Bonn Summer School
Advances in Empirical Macroeconomics

Karel Mertens

Cornell, NBER, CEPR

Bonn, June 2015

Overview

1. Estimating the Effects of Shocks Without Much Theory
   1.1 Structural Time Series Models
   1.2 Identification Strategies
2. Applications to Fiscal Shocks
   2.1 Tax Policy Shocks
   2.2 Government Spending Shocks
   2.3 Austerity Measures
3. **Two Difficulties in Interpreting SVARs**
   3.1 **Noninvertibility**
   3.2 **Time Aggregation**
4. Systematic Tax Policy and the ZLB

3. Two Difficulties in Interpreting SVARs

Hansen and Sargent, 1991, "Two Difficulties in Interpreting Structural Vector Autoregressions", Rational Expectations Econometrics

3.1 Noninvertibility (nonfundamentalness)

3.2 Time aggregation

3.1 Noninvertibility

Invertibility of MA representation

$$z_t = M(L)\upsilon_t$$

requires that $\det(M(L)) \neq 0$ for $\mid L \mid \leq 1$.

If so, then υ_t is **fundamental** white noise for z_t, i.e. υ_t is contained in the linear space spanned by current and lagged z_t.

There are infinitely many other MA representations

$$z_t = \tilde{M}(L)\tilde{\upsilon}_t$$

in which $\tilde{\upsilon}_t$ is nonfundamental and is not contained in the linear space spanned by current and lagged z_t.

Economic agents make decisions based on current and lagged e_t living in a space $\mathcal{I}_t^a$, the agent's information set.

Econometricians make inference based on current and lagged z_t living in a space $\mathcal{I}_t^e$, the econometrician's information set.

e_t must be fundamental for z_t, i.e. the information sets must be the same.

If not, there is no hope for the econometrician to identify e_t without more assumptions.

Model Example I

Consider again the simple NK model

$$\begin{bmatrix} \hat{y}_t^{gap} \\ \pi_t \end{bmatrix} = E_t \sum_{j=0}^{\infty} \mathcal{C}^{-(j+1)} \begin{bmatrix} u_{t+j} \\ v_{t+j} \end{bmatrix}$$

where $\mathcal{C}^{-1} = \frac{1}{1+\phi_\pi \kappa} \begin{bmatrix} 1 & 1-\beta\phi_\pi \\ \kappa & \beta+\kappa \end{bmatrix}$

If shocks follow a VAR(1) process $\begin{bmatrix} u_t \\ v_t \end{bmatrix} = \Lambda \begin{bmatrix} u_{t-1} \\ v_{t-1} \end{bmatrix} + \Sigma e_t$, there is a VAR(1) representation for $\hat{y}_t^{gap}$ and π_t

$$\begin{bmatrix} \hat{y}_t^{gap} \\ \pi_t \end{bmatrix} = (\mathcal{C}-\Lambda)^{-1} \Lambda (\mathcal{C}-\Lambda) \begin{bmatrix} \hat{y}_{t-1}^{gap} \\ \pi_{t-1} \end{bmatrix} + (\mathcal{C}-\Lambda)^{-1} \Sigma e_t$$

The assumed shock process delivers invertibility in the past.

But now assume suppose v_t is white noise and

$$u_t = e^u_{t-1} \text{ , } e^u_t \text{ is white noise}$$

The solution is :

$$\begin{aligned} \begin{bmatrix} \hat{y}_t^{gap} \\ \pi_t \end{bmatrix} &= \left(\mathcal{C}^{-1} \begin{bmatrix} L & 0 \\ 0 & 1 \end{bmatrix} + \mathcal{C}^{-2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} e_t^u \\ v_t \end{bmatrix} \\ &= \mathcal{C}^{-1} \begin{bmatrix} L + \frac{1}{1+\phi_\pi \kappa} & \frac{\kappa}{1+\phi_\pi \kappa} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e_t^u \\ v_t \end{bmatrix} \end{aligned}$$

The matrix on the left looses rank at $L = -\frac{1}{1+\phi_\pi \kappa}$, which is inside the unit circle.

Hence there is no SVAR representation for $[\hat{y}_t^{gap}, \pi_t]$ for this shock process.

Model Example II

Consider RBC model:

$$\max \quad E_0\left[\sum_{t=0}^{\infty}\beta^t\left(\frac{C_t^{1-\sigma}-1}{1-\sigma}-\psi N_t\right)\right] \quad , \quad 0<\beta<1, \psi, \sigma>0$$

$$\text{s.t.} \quad C_t+K_{t+1}\leq A_t K_t^{\alpha} N_t^{1-\alpha} \quad , \quad \ln(A_t)=\varepsilon_t^u+e_{t-q}^a$$

where e_t^u and e_{t-q}^a are white noises with unit variance.

Optimality requires

$$\begin{aligned} C_t^{-\sigma} &= \beta E_t\left[C_{t+1}^{-\sigma}\alpha\frac{Y_{t+1}}{K_{t+1}}\right] \\ \psi C_t^{\sigma} &= (1-\alpha)\frac{Y_t}{N_t} \end{aligned}$$

where $Y_t = C_t + K_{t+1} = A_t K_t^{\alpha} N_t^{1-\alpha}$.

Loglinearizing and simplifying

$$
\begin{aligned}
-\sigma \hat{c}_t &= -\frac{\sigma}{\alpha} E_t \hat{c}_{t+1} + \frac{1}{\alpha} E_t[\hat{a}_{t+1}] \\
\hat{k}_t + \frac{1}{\alpha}\hat{a}_t &= \left(1 - \alpha\beta + \frac{1-\alpha}{\alpha}\sigma\right)\hat{c}_t + \alpha\beta\hat{k}_{t+1}
\end{aligned}
$$

Suppose $q = 1$, i.e. $\hat{a}_t = e^u_t + e^n_{t-1}$

The solution is

$$
\begin{bmatrix} \hat{c}_t \\ \hat{k}_{t+1} \end{bmatrix} = \begin{bmatrix} \alpha\xi \\ \alpha \end{bmatrix} \hat{k}_t + \begin{bmatrix} \xi & \theta\left[\xi - \frac{1}{\sigma}\right] \\ 1 & -\frac{1-\alpha\theta}{\alpha\xi}\left[\xi - \frac{1}{\sigma}\right] \end{bmatrix} \begin{bmatrix} \hat{a}_t \\ e^n_t \end{bmatrix}
$$

where $\theta = \alpha\beta$ and $\xi = \frac{1-\alpha\theta}{\alpha(1-\theta+\sigma\frac{1-\alpha}{\alpha})}$.

For $q = 1$, the MA representation is

$$\begin{bmatrix} \hat{c}_t \\ \hat{k}_{t+1} \end{bmatrix} = (1 - \alpha L)^{-1} \begin{bmatrix} \xi & \frac{1}{\sigma} L + \theta \left(\xi - \frac{1}{\sigma} \right) \\ 1 & L - \frac{1-\alpha\theta}{\alpha\xi} \left(\xi - \frac{1}{\sigma} \right) \end{bmatrix} \begin{bmatrix} e_t^s \\ e_t^n \end{bmatrix}$$

where the determinant of the MA term is a constant and therefore has no roots inside the unit circle (provided $\sigma \neq 1$).

Hence there exists a SVAR representation for c_t and k_{t+1}, which can be obtained by inverting the MA polynomial matrix.

Now suppose $q = 2$, i.e. $\hat{a}_t = e^u_t + e^n_{t-2}$

The solution is

$$\left[\begin{array}{c} c_t \\ k_{t+1} \end{array}\right] = \left[\begin{array}{c} \alpha\xi \\ \alpha \end{array}\right] k_t + \left[\begin{array}{ccc} \xi & \theta\left[\xi - \frac{1}{\sigma}\right] & \theta^2\left[\xi - \frac{1}{\sigma}\right] \\ 1 & -\frac{1-\theta\alpha}{\alpha\xi}\left[\xi - \frac{1}{\sigma}\right] & -\theta\frac{1-\theta\alpha}{\alpha\xi}\left[\xi - \frac{1}{\sigma}\right] \end{array}\right] \left[\begin{array}{c} a_t \\ e^n_{t-1} \\ e^n_t \end{array}\right]$$

The MA representation is

$$\left[\begin{array}{c} c_t \\ k_{t+1} \end{array}\right] = (1-\alpha L)^{-1} \left[\begin{array}{cc} \xi & \frac{1}{\sigma}L^2 + \theta^2\left(\xi - \frac{1}{\sigma}\right) \\ 1 & L^2 - \left[\frac{1-\alpha\theta}{\alpha\xi}\left(\xi - \frac{1}{\sigma}\right)\right] L - \theta\frac{1-\alpha\theta}{\alpha\xi}\left(\xi - \frac{1}{\sigma}\right) \end{array}\right] \left[\begin{array}{c} e^s_t \\ e^n_t \end{array}\right]$$

where the determinant of the MA term is proportional to $-(L+\theta)$ and thus has a root $0 > -\theta > -1$.

Unambiguously inside the unit circle, so no SVAR representation for c_t and k_{t+1}.

The constant θ is the *anticipation rate*: the rate at which news about the future is discounted by rational forward looking agents. (Ljungqvist and Sargent, 2004 chapter 11).

There are many other theoretical examples

See e.g. Sargent and Hansen (1991), Fernandez-Villaverde et al. (2007), Mertens and Ravn (2010), Leeper Walker and Yang (2013),...

The problem typically arises when the shock process follows noninvertible MA, as is the case in 'news shock' literature.

The problem is not the VAR methodology, but the insufficiency of z_t.

Residuals in VAR will be white noise, but are generally still linear combinations of all current and lagged structural shocks.

Addressing the Problem

1. Find variables that contain relevant information;

   e.g. 'commodity price index' and the price puzzle (Sims 1992)

   Ramey defense news variable (Ramey 2011), Tax news narrative (Mertens and Ravn 2011, 2012), Municipal bond spread (Leeper Walker and Yang, 2013)

2. Flip roots with Blaschke matrices

   Lippi and Reichlin 1994, Mertens and Ravn (2010)

3. Large datasets and dimensionality reduction;

   State-Space Models, Factor Augmented VARs

4. Impose structure of a DSGE model.

Application: Anticipated Tax Changes

Based on revision Mertens and Ravn, 2012, Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated U.S. Tax Policy Shocks, American Economic Journal: Economic Policy

Matlab codes and data available on my webpage (look for the 2011 RED companion paper)

Tax policy interventions are often associated with **implementation lags**

- Some preference for phasing-in of changes in tax rates
- Implementation lags are common and can be quite long
- Therefore, tax policy shocks may often be to a large extent anticipated

Do implementation lags matter?

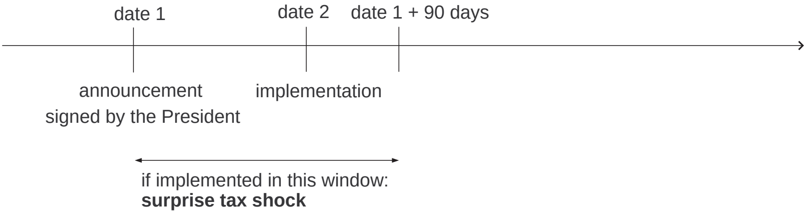
Do agents respond to news about future tax rates?

Traditional SVAR analysis can be problematic because of nonfundamentalness.

The Measurement of Tax Shocks

Romer and Romer (2010) narrative tax changes.

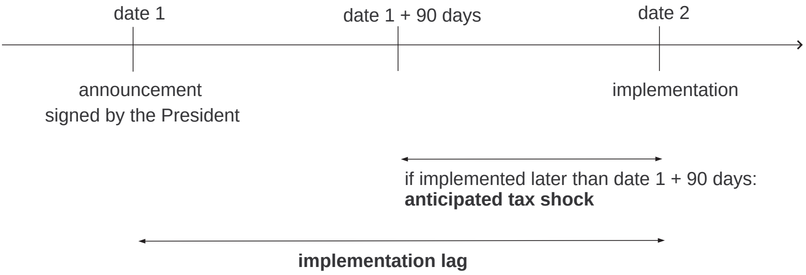
Distinction between **anticipated** and **surprise** tax shocks



The Measurement of Tax Shocks

Romer and Romer (2010) narrative tax changes.

Distinction between **anticipated** and **surprise** tax shocks



Example: The Reagan Tax Cut

August 13, 1981: U.S. Congress passes the *Economic Recovery Tax Act of 1981*, signed by President Reagan

Reduction in marginal tax rates, reduction in corporate taxes and new depreciation guidelines.

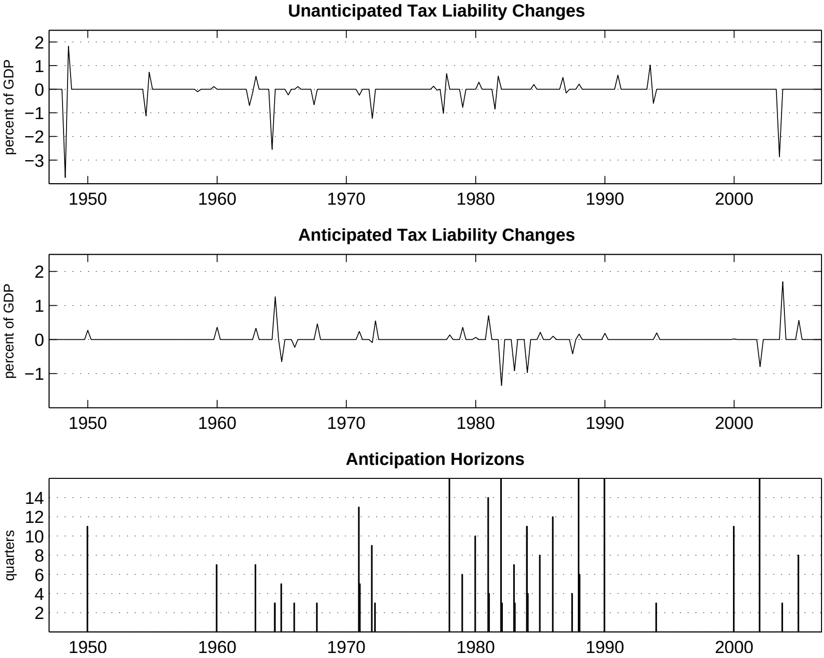
Phasing-in of the tax changes over time:

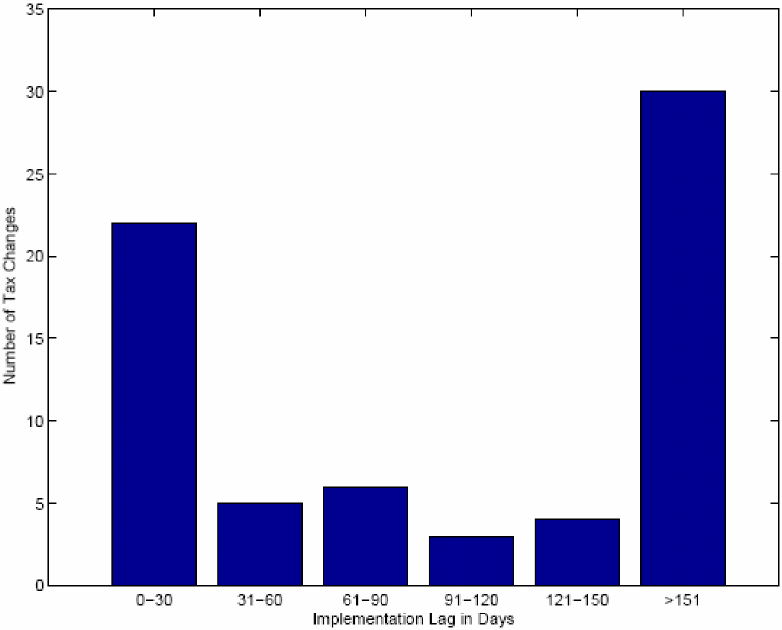
1981Q3	\$26.7 billion tax liability cut	surprise
1981Q4	\$17.8 billion tax liability increase	
1982Q1	\$48.8 billion tax liability cut	anticipated
1983Q1	\$57.3 billion tax liability cut	
1984Q1	\$36.1 billion tax liability cut	

With this classification we find:

- A total of **70 exogenous tax liability changes**
- **33** are classified as **unanticipated**
- **37** are classified as **anticipated**
- the **median implementation lag** is 6 quarters
- the **minimum implementation lag** is 2 quarters
- the **maximum implementation lag** is 21 quarters

Kennedy, Reagan and Bush tax acts associated with substantial anticipated tax changes


Unanticipated Tax Liability Changes
percent of GDP
2
1
0
−1
−2
−3
1950
1960
1970
1980
1990
2000
Anticipated Tax Liability Changes
percent of GDP
2
1
0
−1
1950
1960
1970
1980
1990
2000
Anticipation Horizons
quarters
14
12
10
8
6
4
2
1950
1960
1970
1980
1990
2000


35
30
25
20
15
10
5
0
Number of Tax Changes
0−30
31−60
61−90
91−120
121−150
>151
Implementation Lag in Days

VAR-X Specification

$$z_t = d_t + C(L)z_{t-1} + D(L)\tau_t^u + F(L)\tau_{t,0}^a + \sum_{i=1}^{K} G_i \tau_{t,i}^a + u_t$$

τ_t^u : Unanticipated tax shocks implemented at date t

$\tau_{t,i}^a$: Anticipated tax shocks "known" at date t and implemented at date $t+i$

K : Maximum anticipation horizon that we allow for

Inspired by VARMA representation.

We measure the anticipated shocks as:

$$\tau_{t,i}^{a} = \sum_{j=0}^{M-i} s_{t-j}^{a,i+j}$$

$s_{t-j}^{a,i+j}$: Tax liability changes signed at date $t-j$ with an implementation lag of $i+j$ quarters

M : Maximum implementation lag in the data

- Therefore, we measure the anticipated shocks on the basis of their remaining implementation lag.
- Ideally, one would like to distinguish between tax shocks on the pure basis of their anticipation horizon but this would require many more observations

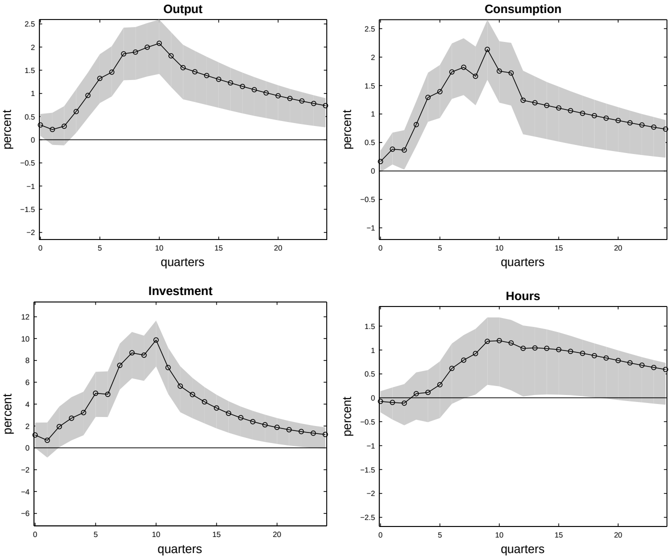
Specification

Oberservables z_t, quart. sample 1947:1 - 2006:4

- Real GDP per capita
- Real consumption p.c.
- Real investment per capita
- Hours worked per capita
- Real wages

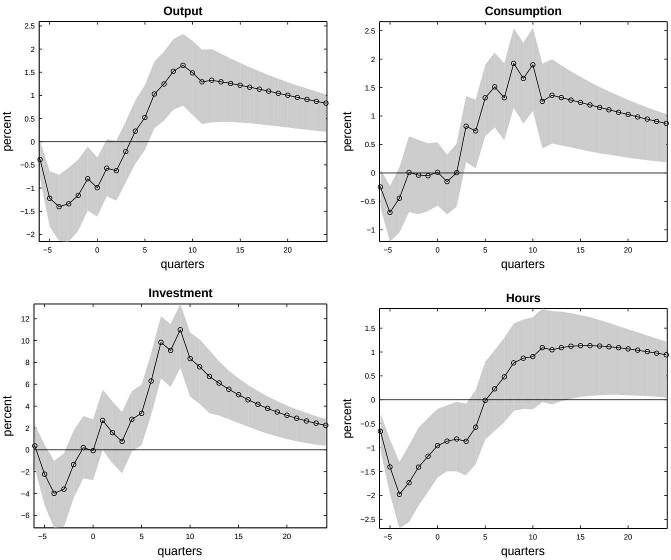
K=6 (6 quarter maximum anticipation)
$C(L)$ includes one lag
$D(L)$ and $F(L)$ include 12 lags of implemented tax changes

A 1% Unanticipated Tax Cut Gives Rise To



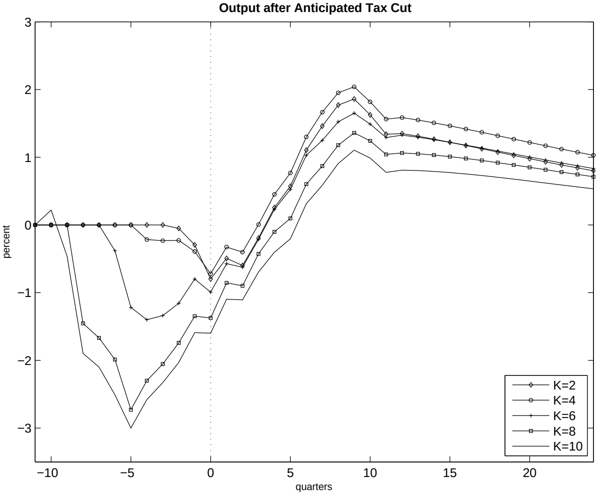
68% bootstrapped confidence intervals.

A 1% Anticipated Tax Cut Gives Rise To

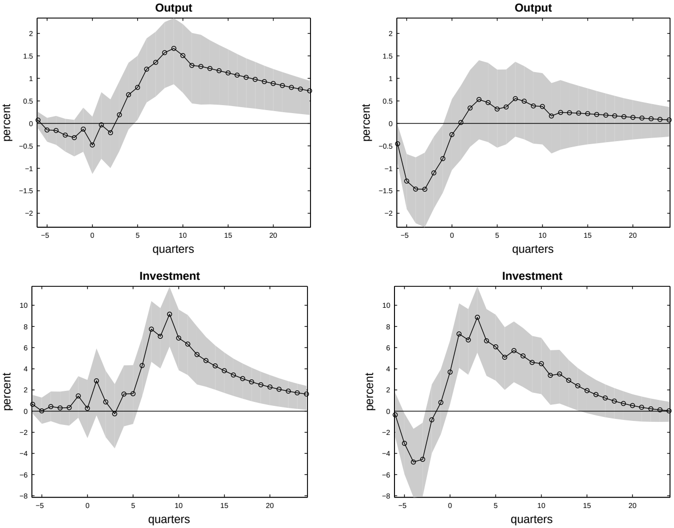


68% bootstrapped confidence intervals.

Sensitivity to Anticipation Horizon K



Anticipation Effects of Surprise Tax Changes



68% bootstrapped confidence intervals.

Implications for the US Business Cycle

Tax liability shocks bring about important adjustment dynamics of the economy.

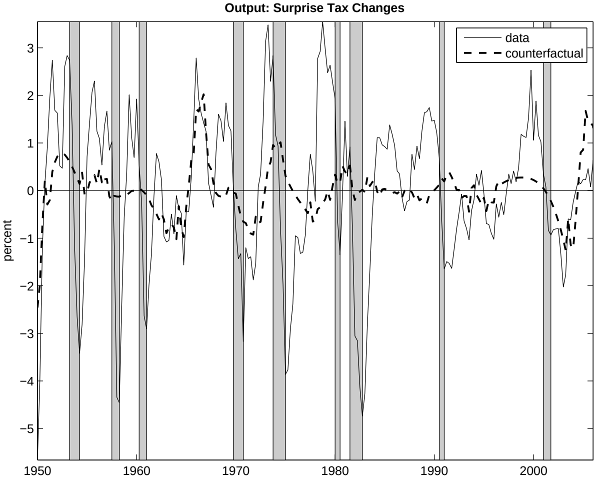
But have these shocks been important for US business cycles?

Counterfactual: simulate the vector of endogenous variables allowing only for tax shocks

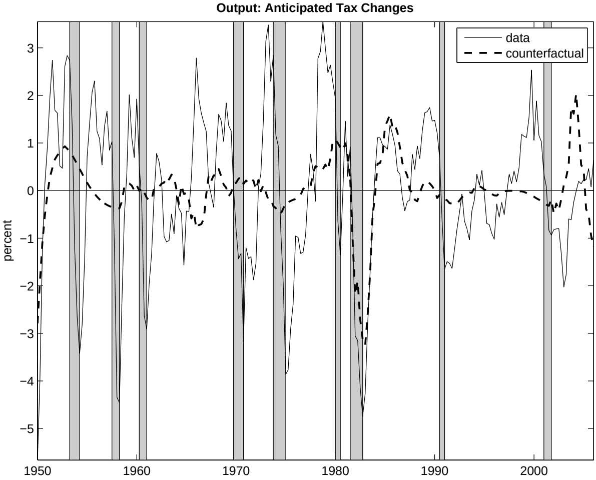
Larger VAR system with monetary variables (see paper)

Resulting time series are Hodrick-Prescott filtered.

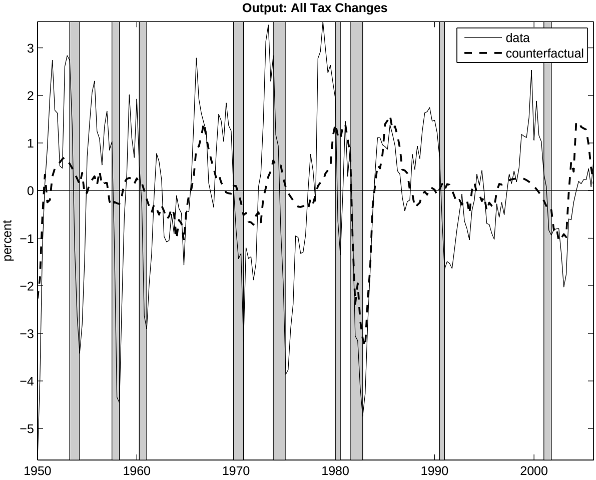
Only Surprise Changes



Only Anticipated Changes



All Tax Changes



Alternative Approach: Municipal Bond Spreads

Leeper, Walker and Yang (2013)

In the US, municipal bonds are exempt from federal taxes.

$\mathcal{Y}_t^m$: yield on a municipal bond at t
$\mathcal{Y}_t$ yield on a taxable bond at t

Define an implicit tax rate

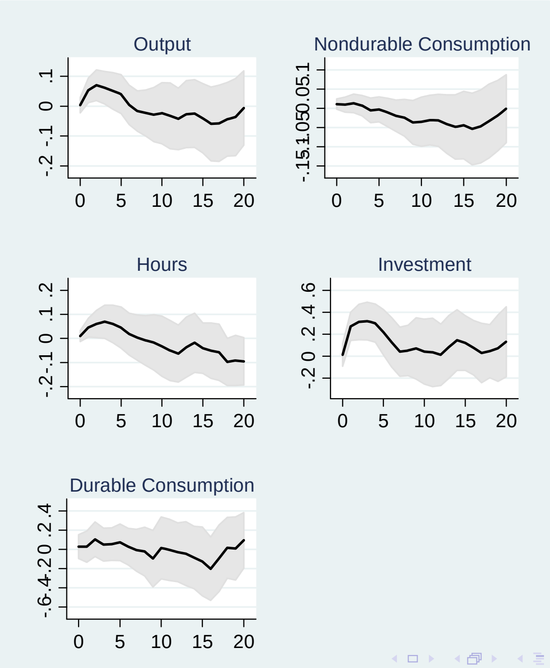
$$\tau_t^I = 1 - \mathcal{Y}_t^m / \mathcal{Y}_t$$

Assuming bonds are otherwise identical, arbitrage implies τ_t^I is a weighted average of discounted expected future tax rates from t to maturity.

Add τ_t^I to a VAR and order first to a Choleski decomposition (but A2 is problematic).

Ramey (2015) uses LP-IV approach.

Figure 4.4 Effect of News of Future Tax Increase, Leeper, Richter, Walker (2011) Measure



State Space/Factor Models

Given available samples, VARs contain limited # of variables in z_t.

Consider again z_t is $n \times 1$ and e_t is $l \times 1$, but now n large and $n \gg l$.

Also assume

$$z_t = z_t^* + \xi_t$$

where ξ_t is uncorrelated white noise measurement error

True data is from a linear model such that

$$\begin{aligned} s_t &= \mathcal{G} s_{t-1} + \mathcal{F} e_t \\ z_t &= \mathcal{A} s_{t-1} + \mathcal{D} e_t + \xi_t \end{aligned}$$

with s_t is $m \times 1$ state vector.

State space/factor model estimation: keep m relatively low.

See Stock and Watson (2011, 2015) for surveys.

Factor Augmented VARs

Bernanke, Boivin, and Eliasz (2005) consider a reduced form FAVAR(1):

$$\begin{pmatrix} f_t \\ z_t \end{pmatrix} = C(L) \begin{pmatrix} f_{t-1} \\ z_{t-1} \end{pmatrix} + \upsilon_t$$
$$x_t = \Lambda^f f_t + \Lambda^y z_t + \epsilon_t$$

- f_t are $m \times 1$ unobservable factors
- z_t are $n \times 1$ observable core variables of interest
- x_t are $k \times 1$ additional informational variables (stationary)
- ϵ_t are error terms (asymptotically) uncorrelated
- Λ^f is $k \times m$ and Λ^z is $k \times n$

Note: n is 'small' but now $k > n + m$ is large

Again a state space model.

Principal Components Estimation

A simple estimation procedure:

1. Estimate p_t, the first $n + m$ principal components of $E[x_t x_t']$

   Normalize $E[p_t p_t'] = I$.

2. Estimate VAR system replacing f_t by p_t.

Note: Identification restrictions may require additional steps in either stage (e.g. Bernanke, Boivin, and Eliasz (2005)).

Factor models with external instruments seem particularly appealing (see Stock and Watson, 2012)

Forni and Gambetti (2014) propose some simple testing procedures

Test for Informational Sufficiency:

- Estimate simple VAR for z_t and test whether p_t Granger causes z_t. If rejected, than z_t is informationally sufficient.
- If not rejected, add factors to the VAR one at a time in decreasing order until Granger-causality is rejected.

Even if informational sufficiency is rejected, identification of single shock may still be OK.

Test for 'Structuralness' of an Estimated Shock:

- Test for orthogonality of identified shock to lags of p_t.

3.2 The Time Aggregation Problem

Suppose the model for high frequency data is a VAR(p)

$$G(Z)z^*_\tau \quad = \quad \mathcal{D}e_\tau \ , \ \tau = 1, ..., kT$$

where $G(Z) = I_n - \mathcal{G}_1 Z - ... - \mathcal{G}_p Z^p$, Z is the lag operator $Z^j x_\tau = x_{\tau-j}$.

Suppose the econometrician observes average sampled data

$$z_t = (I + Z + ...Z^{k-1})z^*_{tk} \ , \quad t = 1, .., T$$

where t indexes the lower frequency. For concreteness, assume that $p \geq k - 1$.

Can we fit a VAR for z_t and do structural VAR analysis?

Generally no.

The time aggregated data has a VARMA(p,q) representation

$$B(L)z_t \quad = \quad H(L)\upsilon_t \ , \ t = 1, ..., T$$

where

$$E[\upsilon_t] = 0 \ , \ E[\upsilon_t \upsilon_t'] = \Sigma \ , \ E[\upsilon_t \upsilon_s'] = 0 \text{ for } s \neq t$$

Based on results in Marcellino (1999):

- $B(L)$ has order p or less, i.e. generally the same order as $G(L)$
- The order of $H(L)$ is bounded by p if $p = k - 1$, or otherwise by $p + 1 + q$ where q is the smallest positive integer that satisfies $qk \leq k - p - 2 < (q + 1)k$
- $\upsilon_t \neq \mathcal{D}(I + Z + ...Z^{k-1})e_t$. Instead υ_t is a linear combination of current and up to $(p + 1)(k - 1)$ lags of shocks.

Similar results hold for point in time sampling.

We cannot expect to uncover high frequency dynamics with low frequency data.

Again, the problem is not the VAR, but insufficient data.

In practice, implications for the interpretation of VAR residuals are potentially serious.

Recent developments:

- Mixed-Frequency VARs and Nowcasting models

  (e.g. Mariano and Murasawa (2009), Banbura, Giannone, Modugno and Reichlin (2012), Foroni, Ghysels and Marcellino (2013), Schorfheide and Song (2014))

- MIDAS-VARs: Ghysels (2012)

Few applications of structural mixed frequency models, see Ghysels (2012) and Marcellino (2014) for exceptions.