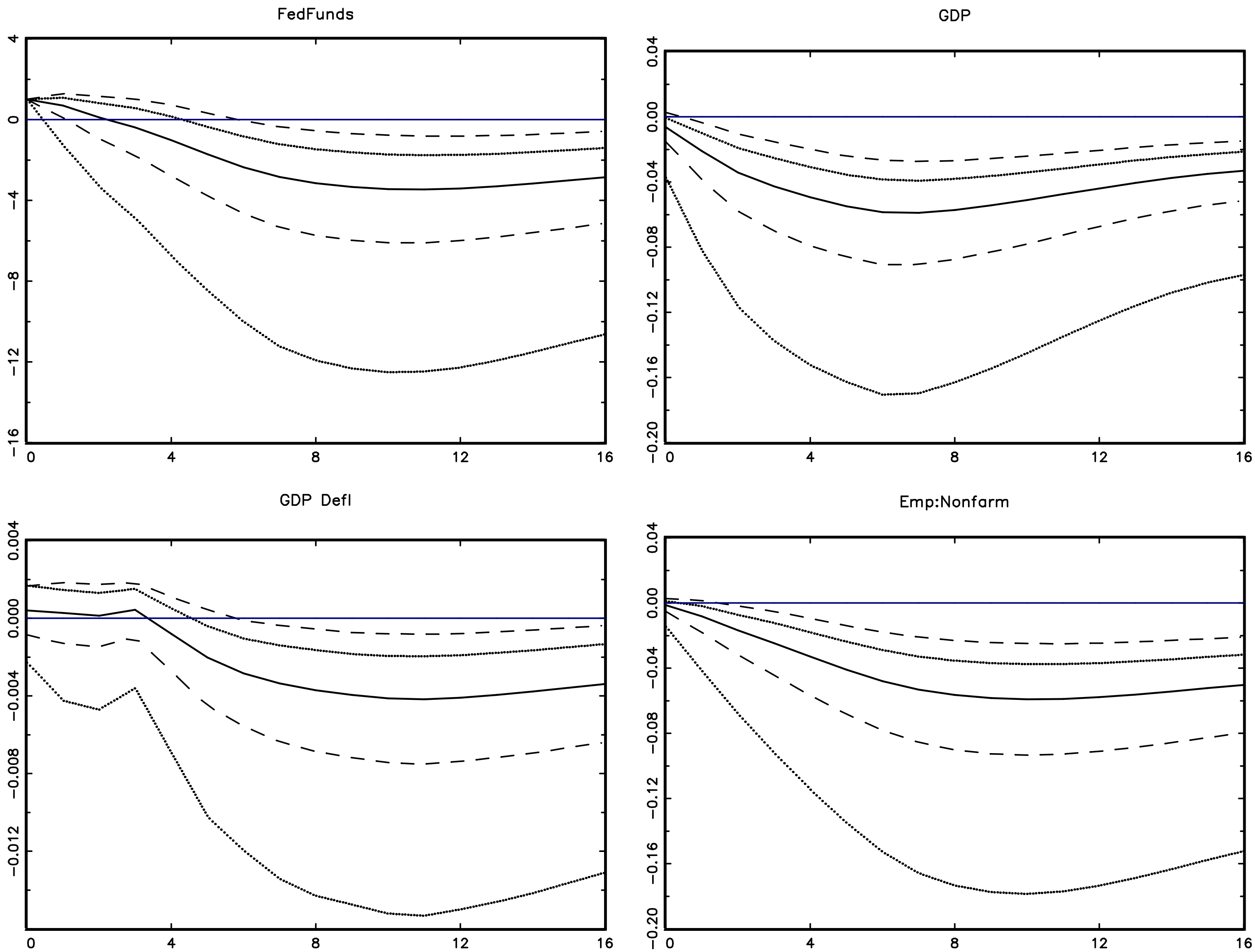
Hamilton (1996, 2003) oil shock (F_2 = 15.7)



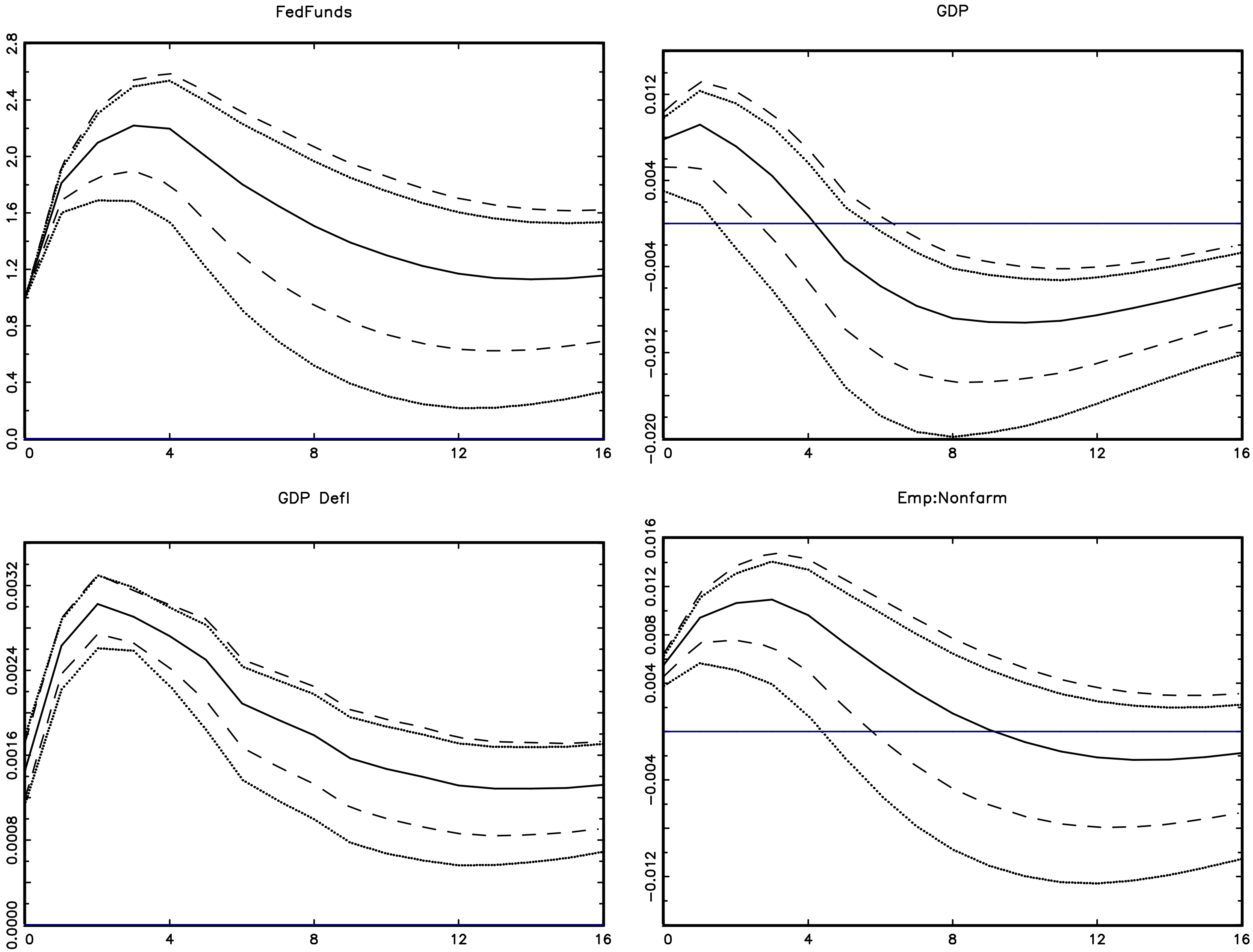
Ramey-Vine (2010) oil shock (F_2 = 0.6)



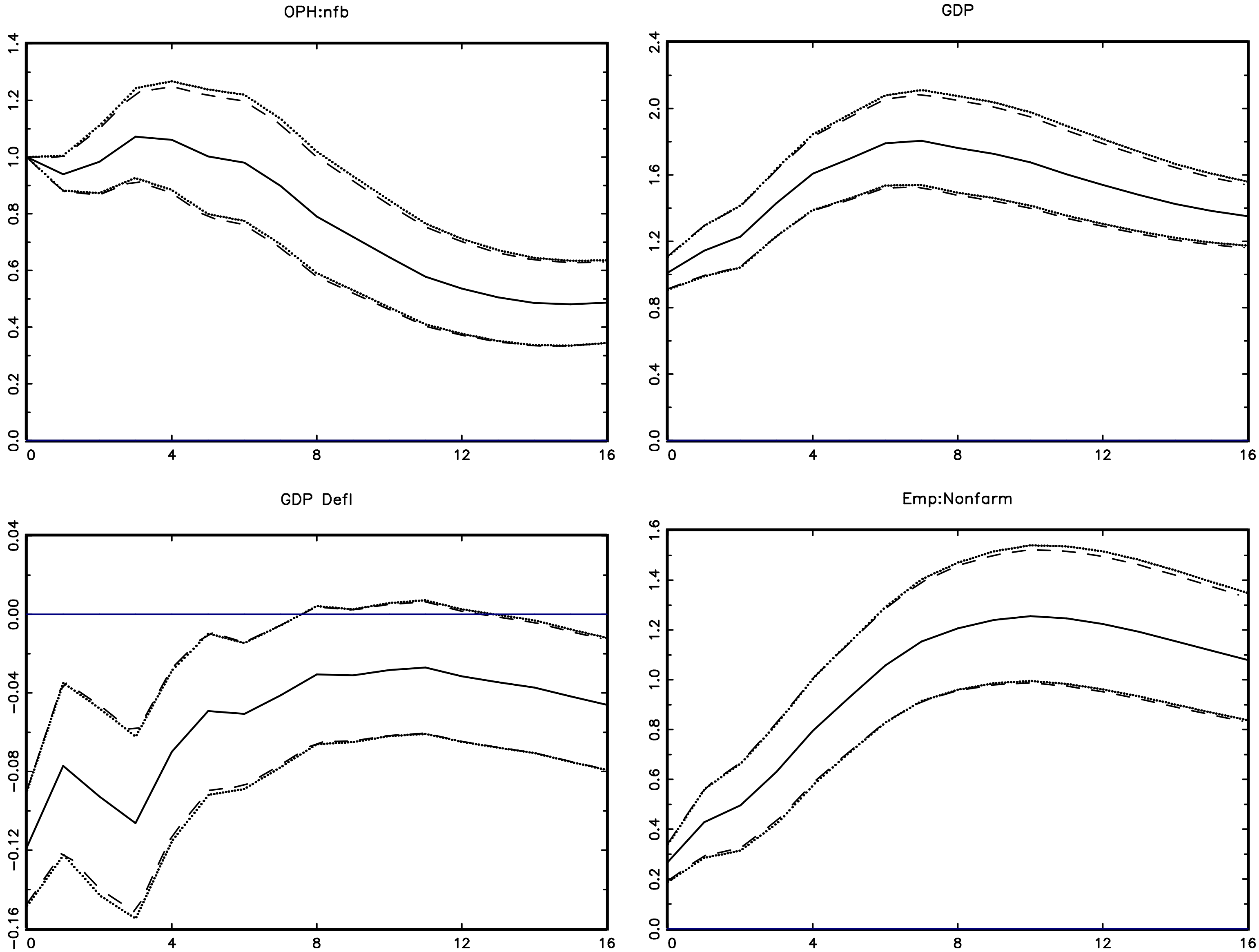
Romer and Romer (2004) monetary policy shock (F_2 = 21.4)



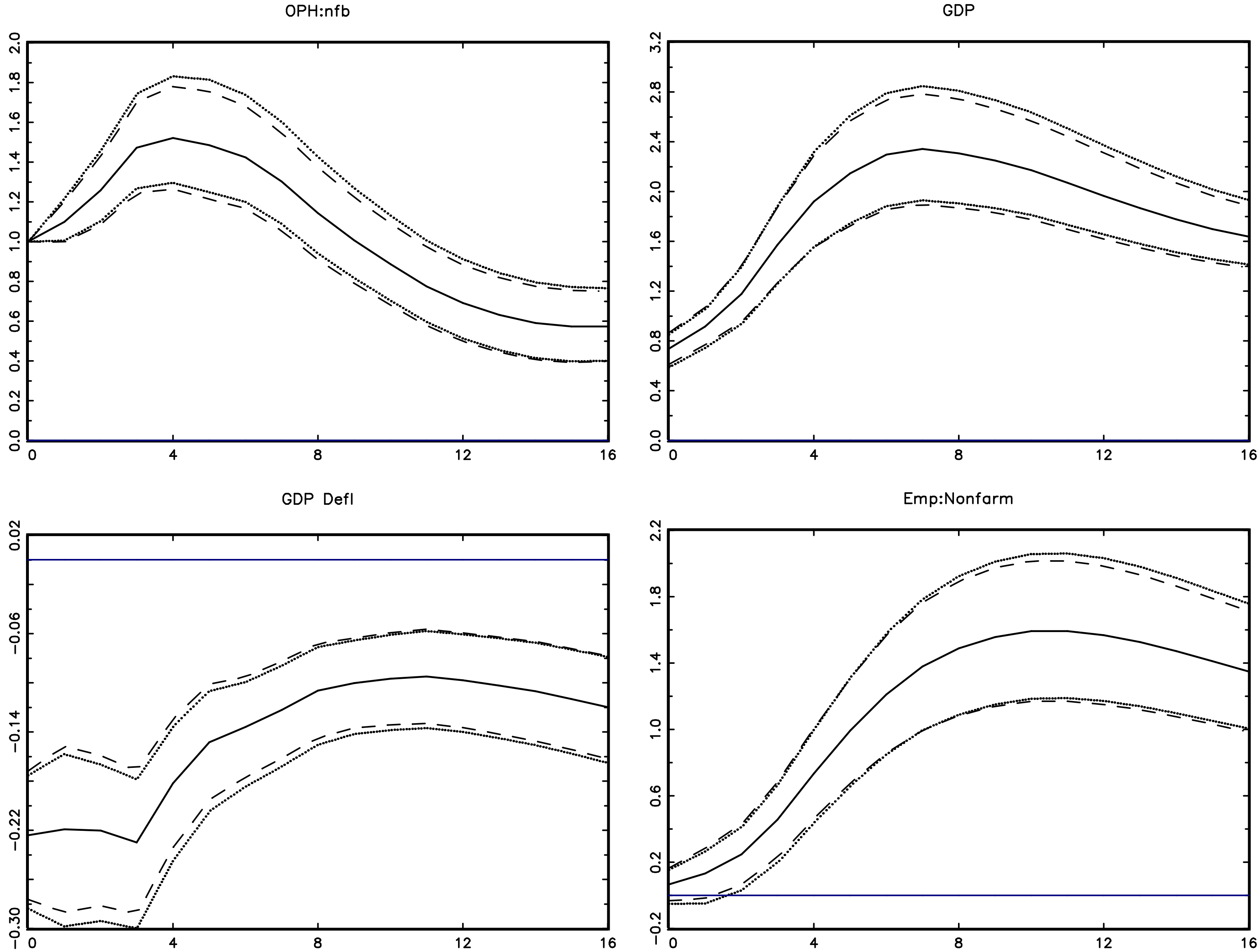
Smets-Wouters (2007) monetary policy shock ($F_2 = 5.3$)



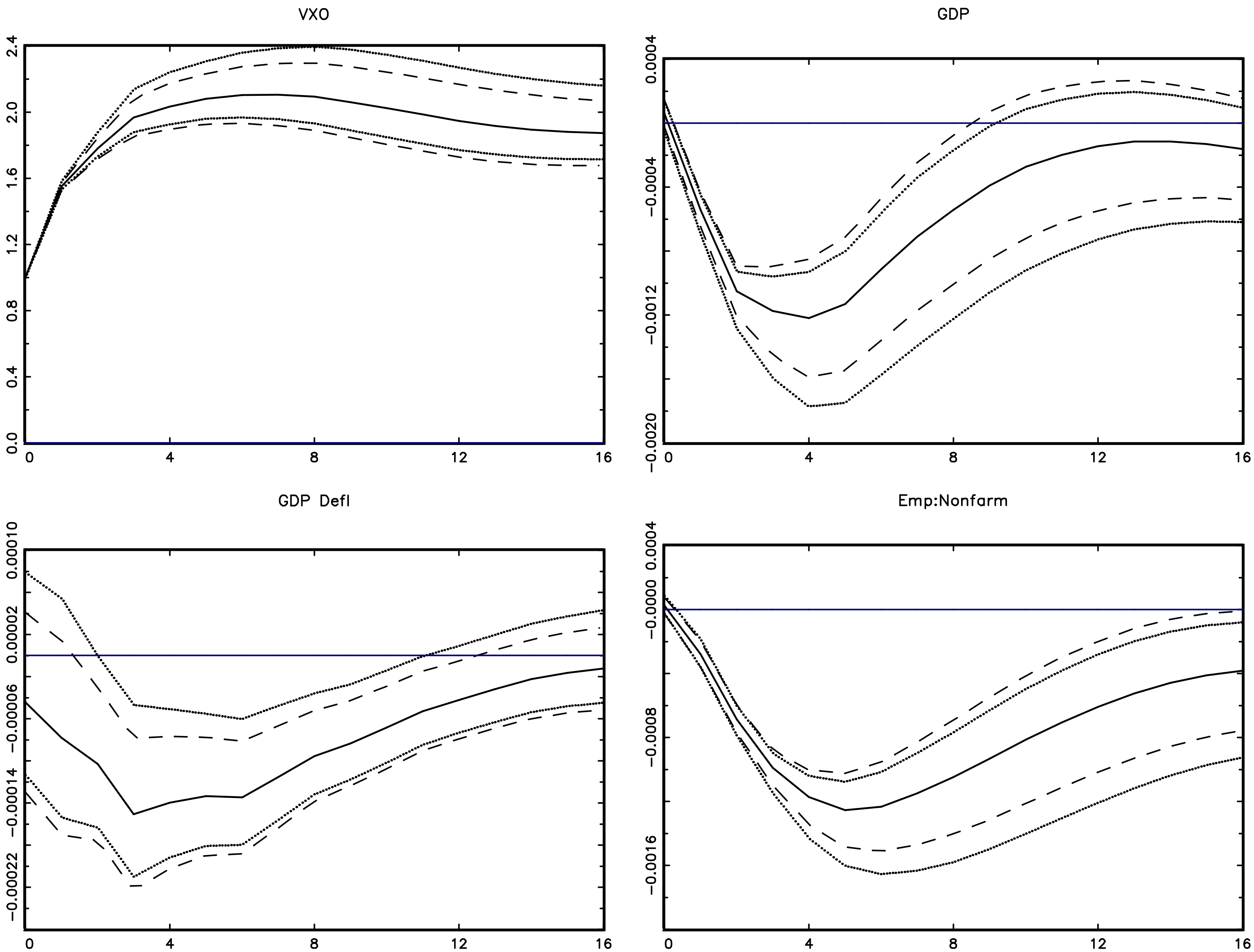
Sims-Zha (2006) monetary policy shock ($F_2 = 32.5$)



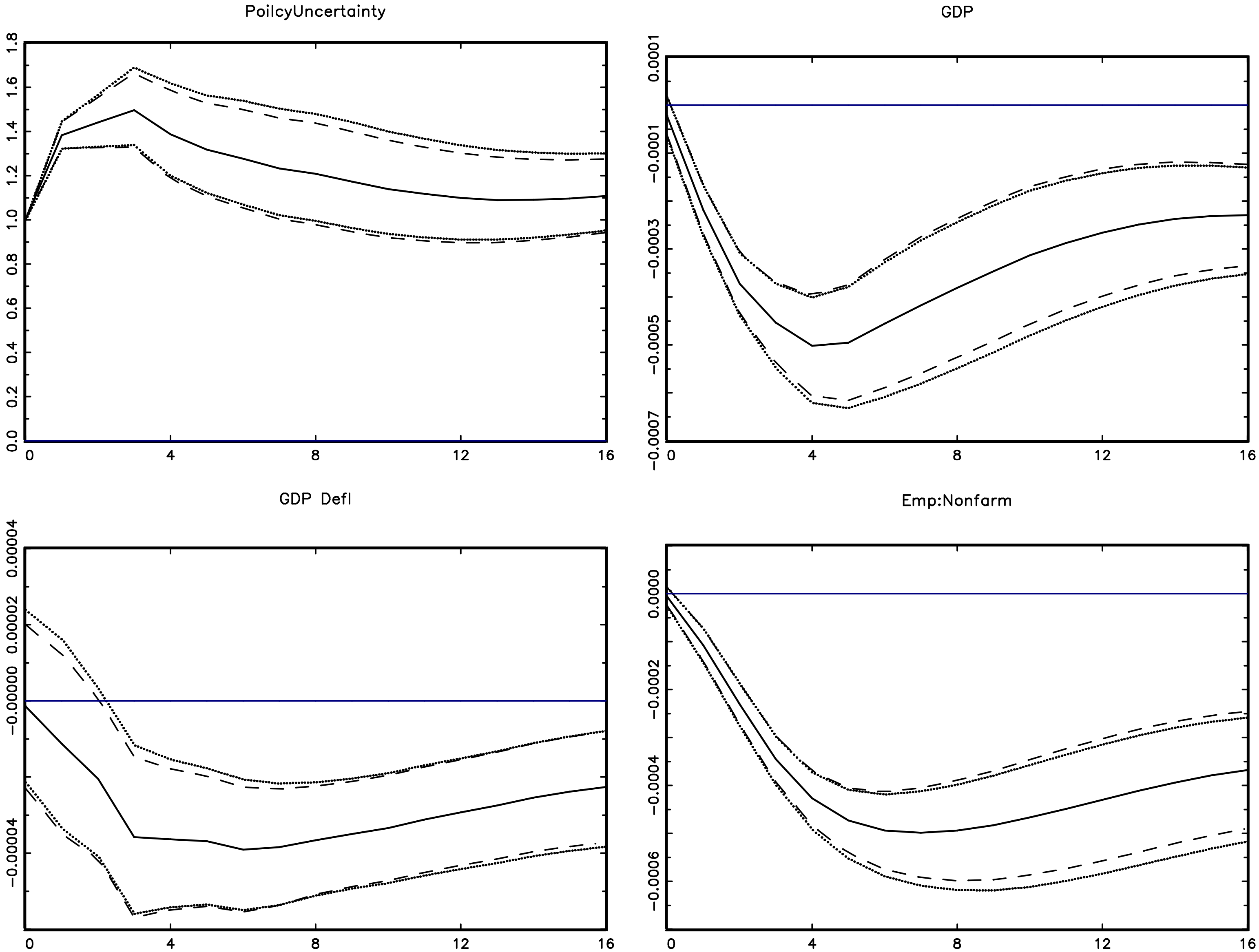
Fernald (2009) productivity shock ($F_2 = 59.6$)



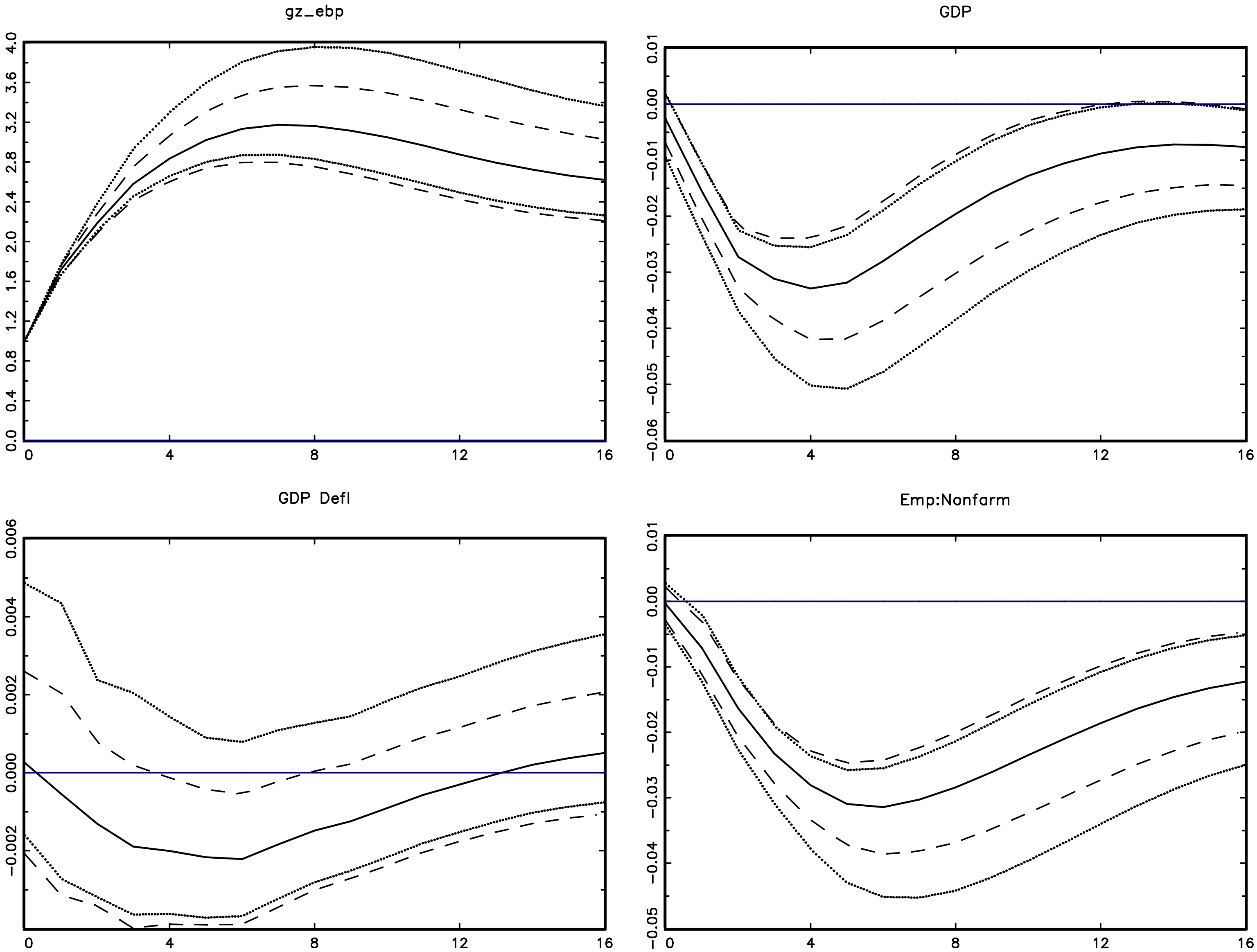
Smets-Wouters (2007) productivity shock (F_2 = 32.3)



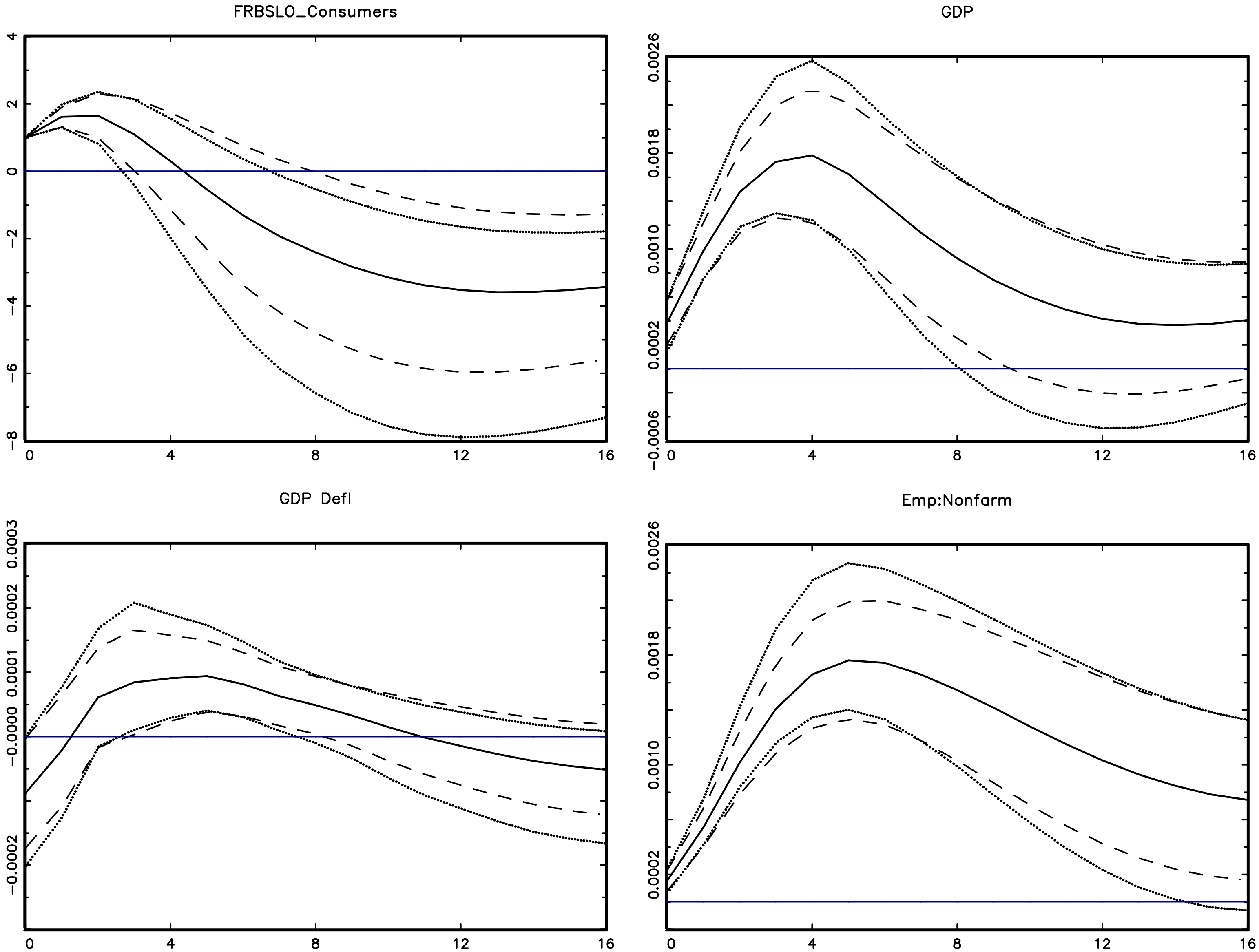
Bloom (2009) (VIX) uncertainty shock (F_2 = 239.6)



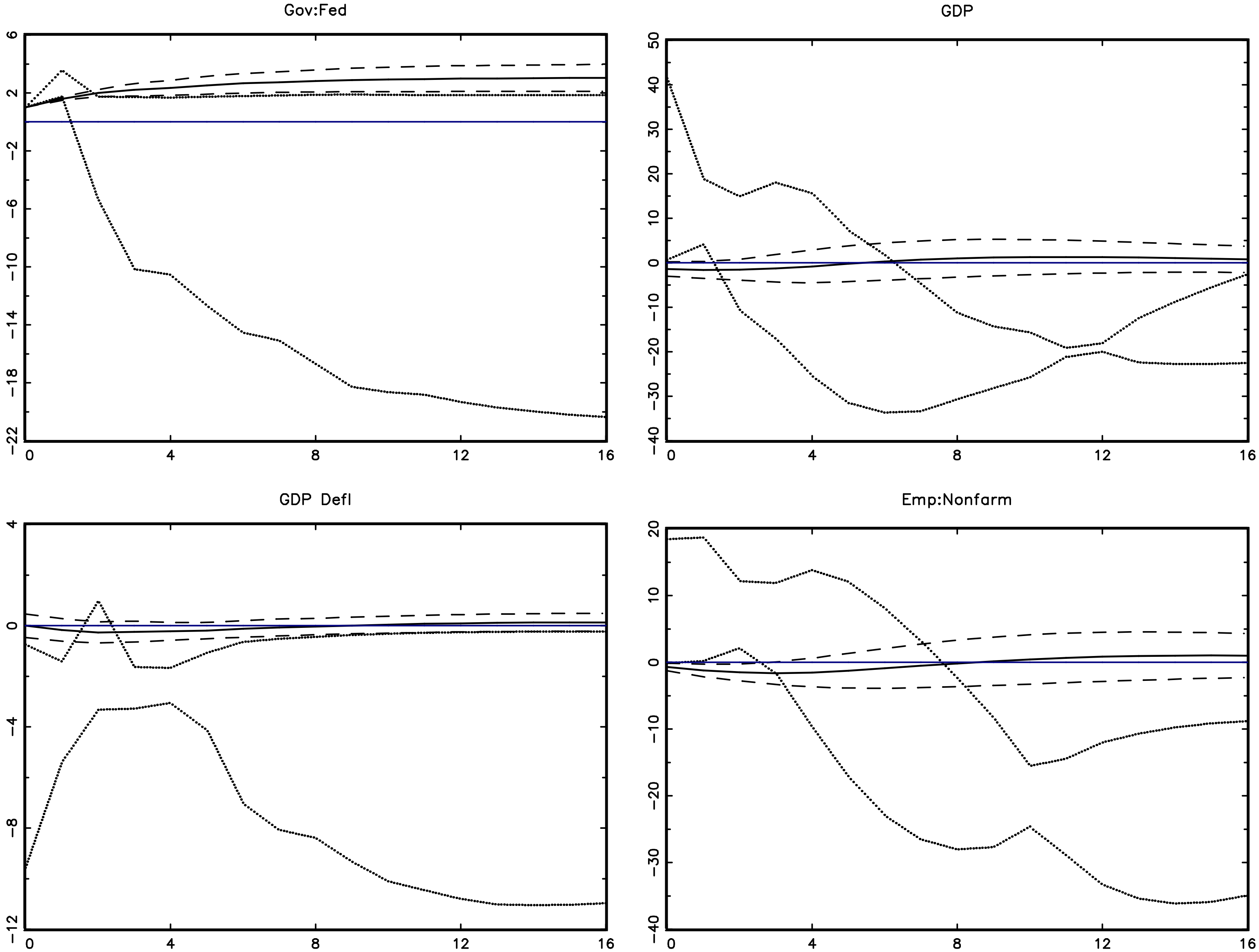
Baker, Bloom, Davis (2012) policy uncertainty shock (F_2 = 73.1)



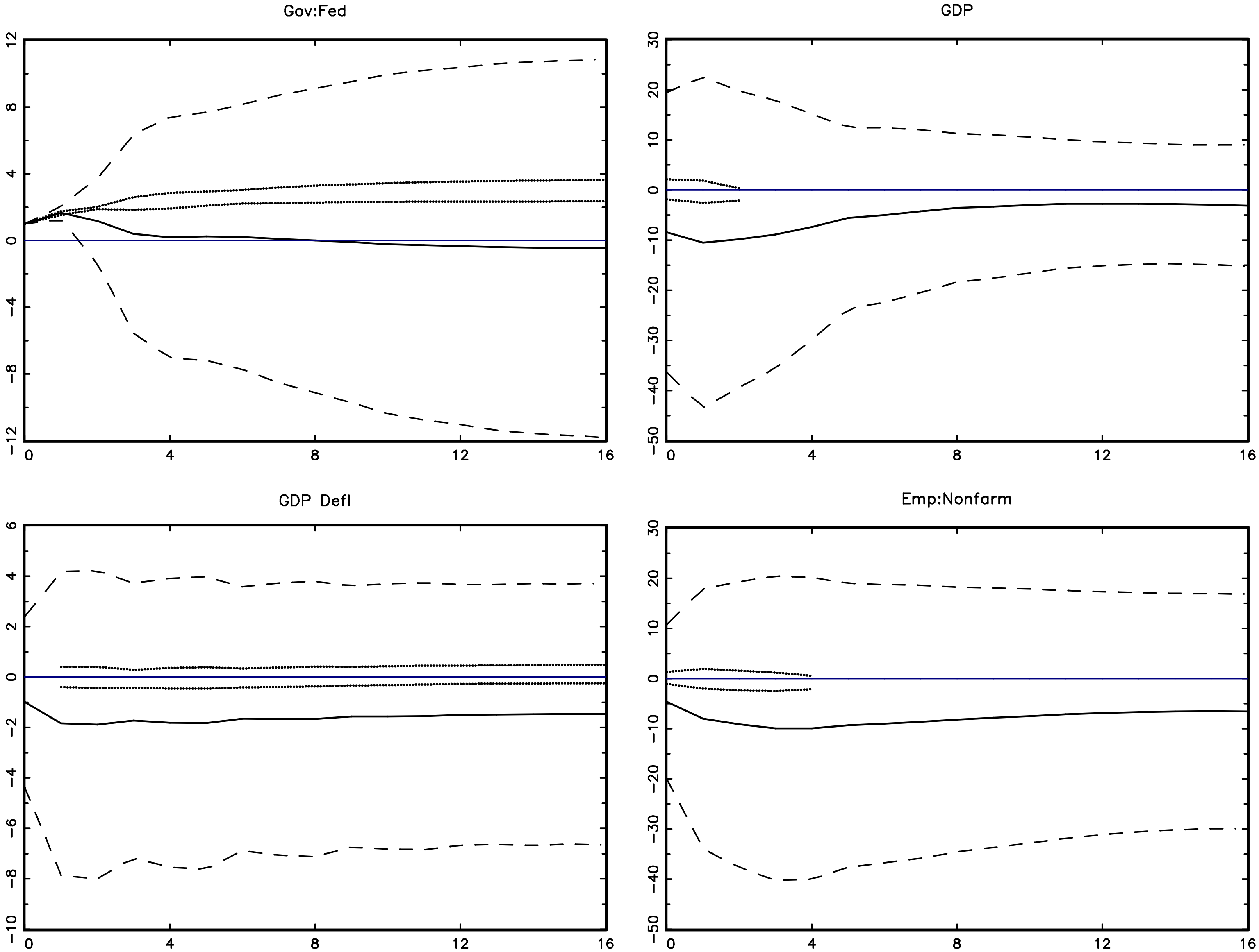
Gilchrist and Zakrajšek (2011) excess bond premium liquidity/risk shock (F_2 = 23.8)



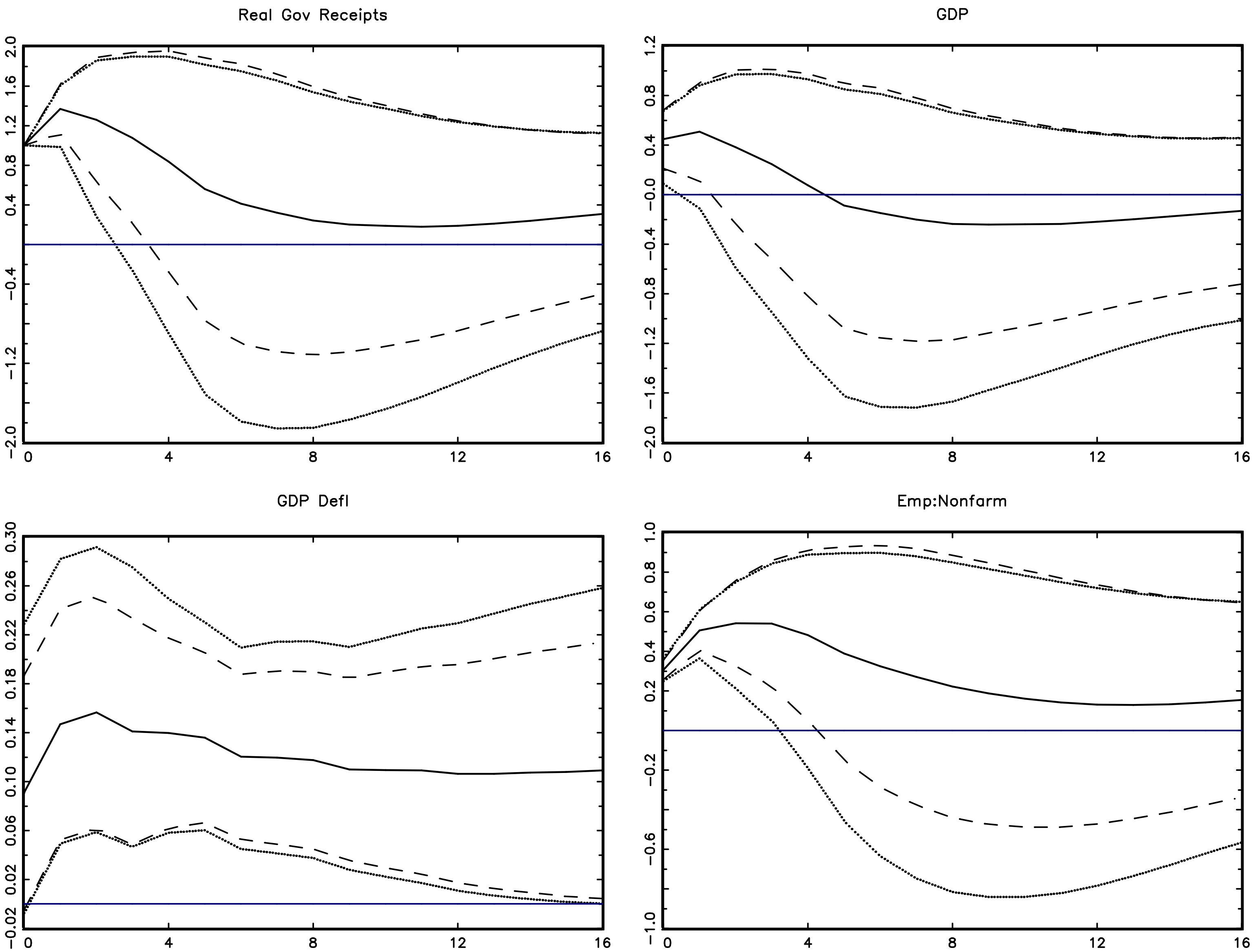
Bassett, Chosak, Driscoll, and Zakrajšek (2011) bank loan supply liquidity/risk shock (F_2 = 4.2)



Ramey (2011) fiscal (spending) shock (F_2 = 1.0)



Fisher and Peters (2010) fiscal (spending) shock ($F_2 = 0.1$)



Romer and Romer (2010) fiscal (tax) schock (F_2 = 2.1)

Outline

1) VARs, SVARs, and the Identification Problem

2) Classical approaches to identification

 2a) Identification by Short Run Restrictions

 2b) [Identification by Long Run Restrictions]

3) New approaches to identification (post-2000)

 3a) Identification from Heteroskedasticity

 3b) Direct Estimation of Shocks from High Frequency Data

 3c) External Instruments

 3d) Identification by Sign Restrictions

4) Inference: Challenges and Recently Developed Tools

3d) 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 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, which is the most popular method (the first steps are the same as #3; #1 has only been used a few times)

Sign restrictions, ctd.

It is useful to rewrite the identification problem after normalizing by a Cholesky factorization (and setting $\Sigma_\varepsilon = I$):

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

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

Where $R_c^{-1} = Chol(\Sigma_u)$ and Q is a $n\times n$ orthonormal matrix so $QQ' = I$. Then

Structural errors: $u_t = R_c^{-1}Q\varepsilon_t$

Structural IRF: $\mathrm{D(L)} = \mathrm{C(L)}R_c^{-1}Q$

Let **D** denote the set of acceptable IRFs (IRFs that satisfy the sign restrictions)

Sign restrictions, ctd.

Structural IRF: $\mathrm{D(L)} = \mathrm{C(L)}R_c^{-1}Q$

Uhlig's algorithm (slightly modified):

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

(ii) Compute the IRF $\tilde{D}(L) = \mathrm{D(L)} = \mathrm{C(L)}R_c^{-1}\tilde{Q}$

(iii) If $\tilde{D}(L) \notin \mathbf{D}$, discard this trial $\tilde{Q}$ and go to (i). Otherwise, if $\tilde{D}(L) \in \mathrm{D}$, retain $\tilde{Q}$ then go to (i)

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

This algorithm implements Bayes inference using a prior proportional to

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

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

n = 2 example

Consider a n = 2 VAR: A(L)$Y_t = u_t$ and structural IRF

$$\mathrm{D(L)} = \begin{pmatrix} D_{11}(L) & D_{12}(L) \\ D_{21}(L) & D_{22}(L) \end{pmatrix} = \mathrm{A(L)}^{-1} R_c^{-1} Q.$$

The sign restriction is $D_{21,I} \geq 0$, $I = 1,\ldots, 4$ (shock 1 has a positive effect on variable 2 for the first 4 quarters).

Suppose the population reduced form VAR is A(L)$Y_t = u_t$ where

$$\mathrm{A(L)} = \begin{pmatrix} (1-\alpha_1 L)^{-1} & 0 \\ 0 & (1-\alpha_2 L)^{-1} \end{pmatrix} \text{ and } \Sigma_u = I \text{ so } R_c^{-1} = I.$$

What does set-identified Bayesian inference look like for this problem, in a large sample?

- With point-identified inference and nondogmatic priors, it looks like frequentist inference (Bernstein-von Mises theorem)

n = 2 example, ctd.

Step 1: use n =2 to characterize Q

In the n = 2 case, the restriction QQ' = I implies that there is only one free parameter in Q, so that all orthonormal Q can be written,

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ [check: } \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = I\text{]}$$

- The standard method, used here, is to draw Q by drawing $\theta \sim U[0,2\pi]$
- The main point of this example is that the uniform prior on θ ends up being informative for what matters, D(L), so much so that the prior induced a Bayesian posterior coverage region strictly inside the identified set.

Step 2: Condition for checking whether Q is retained:

$$\hat{D}_{21}(L) = \left[\hat{A}(L)^{-1}\hat{R}_c^{-1}Q\right]_{21} \geq 0 \text{ for first 4 lags}$$

Step 3: In a very large sample, A(L) and Σ_u will be essentially known (WLLN), so that

$$\hat{A}(L)^{-1}\hat{R}_c^{-1}Q \approx \begin{pmatrix} (1-\alpha_1 L)^{-1} & 0 \\ 0 & (1-\alpha_2 L)^{-1} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

$$= \begin{pmatrix} (1-\alpha_1 L)^{-1}\cos\theta & -(1-\alpha_1 L)^{-1}\sin\theta \\ (1-\alpha_2 L)^{-1}\sin\theta & (1-\alpha_2 L)^{-1}\cos\theta \end{pmatrix}$$

so $\qquad \hat{D}_{21}(L) = \left[\hat{A}(L)^{-1}\hat{R}_c^{-1}Q\right]_{21} \approx (1-\alpha_2\mathrm{L})^{-1}\sin\theta$

Thus the step, keep Q if $\hat{D}_{21,i} \geq 0$, $i = 1,\ldots,4$ reduces to keep Q if $\sin\theta \geq 0$, which is equivalent to $0 \leq \theta \leq \pi$.

Thus, in large samples the posterior of $\hat{D}_{21}(L)$ is $\approx (1-\alpha_2\mathrm{L})^{-1}\sin\theta$, for $\theta \sim \mathrm{U}[0,\pi]$.

Characterization of posterior

A draw from the posterior (for a retained θ is): $D_{21}(\mathrm{L}) = (1-\alpha_2\mathrm{L})^{-1}\sin\theta$

Posterior mean for $D_{21,i}$:

$$
\begin{aligned}
E[D_{21,i}] &= E\left(\alpha_2^i \sin\theta\right) \\
&= \alpha_2^i E\left(\sin\theta\right) \\
&= \alpha_2^i \int_0^\pi \frac{1}{\pi}\sin\theta d\theta \\
&= \frac{\alpha_2^i}{\pi}(-\cos\theta\big|_0^\pi) = \frac{2}{\pi}\alpha_2^i \approx .637\alpha_2^i
\end{aligned}
$$

Posterior distribution: drop scaling by α_2^i and focus on $\sin\theta$ part

$$\Pr[\sin\theta \le x] = \Pr[\theta \le \mathrm{Sin}^{-1}(x)] \text{ for } \theta \sim U[0,\pi/2]$$

$$= 2\mathrm{Sin}^{-1}(x)/\pi$$

So the pdf of x is: $f_X(x) = \frac{d}{dx}\frac{2}{\pi}\mathrm{Sin}^{-1}(x) = \frac{2}{\pi\sqrt{1-x^2}}$

So the posterior of $\hat{D}_{21,i}$ is: $p(\hat{D}_{21,i}|Y) \propto \dfrac{2}{\pi\sqrt{1-x^2}}\alpha_2^i$

67% posterior probability interval with equal mass in each tail:

Lower cutoff:

$\Pr[\sin\theta \le x] = 1/6 \rightarrow x_{lower} = \sin(\pi/12) = .259$

$\Pr[\sin\theta \le x] = 5/6 \rightarrow x_{upper} = \sin(5\pi/12) = .966$

so 67% posterior coverage interval is [$.259\alpha_2^i$, $.966\alpha_2^i$], with mean $.637\alpha_2^i$

What's wrong with this picture?

- Posterior coverage interval: [$.259\alpha_2^i$, $.966\alpha_2^i$], with mean $.637\alpha_2^i$
- Identified set is [0, α_2^i]
- What is the frequentist confidence interval here?
- Why don't Bayesian and frequentist coincide?

Recent references on sign-restriction VARs:

Baumeister and Hamilton (WP, 2014)

Fry and Pagan (2011)

Kilian and Murphy (*JEEA*, 2012)

Moon and Schorfheide (*ECMA*, 2012)

Moon, Schorfheide, and Granziera (WP, 2013)