Estimating Causal Effects in Macroeconomics:

General Methods and Pitfalls

By Valerie A. Ramey

Question we want to answer:

How can we estimate empirically the key parameters that help us answer the following type of question:

"What is the causal effect of

* changes in government spending, taxes, fiscal consolidations
* changes in monetary policy in the domestic country or the U.S., Europe
* changes in technology, preferences, etc.

on

* GDP, consumption, investment, exchange rates, prices, interest rates,
* other variables of interest?"

Outline

I. Estimating the Causal Effect on Macroeconomic Variables

A. Identification: Conceptual Issues

B. Approaches to Identification in Times Series Models

II. Estimating Impulse Responses

III. Pitfalls

A. Anticipations
B. Trends, lags, outliers and influential observations
C. Counterfactuals
D. Nonlinearities

I. Estimating the causal effect on macroeconomic variables

A. Identification: Conceptual Issues

Although macroeconomists don't talk about identification much, it is at the heart of every empirical effort.

Identification turns correlations into causal relationships

Simple Static Example

How much will GDP increase if the government raises government spending or raises taxes?

Consider a simple model in a static setting:

$$\tau_t = b_{\tau g} g_t + b_{\tau y} y_t + \varepsilon_{\tau t}$$
$$g_t = b_{g\tau} \tau_t + b_{gy} y_t + \varepsilon_{gt}$$
$$y_t = b_{y\tau} \tau_t + b_{yg} g_t + \varepsilon_{yt}$$

τ = net taxes
g = government spending
y = GDP

ε's are macroeconomic and policy shocks. Assume they are uncorrelated with each other.

$\varepsilon_{\tau t}$ = tax shock
ε_{gt} = government spending shock
ε_{yt} = macroeconomic shock, such as technology, confidence, etc.

Estimates of b_{yg} and $b_{y\tau}$ help us answer the question we posed earlier.

Problems in estimating b_{yg} and $b_{y\tau}$

$$\tau_t = b_{\tau g} g_t + b_{\tau y} y_t + \varepsilon_{\tau t}$$
$$g_t = b_{g\tau} \tau_t + b_{gy} y_t + \varepsilon_{gt}$$
$$y_t = b_{y\tau} \tau_t + b_{yg} g_t + \varepsilon_{yt}$$

τ = net taxes

g = government spending

y = GDP

- We expect $b_{yg} > 0$ and $b_{y\tau} < 0$, but government spending and taxes may also respond to GDP, e.g., $b_{gy} < 0$ and $b_{\tau y} > 0$.

- Thus, a simple OLS regression of GDP on government spending and taxes will not uncover b_{yg} and $b_{y\tau}$ because g_t and τ_t are correlated with the shock to GDP, ε_{yt}.

- For example, we might observe no correlation between GDP and government spending, but this correlation is consistent with both no structural relationship between GDP and government spending (i.e. $b_{yg} = b_{gy} = 0$) and with b_{yg} and b_{gy} being large, with equal but opposite signs.

Problems in estimating b_{yg} and $b_{y\tau}$ *(cont.)*

$$\tau_t = b_{\tau g} g_t + b_{\tau y} y_t + \varepsilon_{\tau t}$$
$$g_t = b_{g\tau} \tau_t + b_{gy} y_t + \varepsilon_{gt}$$
$$y_t = b_{y\tau} \tau_t + b_{yg} g_t + \varepsilon_{yt}$$

τ = net taxes

g = government spending

y = GDP

- Without further assumptions or data, we cannot identify the parameters – neither the order nor the rank condition is met.
- But what if we can identify some of the shocks, ε's ?
- Making assumptions to identify the ε's has been the main way that macroeconomists deal with identification.

What if macroeconomists approached identification like modern applied microeconomists?

We would:

* Try to find good natural experiments.
* Run randomized control trials.

Suppose we wanted to identify the causal effects of an increase in government spending on GDP?

Natural Experiment Methods

* Find examples when other factors *which were uncorrelated with the state of the economy* led to changes in government spending.

* Use Diff-in-Diff methods to estimate the causal effects of government spending.

* Note – implicitly Diff-in-Diff is comparing treatment and control.

Randomized Controlled Trial

The IMF would run a randomized controlled trial:

- It would randomly assign a large group of countries to treatment and control groups.
- In the treatment group, government spending would be increased for 2 years starting immediately and financed by deficit spending. The government would commit to a future increase in tax rates to finance the deficit.
- There would be no change in the countries in the control group.
- After 2 years, the IMF economists would simply test whether the growth rate of output over the two years was different across the treatment and control groups.

In a sense, macroeconomists' methods are trying to reproduce the structure of a randomized controlled trial

Treatment Group - quarterly or country-year observations in which government spending changes.

Control Group – all other quarterly or country-year observations; note that we often call this the "counterfactual"

Random Assignment - to mimic random assignment, we need to:

- Identify unanticipated, "exogenous" shocks

- Include control variables so that we compare to the right counterfactual or control group.

Why identification is so difficult in macro

* Angrist and Pischke (2010) – econometric revolution in applied micro, claim that macroeconomics is lagging behind on econometric revolution.

* In macroeconomists' defense, identification is particularly challenging in macroeconomics because researchers ask questions for which:

- dynamics are all-important
- general equilibrium effects are crucial
- expectations have powerful effects.

Back to our simple macro model

Static fiscal Model

$$\tau_t = b_{\tau g} g_t + b_{\tau y} y_t + \varepsilon_{\tau t}$$
$$g_t = b_{g\tau} \tau_t + b_{gy} y_t + \varepsilon_{gt}$$
$$y_t = b_{y\tau} \tau_t + b_{yg} g_t + \varepsilon_{yt}$$

τ = net taxes

g = government spending

y = GDP

Now let's extend it to a general trivariate model with dynamics.

Simple Trivariate Dynamic Model

- $Y_t = [Y_{1t}, Y_{2t}, Y_{3t}]$ be the vector of endogenous variables.

- Dynamic behavior of Y_t described by the following structural model:

$$Y_t = B(L)Y_t + \varepsilon_t$$

where $B(L) = B_0 + \sum_{k=1}^{p} B_k L^k$ and $E[\varepsilon_t \varepsilon_s'] =$ D if $t = s$, and 0 otherwise, where D is a diagonal matrix. The ε's are the primitive structural shocks.

- The elements of B_0 are the same as the b's from the simple static model with $b_{jj} = 0$.

Simple Trivariate Dynamic Model (cont.)

Thus, the easiest way to address the dynamics is to recast the problem in terms of the *innovations* from a reduced form vector autoregression (VAR):

$$A(L)Y_t = \eta_t$$

where A(L) is a polynomial in the lag operator and $A(L) = I - \sum_{k=1}^{p} A_k L^k$.

$\eta_t = [\eta_{1t}, \eta_{2t}, \eta_{3t}]$ are the reduced form VAR innovations.

Simple Trivariate Dynamic Model (cont.)

We can then link the innovations η_t back to the underlying structural shocks ε :

$$\eta_{1t} = b_{12}\eta_{2t} + b_{13}\eta_{3t} + \varepsilon_{1t}$$

$$\eta_{2t} = b_{21}\eta_{1t} + b_{23}\eta_{3t} + \varepsilon_{2t}$$

$$\eta_{3t} = b_{31}\eta_{1t} + b_{32}\eta_{2t} + \varepsilon_{3t}$$

(I have imposed the restrictions that (i) each shock enters only one equation; and (ii) each shock has unit impact.)

The interpretations of the *b*'s are the same as in the static model if the structural relationships depend on contemporaneous interactions.

But we still face the same identification problem!

$$\eta_{1t} = b_{12}\eta_{2t} + b_{13}\eta_{3t} + \varepsilon_{1t}$$

$$\eta_{2t} = b_{21}\eta_{1t} + b_{23}\eta_{3t} + \varepsilon_{2t}$$

$$\eta_{3t} = b_{31}\eta_{1t} + b_{32}\eta_{2t} + \varepsilon_{3t}$$

η_{1t} = innovation to taxes
η_{2t} = innovation to government spending
η_{3t} = innovation to GDP

- We still have the simultaneous equation problem.
- We have no excluded instruments.
- However, if we can identify the shocks, we can identify the parameters.

B. Approaches to Identification in Time Series Models

A reminder of what we want in an instrument:

Simple equation: $y_t = b \cdot x_t + \varepsilon_t$

Suppose x_t is correlated with ε_t.

A series Z is a valid instrument for identifying parameter b if the following two conditions hold:

Relevance: $E[Z_t x_t] \neq 0$

Exogeneity: $E[Z_t \varepsilon_t] = 0$

The importance of instrument relevance

* Work by Bound, Jaeger, Baker (1995) and Staiger-Stock (1997) show how far wrong you can go if your instruments have low relevance.

* Even with gigantic data sets, (Angrist-Krueger had 330,000 observations), the IV will be severely biased towards OLS if the first-stage F-statistic is low.

* Typically, you need first-stage F-statistics to be 18 or above!

1. Cholesky Decompositions

- Identification based on timing.
- Imposes recursive zero restrictions.
- Examples:

i. Blanchard-Perotti: Government spending does not respond to contemporaneous shocks to output or taxes. Thus, $b_{g\tau} = b_{gy} = 0$ in

$$\eta_{gt} = b_{g\tau}\eta_{\tau t} + b_{gy}\eta_{yt} + \varepsilon_{gt}$$

In practice, this means that we regress g on lags of g, τ, and y to obtain the reduced form innovation η_{gt} and we assume $\eta_{gt} = \varepsilon_{gt}$, the structural shock to g.

This identified ε_{gt} is like an *instrument* for η_{gt} in the following equation:

$$\eta_{yt} = b_{y\tau}\eta_{\tau t} + b_{yg}\eta_{gt} + \varepsilon_{yt}$$

1. Cholesky Decompositions (cont.)

ii. Bernanke-Blinder: They assume that the other endogenous variables do not respond to the policy variable.

In their case, they were looking at the effects of monetary policy.

Consider a trivariate system with output, prices, and the federal funds rate (the policy variable).

To identify the shock to monetary policy, they regress the federal funds rate on the lags of output, prices, and the federal funds rate *and on the contemporaneous values of output and prices*. They call the residual from this equation the monetary policy shock.

This is the recursiveness assumption and we will discuss how it is not innocuous.

Stata Demonstration: Cholesky Decomposition Intuition

Trivariate fiscal model

1. Regress y, t, g on lagged values of all variables and extract the reduced form residuals, η

2. Use identification assumption that government spending does not respond to contemporaneous shocks to y and t: $\varepsilon_{gt} = \eta_{gt}$

3. Compare the impact coefficient in the following 4 regressions:

i. Regress η_{gt} on η_{gt}
ii. Regress y on η_{gt}, plus lags(g, t, y).
iii. Regress y on g_t, plus lags(g, t, y).
iv. Run an IV regression of y on g_t, plus lags(g, t, y), using ε_{gt} as an instrument for g_t.

Suppose we identified all three shocks using a Cholesky decomposition

- Decide ordering: g, τ, y

  - g does not respond to contemporaneous τ or y.
  - τ can respond to contemporaneous g, but not contemporaneous y. (silly)
  - y can respond to contemporaneous g and τ.

- In practice, you are running the following regressions:

  Reg g_t on lags(g, τ,y).
  Reg τ_t on g_t, lags(g, τ,y).
  Reg y_t on g_t, τ_t, lags(g, τ,y).

- Then you use the coefficients to dynamically simulate.

2. Structural VARs (SVARs)

- Uses either economic theory or outside estimates to constrain parameters.

- Example: Blanchard-Perotti: Use outside evidence on cyclical sensitivity of taxes and set $b_{\tau y}$=2.08.

$$\eta_{\tau t} = b_{\tau g}\eta_{gt} + b_{\tau y}\eta_{yt} + \varepsilon_{\tau t}$$

$$\eta_{gt} = b_{g\tau}\eta_{\tau t} + b_{gy}\eta_{yt} + \varepsilon_{gt}$$

$$\eta_{yt} = b_{y\tau}\eta_{\tau t} + b_{yg}\eta_{gt} + \varepsilon_{yt}$$

Recall that they also set $b_{g\tau}$= b_{gy} = 0

Stata Demonstration: SVAR intuition

1. Regress y, t, g on lagged values of all variables and extract the reduced form residuals, η

2. Construct the variable $\eta_{\tau t} - 2.08 \cdot \eta_{yt}$.

3. Either:

a. Regress η_{yt} on η_{gt} and $\eta_{\tau t}$ and use $\eta_{\tau t} - 2.08 \cdot \eta_{yt}$ as the instrument for $\eta_{\tau t}$.

b. Regress y on g_t and τ_t plus lags(g, τ,y), and use $\eta_{\tau t} - 2.08 \cdot \eta_{yt}$ as the instrument for τ_t.

Either a or b provides estimates of $b_{y\tau}$ *and* b_{yg}.

3. Narrative Methods

- Use historical documents to identify policy shocks.

  - Find changes in policy that are uncorrelated with the current state of the economy.
  - The idea is that these shocks are uncorrelated with the other shocks.

- Examples

  Monetary: Friedman and Schwartz (1963), Romer-Romer (1989, 2004)

  Fiscal: Poterba (1986), Ramey-Shapiro (1998), Ramey (2011), Romer-Romer (2010), Leigh et al. (2010)

- Caution: Narratives alone do not provide exogeneity. Consider the following:

Example in which Narrative Methods Don't Necessarily Solve the Exogeneity Problem

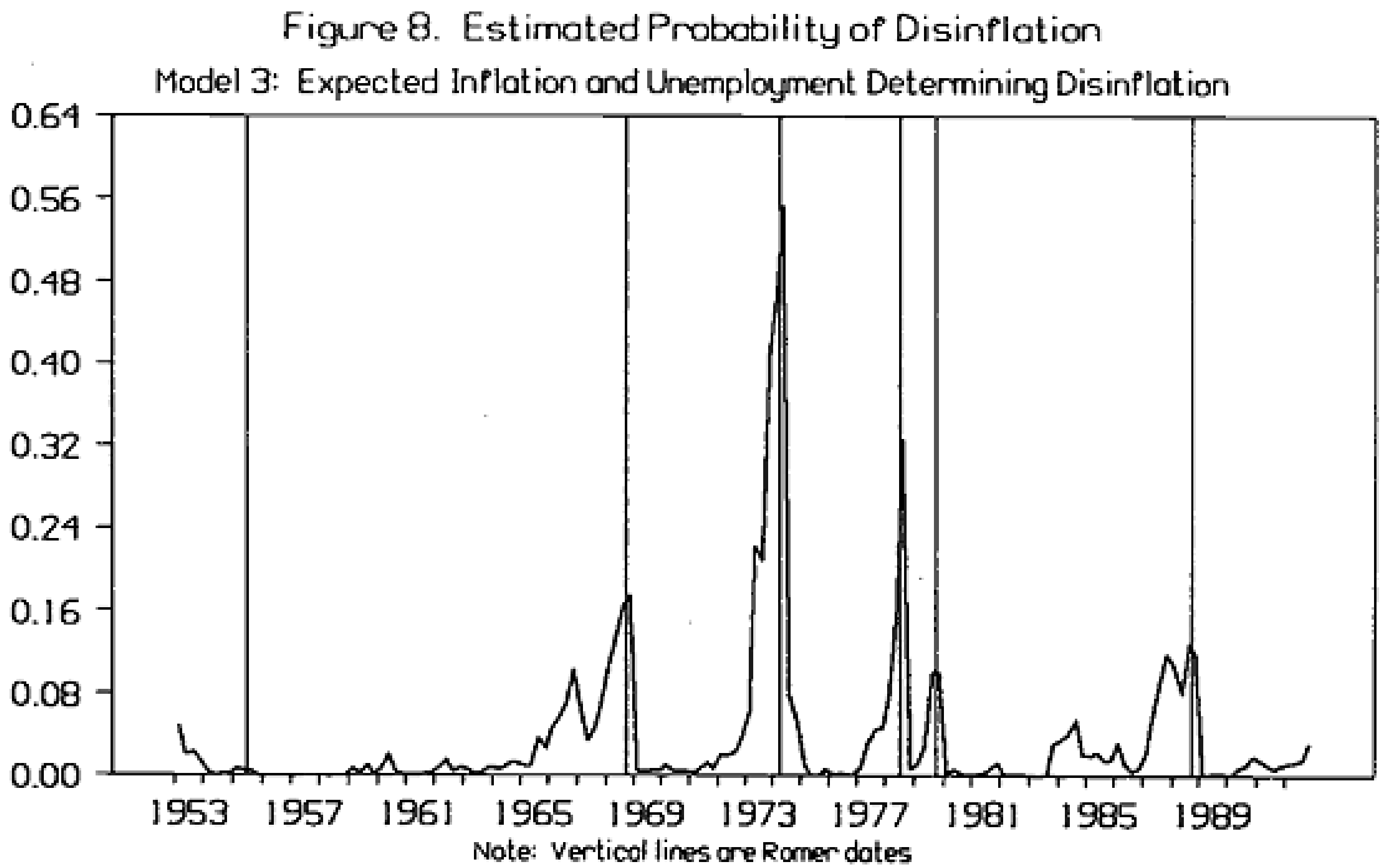
- Romer and Romer (1989) used the narrative approach to identify dates at which Fed decided to reduce inflation.
- They took this as an exogenous shock to policy and then studied the effects.
- We now know that they were estimating the reaction part of policy, not an exogenous shock.

$$i_t = .04 + 1.5(\pi_t - .02) + 0.5(y_t - ybar_t)$$

- In fact, Shapiro (1994) showed that the dates were predictable from expectations about future unemployment and inflation:

From Shapiro (1994)

"Federal Reserve Policy: Cause and Effect"



Thus, these dates can't be used to answer the question: What is the independent effect of the Federal Reserve raising interest rates?

4. High Frequency Identification

- Uses high frequency data from financial instruments during key windows.

  e.g. news announcements around FOMC dates and the movement of federal funds futures to identify unexpected Fed policy actions.

- This identification is also based in part on timing, but because the timing is so high frequency (daily or higher), the assumptions are more plausible than those employed at the monthly or quarterly frequency.

- However, without additional assumptions the unanticipated shock is not necessarily exogenous to the economy.

  e.g. Fed's private information about the state of the economy might be driving its policy changes

5. External Instrument/Proxy SVAR

- Developed independently by Stock-Watson and Mertens-Ravn.

- Uses external instruments for identification in an SVAR.

- Idea:

Suppose that Z_t represents one of these external series. Then this series is a valid instrument for identifying the shock ε_{1t} if the following two conditions hold:

Relevance: $E[Z_t \varepsilon_{1t}] \neq 0,$

Exogeneity: $E[Z_t \varepsilon_{it}] = 0$ i = 2, 3

5. External Instrument/Proxy SVAR (cont.)

To identify ε_{1t}

- Step 1: Estimate the reduced form system to obtain estimates of the reduced form residuals, η_t.

- Step 2: Regress η_{2t} and η_{3t} on η_{1t} using the external instrument Z_t as the instrument. These regressions yield unbiased estimates of $b_{21}$and b_{31}. Define the residuals of these regressions to be ν_{2t} and ν_{3t}.

- Step 3: Regress η_{1t} on η_{2t} and η_{3t} , using the ν_{2t} and ν_{3t} estimated in Step 2 as the instruments. This yields unbiased estimates of b_{12} and b_{13}.

4 other leading methods

6. Long-run restrictions

- Impose theoretical restrictions about the long run, such as no long-run effects of demand shocks.
- Can be recast as an IV problem.
- Instrument relevance can be a problem.

7. Sign restrictions

- Also known as set identification.
- Several recent innovations, such as Bayesian methods, influential observations.

8. FAVARs

- Tries to address the problem of including enough conditioning variables without saturating the regressions.

9. Estimated DSGE

- Uses theory and assumptions on the structure of the shocks to identify parameters and shocks.

II. Impulse Response Functions

Suppose we have identified an exogenous shock, ε_{1t} .

There are 3 common methods for estimating dynamic effects of shocks:

(A) Standard VAR

(B) Dynamic simulation (used by Romer-Romer (1989, 2010), Ramey-Shapiro (1998))

(C) Jorda local projections

A. Standard VAR: what is dynamic effect of ε_{1t}?

$$Y_t = B(L)Y_t + \varepsilon_t$$

with $Y_t = [Y_{1t}, Y_{2t}, Y_{3t}]$, $\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}]$

2 equivalent methods:

a. Reduced form parameters provide elements of moving average representation. The IRFs are nonlinear functions of the reduced form parameters.

b. Set $e_{1t} = 1$ (or its standard deviation) and calculate the responses recursively from the estimated B's, allowing all three variables to respond dynamically.

B. Dynamic simulations

$$Y_{1,t} = \sum_{i=0}^{p} a_i \, \varepsilon_{1,t-i} + \sum_{j=1}^{q} c_j \, Y_{1,t-j} + u_t$$

To calculate IRFs, set $\varepsilon_{1,t} = 1$ and simulate dynamic impact on each y separately.

Romer-Romer typically use this method.

Arezki, Ramey, and Sheng (2017) used this method for oil discoveries.

C. Jorda Local Projections

An IRF is

$$E[Y_{i,t+h} \,/\, \varepsilon_{1t} = 1; controls] - E[Y_{i,t+h} \,/\, \varepsilon_{1t} = 0; controls]$$

Rather than imposing the dynamic structure from a VAR we can instead estimate a simple regression for each horizon h:

$$Y_{i,t+h} = \theta_{i,h} \cdot \varepsilon_{1t} + control\ variables + \xi_{t+h}$$

$\theta_{i,h}$ is the estimate of the IRF at horizon h.

C. Jorda Local Projections (cont.)

* The control variables do not need to be the same in every set of regressions.

* This method imposes fewer restrictions; however, the IRFs are often estimated less precisely.

* By construction, the error terms are moving average, so you need to use HAC standard errors.

* This method is very flexible – can easily extend to:
  * State dependent effects
  * FAVAR
  * IV

* Important thing to note:

Local Projections and VARs Estimate the Same Impulse Responses*

Mikkel Plagborg-Møller
Princeton University

Christian K. Wolf
Princeton University

This version: January 20, 2019
First version: November 16, 2018

Abstract: We prove that linear local projections and Vector Autoregressions (VARs) estimate the same impulse response functions. This nonparametric result only requires the lag structures in the two specifications to be unrestricted. We discuss several implications: (i) Local projection and VAR estimators should not be thought of as conceptually separate procedures; instead, they belong to a spectrum of dimension reduction techniques that share the same estimand but have different finite-sample bias-variance properties. (ii) VAR-based structural estimation can equivalently be performed using local projections, and *vice versa*. (iii) Valid structural estimation with an external instrument (also known as a proxy variable) can be carried out by ordering the instrument first in a recursive VAR, even if the shock of interest is noninvertible. (iv) Local projections are not more "robust to non-linearities" than VARs.

Stata demonstration: Jorda method

We can run the reduced form or a one-step IV

III. Pitfalls

A. Anticipations

B. Trends, lags, outliers and influential observations

C. Counterfactuals

D. Nonlinearities

A. Anticipations or Foresight

* Anticipations or foresight on the part of agents can present serious problems for identification and the correct specification of the counterfactual.

* There are two main foresight problems:

(i) foresight on the part of private agents

(ii) foresight on the part of policy makers

(i) Foresight on the Part of Private Agents:

Examples from Fiscal Literature

This was the whole point of my 2011 QJE paper "Identifying Government Spending Shocks: It's All in the Timing."

I showed that most movements in government spending are anticipated and that failing to incorporate that in a VAR can dramatically change the results.

Consider a simple neoclassical model with government spending:

Effect of Government Spending in a Simple Model

$$y_t = n_t^{0.67} k_t^{0.33}$$

$$c_t + i_t + g_t \leq y_t$$

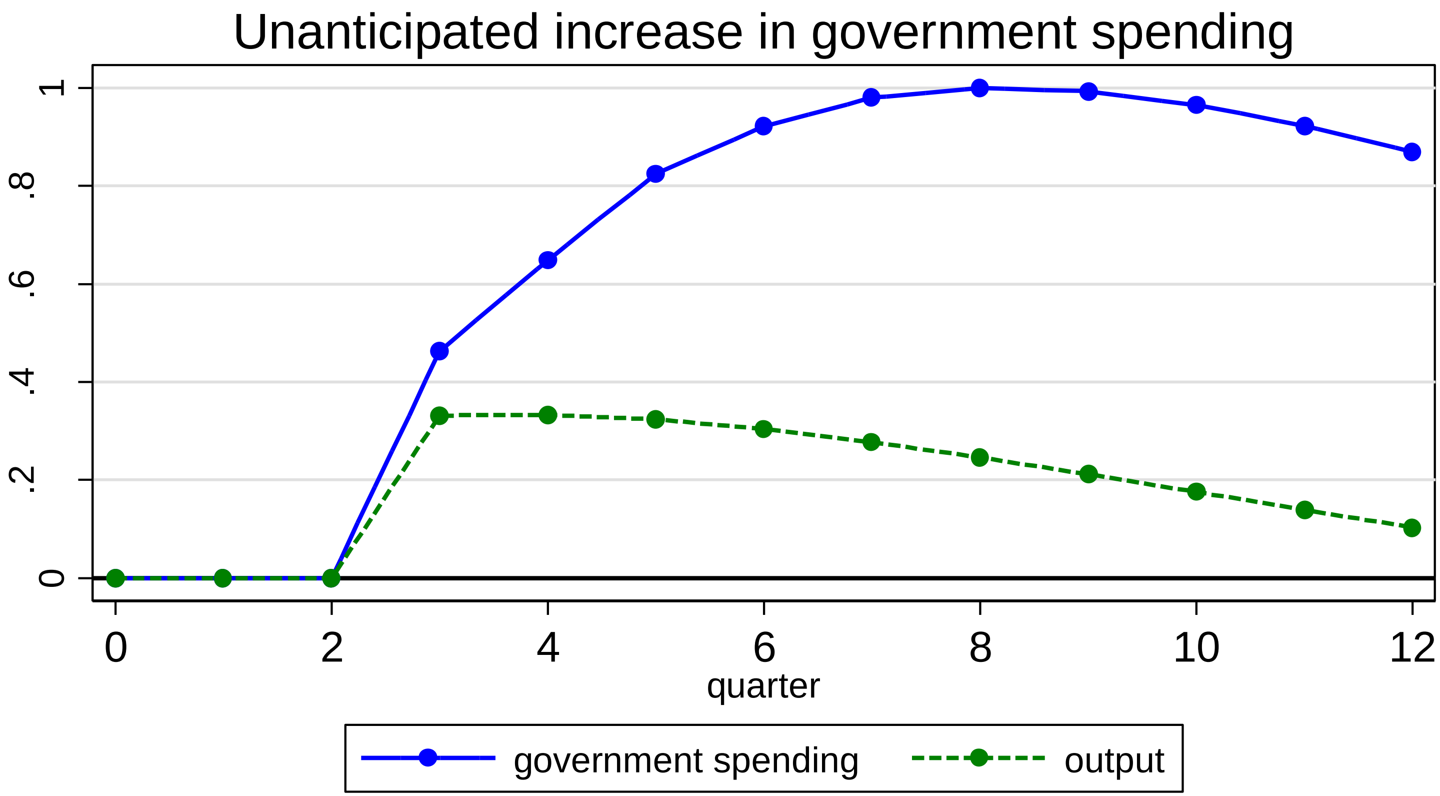
$$u = \ln c_t + 1.75 \cdot \ln(1 - n_t)$$

$$k_{t+1} = i_t + 0.977 \cdot k_t$$

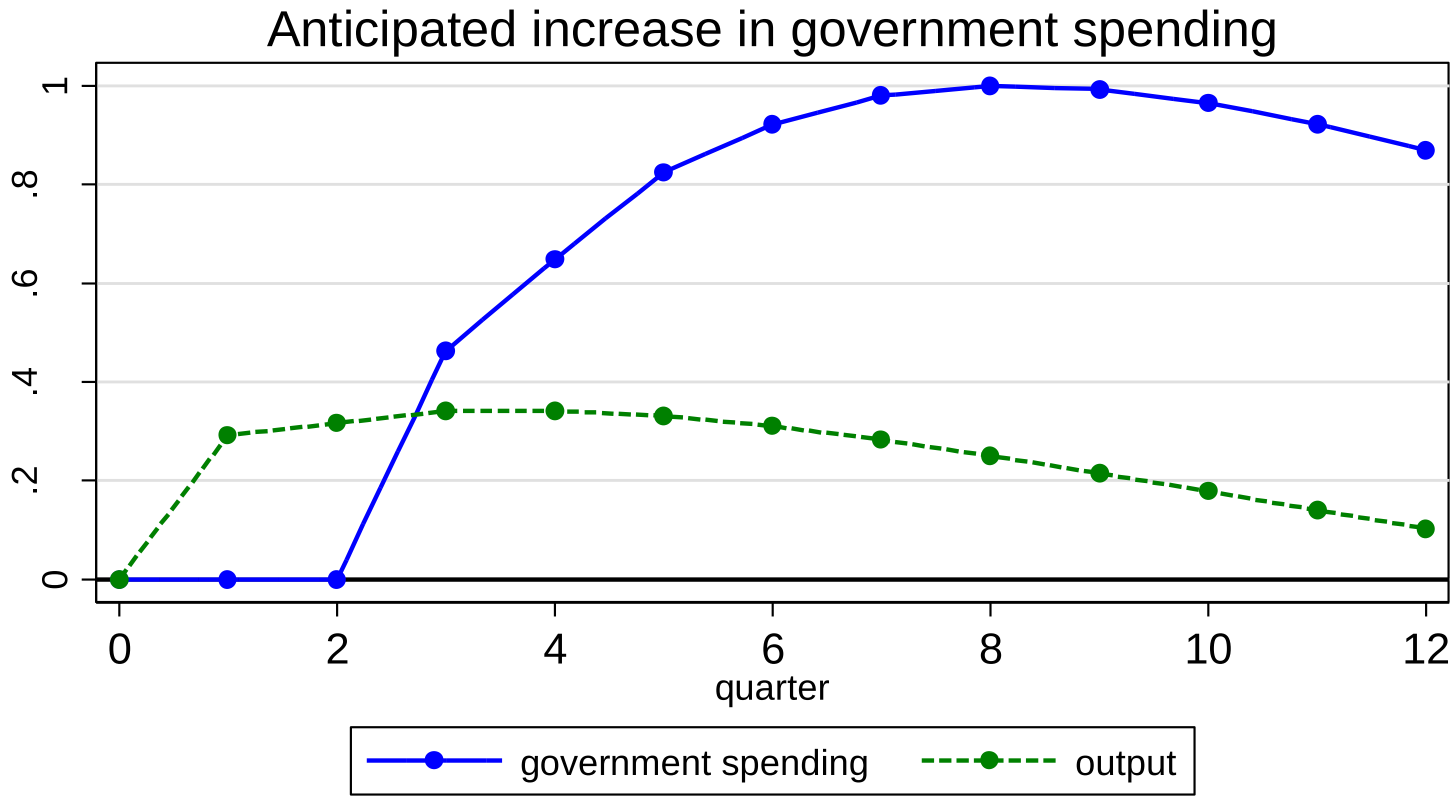
$$g_t = constant + 1.4 g_{t-1} - 0.18 g_{t-2} - 0.25\, g_{t-3} + \varepsilon_t$$

Assume deficit spending, with future lump sum taxes

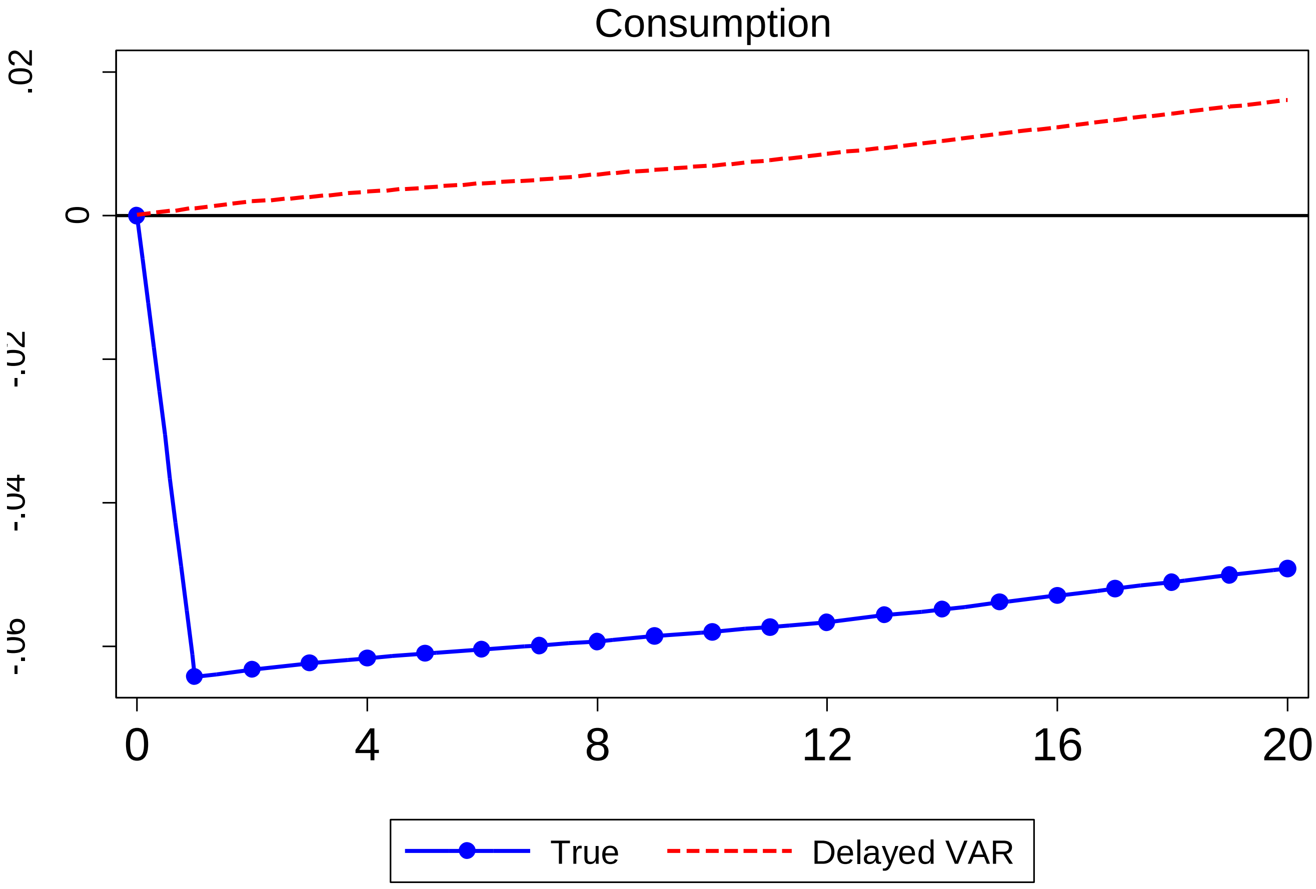
Increase in government spending in quarter 3, unanticipated



Increase in government spending in quarter 3, announced in quarter 1



Errors from using Blanchard-Perotti type identification



Econometrics of Foresight

* Building on work by Hansen and Sargent (1991), Leeper, Walker, and Yang (2013) work out the econometrics of "fiscal foresight" for taxes, showing that foresight can lead to a non-fundamental moving average representation.

* Beaudry, Fève, and Guay (2015) develop a diagnostic to determine whether non-fundamentalness is quantitatively important. They argue that in some cases the non-fundamental representation is close to the fundamental representation.

* The growing importance of "forward guidance" in monetary policy means that many changes in policy rates may be anticipated.

* Mertens-Ravn and my Handbook chapter point out that if foresight leads to a non-fundamental moving average representation in the VAR, one shouldn't use proxy SVAR methods with external instruments.

(ii) Foresight on the Part of Policy Makers

* Sometimes policymakers have more information about the state of the economy than private agents.

* If this is the case, and we do not include that information in the VAR, part of the identified shock may include the endogenous response of policy to expectations about the future path of macroeconomic variables.

* Example: Price puzzle in monetary VARs.

  Sims (1992) argued that the "price puzzle" was the result of typical VARs not including all relevant information for forecasting future inflation. Thus, the identified policy shocks included not only the exogenous shocks to policy but also the endogenous policy responses to forecasts of future inflation.

Application to the Fiscal Context

* Most identification methods have difficulty capturing expectations about the future

* These expectations are key to determining when and how policy makers act.

* Consider the following example in a less subtle setting:

Measuring the Effect of Going to the Hospital

* Question: What is the effect of going to the hospital on the probability of dying in the next 6 months?

* Method: compare individuals who show up at the emergency room to those who don't.

* Controls: body temperature, blood pressure, and pulse rate.

* Comparison: death rate of "treatment group" vs "control group" (those who didn't go to the hospital).

* Result: People who went to the hospital were more likely to die than those who didn't go to the hospital

* Would you refuse to go to the hospital because of this study?

Application to the Fiscal Context
Why do Policy Makers Undertake Fiscal Consolidations?

* Demographics: An increase in the fraction of the population that is older (1) decreases labor supply growth, and hence output growth; (2) increases transfer payments and decreases tax revenues.

* Growth Slowdown: Government tax and transfer programs may have been set up assuming high growth. It takes awhile for politicians to realize the growth slowdown is not temporary. In the meantime, the deficit increases.

* Corrupt leadership: Corrupt leaders pass legislation that lines the pockets of their cronies, distorts economic incentives, raises the deficit, and leads to decreased productivity.

Thus, the action taken may be correlated with the error term, so the instrument may not be exogenous.

B. Trends, Lags, Outliers, and Influential Observations

- Theory can only guide us so far, so we often must make decisions on which theory doesn’t provide enough guidance.

- How variables are detrending can make a big difference.

- Never use a two-sided filter (such as the HP filter) on data used in regressions (See Watson 2008 Mini Time Series notes).

Suppose your data are trending. What should you do?

- Sims, Stock and Watson (1990) demonstrate that even when variables might have stochastic trends and might be cointegrated, the log levels specification will give consistent estimates.

- While one might be tempted to pretest the variables and impose the unit root and cointegration relationships to gain efficiency, Elliott (1998) shows that such a procedure can lead to large size distortions in theory.

- Gospodinov, Herrera, and Pesavento (2013) demonstrate:

- the size distortions can be large in applications.
- impulse response from the levels specification tend to be more robust when the magnitude of the roots is not known.

- Perhaps the safest method is to estimate the SVAR in log levels (usually also including some deterministic trends) as long as the imposition of stationarity is not required for identification. One can then explore whether the imposition of unit roots and cointegration lead to similar results but increase the precision of the estimates.

Suppose you want to estimate the correlation between cyclical components of several series

- Two-sided filters such as Hodrick-Prescott or Baxter-King are the most common way to split a series into trend versus cycle.

- It has long been known that one should never use a two-sided filter (such as the HP filter) on data used in regressions (See Watson 2008 Mini Time Series notes).

- But is it okay to use them to estimate the correlation of the cyclical components of two series?

  - Hamilton (ReStat 2018) says no and proposes a different method.

  - Nekarda-Ramey (in progress) find, however, that Hamilton's alternative method is not robust to the omnipresent low frequency elements of typical macroeconomic series.

Filtering Comparison Example

- Hamilton benchmark filter: To find the cyclical component of variable x, run the following regression and extract the residual as the cyclical component (x^c):

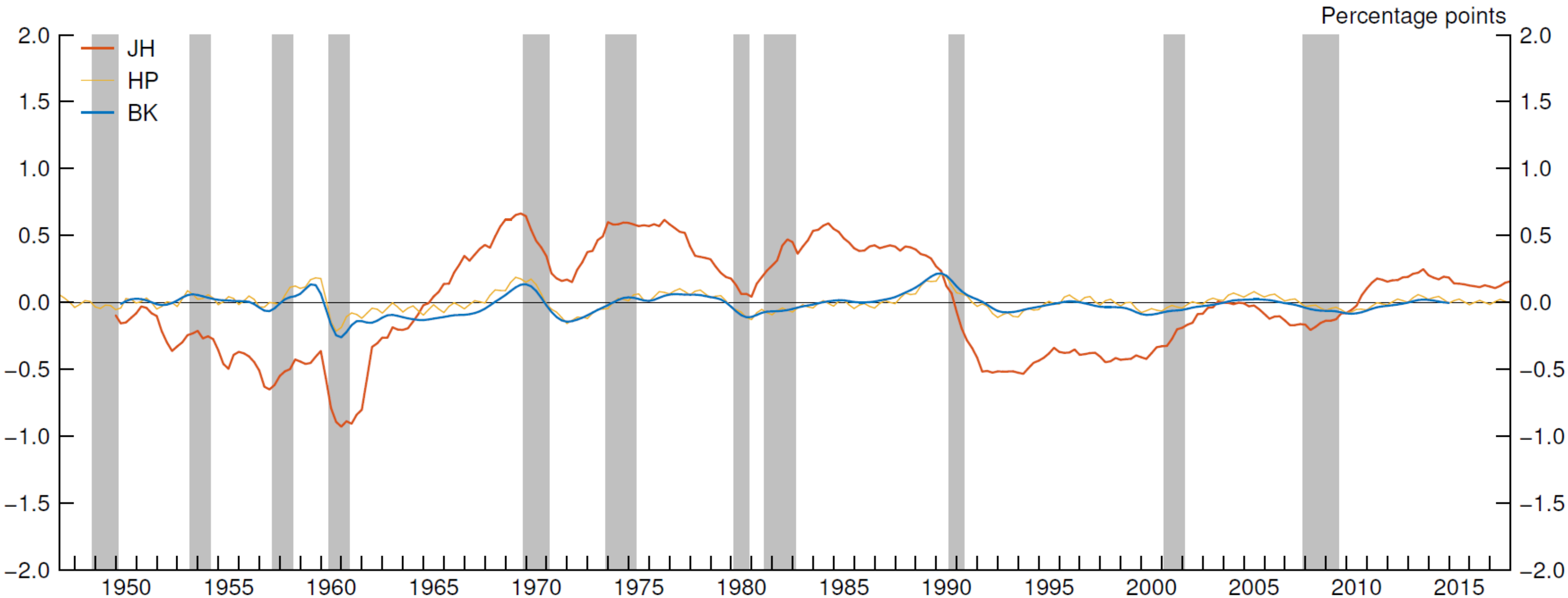
  Reg x on x_{t-8} through x_{t-11} . The residual of that regression is x^c.

- Chris Nekarda and I needed to look at the cyclical behavior of the markup. We looked at several ways to extract the cyclical component from the markup, mu, and GDP (y).

- In one specification, we applied the filtering procedures to log (real GDP), in the other we applied the procedures to log (real GDP/population), i.e. per capita real GDP.

We compared the cyclical component of log real output (ly) and log real output per capita (lyc) using 3 detrending methods (BK, HP, and JH). Figure 1 plots the difference between the detrended series (lyc_cyc – ly_cyc).

Figure 1: Difference in Cyclical Component



For BK and HP, there was essentially no difference between the cyclical components — which is not surprising, because population growth changes at frequencies that those filters remove. However, the JH filter doesn't filter out all of the lower frequency movements.

Bottom Line on Filtering

At this point there is no good solution to the problem of filtering series.

- The problems with HP and BK have long been known and were recently highlighted in Hamilton (2018).

- But Hamilton's proposed alternative might have problems with low frequency movements, as Nekarda and I recently discovered. (We also found issues with deterministic trends and structural breaks.)

In any case, don't filter data before using them in an SVAR!

B. Trends, Lags, Outliers, and Influential Observations (cont.)

- Lags are also an issue – always check sensitivity
- Always plot both the raw data and the partial correlations so you can detect influential observations and outliers.
- Influential observations and outliers: infamous example

QJE May 1991

EQUIPMENT INVESTMENT AND ECONOMIC GROWTH*

J. BRADFORD DE LONG AND LAWRENCE H. SUMMERS

Using data from the United Nations Comparison Project and the Penn World Table, we find that machinery and equipment investment has a strong association with growth: over 1960–1985 each extra percent of GDP invested in equipment is associated with an increase in GDP growth of one third of a percentage point per year. This is a much stronger association than found between growth and any of the other components of investment. A variety of considerations suggest that this association is causal, that higher equipment investment drives faster growth, and that the social return to equipment investment in well-functioning market economies is on the order of 30 percent per year.

Also published extended work in Brookings in 1992

Reassessing the Social Returns to Equipment Investment
Author(s): Alan J. Auerbach, Kevin A. Hassett and Stephen D. Oliner
Source: *The Quarterly Journal of Economics*, Vol. 109, No. 3 (Aug., 1994), pp. 789-802

(1) for the 61-country sample are[1]

$$\underset{}{DYL} = \underset{(0.010)}{-0.017} + \underset{(0.069)}{0.223\, i_E} + \underset{(0.039)}{0.096\, i_S} + \underset{(0.009)}{0.020\, GAP} - \underset{(0.194)}{0.023\, DL}$$

$$N = 61 \quad \overline{R}^2 = 0.322.$$

The central finding is a positive and statistically significant association between the growth of real GDP per worker and the share of real GDP devoted to equipment investment. As shown above, a one percentage point increase in i_E, all else equal, is estimated to boost the average annual growth of real GDP per worker by 0.223 percentage point per year, which cumulates to nearly a 6 percent difference over a 25-year period. The estimate of β_S, though also statistically significant, is less than one-half the size of β_E.

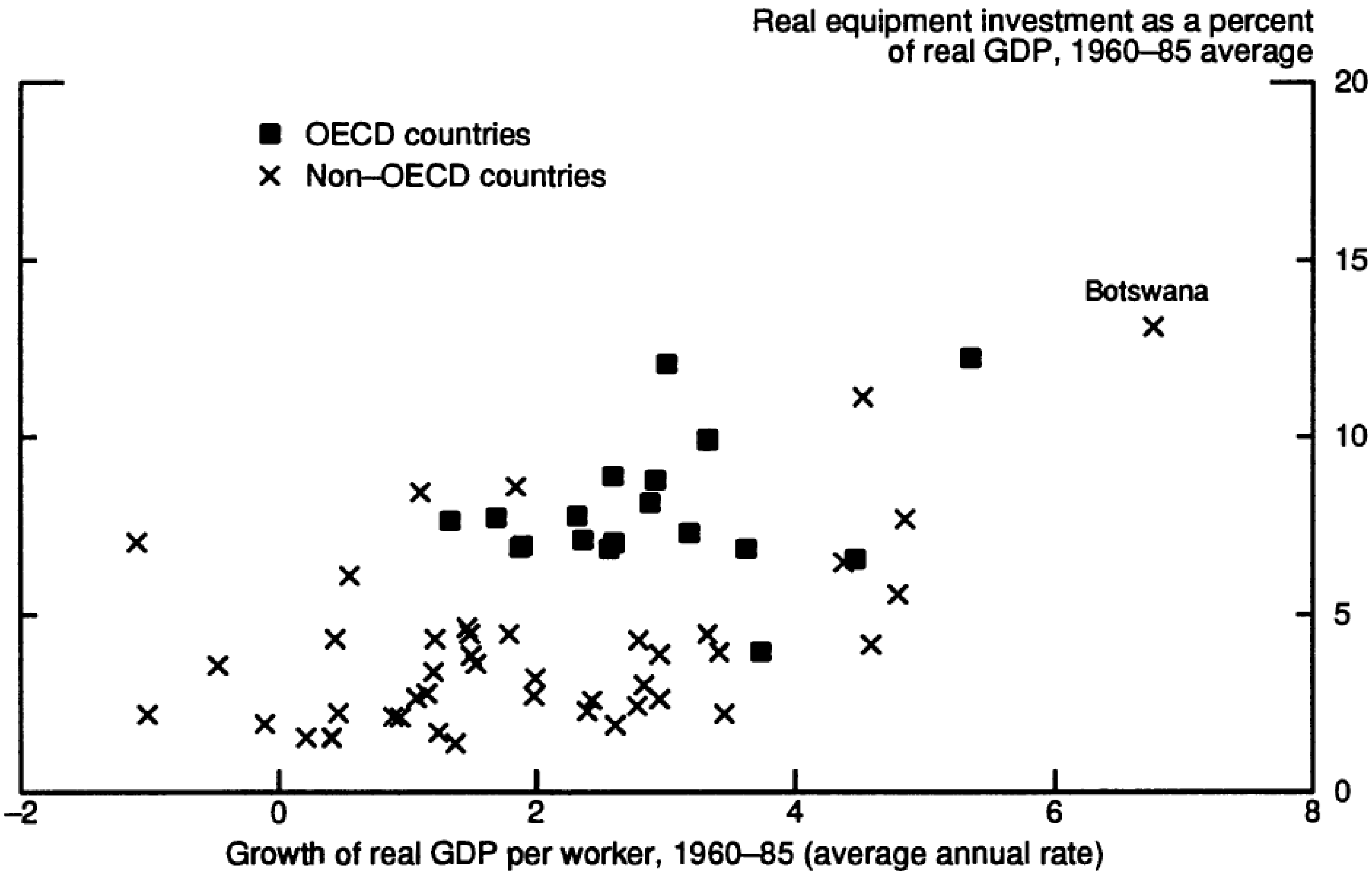


FIGURE I
Real Equipment Investment and the Growth of Real GDP per Worker
Note that data are from the appendix to De Long and Summers [1991].

TABLE I

ESTIMATION OF DE LONG AND SUMMERS' BASIC EQUATION
(STANDARD ERRORS IN PARENTHESES)

Variable	Full sample (1)	High productivity (2)	OECD (3)	Non-OECD (4)	Excl. Botswana: Full sample (5)	Excl. Botswana: Non-OECD (6)
Constant	−0.017	−0.016	−0.001	−0.020	−0.012	−0.020
	(0.010)	(0.009)	(0.008)	(0.021)	(0.010)	(0.020)
Equipment share (β_E)	0.223	0.314	0.028	0.240	0.157	0.114
	(0.069)	(0.071)	(0.083)	(0.099)	(0.075)	(0.115)
Structures share (β_S)	0.096	0.020	0.057	0.094	0.107	0.099
	(0.039)	(0.052)	(0.045)	(0.053)	(0.039)	(0.051)
GDP Gap (Θ)	0.020	0.030	0.035	0.017	0.017	0.015
	(0.009)	(0.009)	(0.008)	(0.020)	(0.009)	(0.019)
Growth of (γ) labor force	−0.023	0.032	−0.025	0.203	−0.046	0.378
	(0.194)	(0.148)	(0.274)	(0.369)	(0.189)	(0.368)
N	61	25	18	43	60	42
$\overline{R}^2$	0.322	0.661	0.674	0.235	0.262	0.129
p-value	0.027	0.004	0.632	0.166	0.179	0.800

Note. All data are from the appendix to De Long and Summers [1991], with corrected values of the structures share for Argentina and Chile. The dependent variable is the average annual growth rate of real GDP per worker over 1960–1985 (DYL). The table shows the OLS estimates of

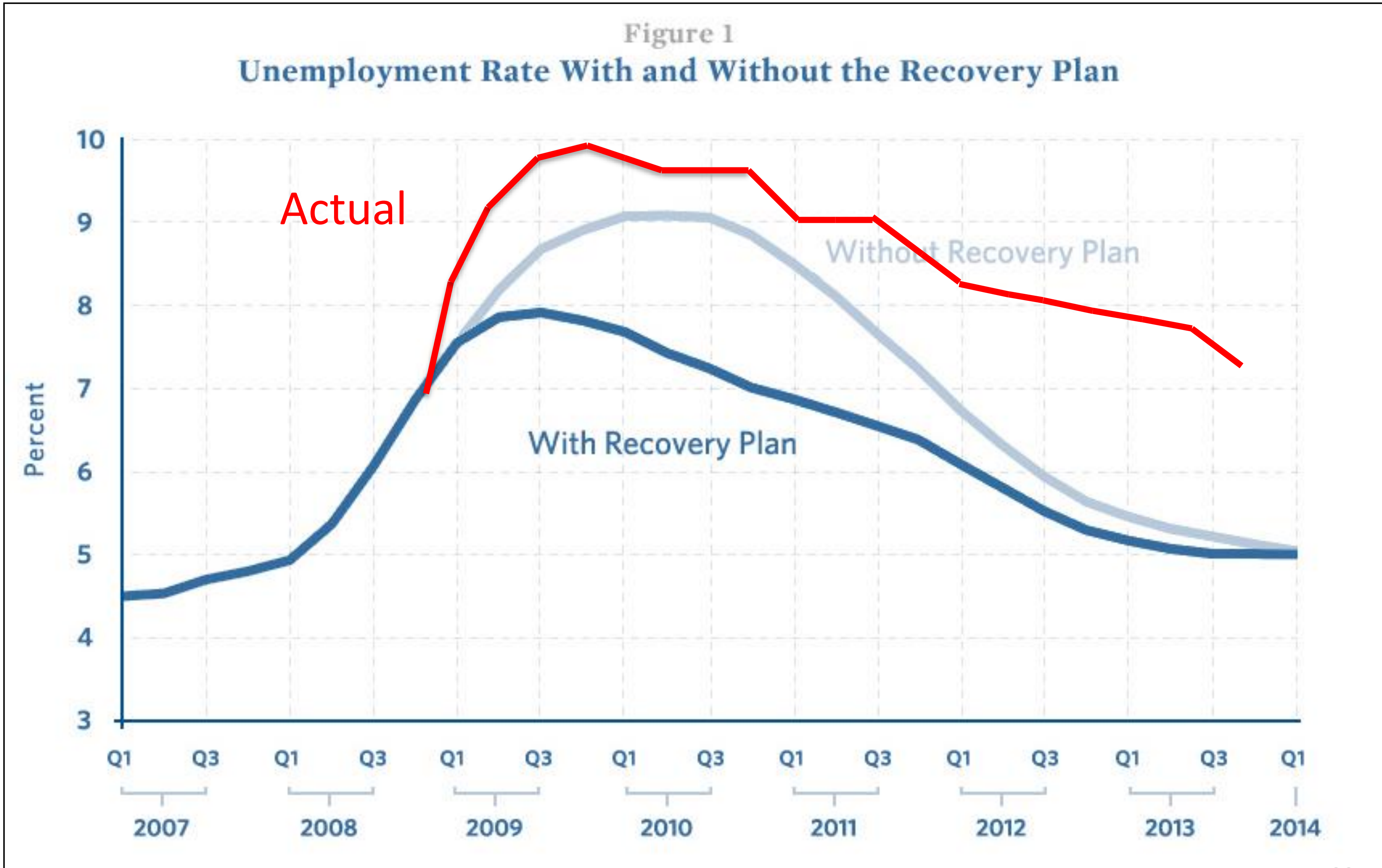
C. Counterfactuals - getting the control group right

- We want to compare the path of the economy after the policy was adopted to what would have happened if the policy hadn't been adopted.

- The typical way to capture the counterfactual is to run an SVAR (using Choleski, narrative, etc.) and include lagged values indicating the state of the economy, such as GDP, hours, interest rates, taxes.

- The impulse response functions then compare what happens after a shock to government spending to *what would have happened had government spending not changed relative to its past.*

From Christina Romer and Jared Bernstein, January 2009



D. Nonlinearities

- There are many cases in which nonlinearities may be important. For example:
  - Asymmetry: positive shocks might have different effects from negative shocks
  - Size effects: effects might not be proportional to the size of the shock
  - State dependence: the effect of a shock might depend on the state of the economy when the shock hits.

- 3 quick points:

1. Koop, Pesaran, and Potter (1996) provide a very useful analysis of the issues that arise when estimating impulse responses in nonlinear models.
2. If one is interested in estimating state dependent models, the Jordà (2005) local projection method is a simple way to estimate such a model and calculate impulse response functions.
3. A caution in testing for a certain kind of nonlinearity:

D. Nonlinearities (cont.)

Kilian and Vigfusson (Quantitative Economics 2011) point about testing for asymmetries:

- Suppose *Y* is a linear function of *X*, where *X* takes on both negative and positive values.

$$Y_t = \beta X_t + \varepsilon_t$$

- Suppose that you instead estimate: $Y_t = \beta' X_t^+ + \varepsilon_t$

  where X^+ are only the positive values of X.

  You find that β' is bigger (in absolute value) than β

- Should you conclude that there are asymmetries, with positive values X having bigger effects than negative values?

- No – this specification leads to bias because you are implicitly If setting all of the negative values of *X* to zero.

- Consider the following graph:

Kilian-Vigfusson (2011)

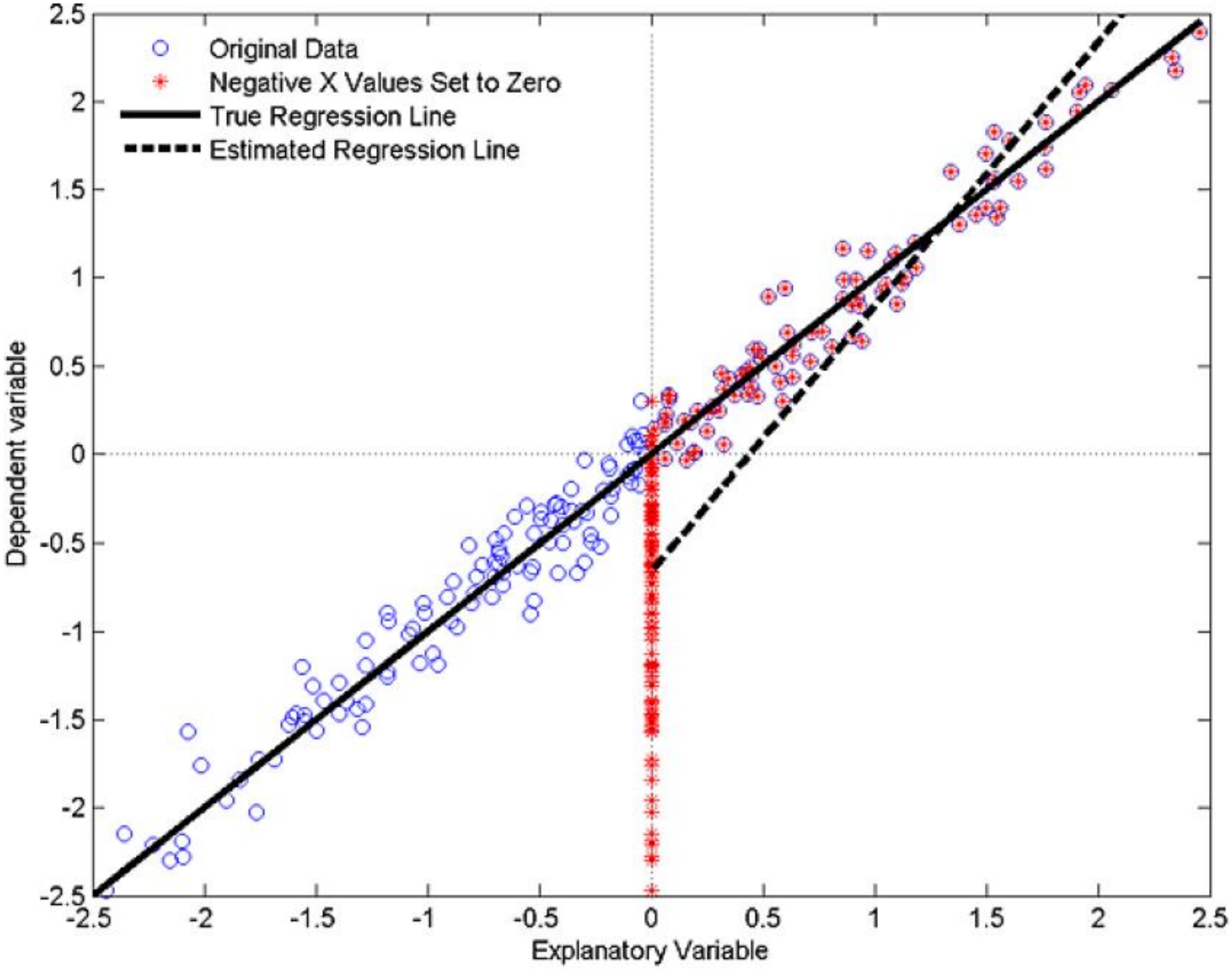


FIGURE 1. The effect of censoring negative values of the explanatory variable.

D. Nonlinearities (cont.)

- Thus, this procedure that truncates on the *X* variable produces slope coefficients that are biased upward in magnitude.

- Thus, one would incorrectly conclude that positive *X*'s have a greater impact than negative *X*'s, even when the true relationship is linear.

- They use an example from the oil literature that had found that oil price increases had bigger effects than oil price decreases.

- To guard against this faulty inference, one should always make sure that the model nests the linear case when one is testing for asymmetries. You should estimate:

$$Y_t = \beta X_t + \beta' X_t^{+} + \varepsilon_t$$

and test whether β' = 0.

D. Nonlinearities (cont.)

- But often narrative methods focus only on one side, such as fiscal consolidations.

- What to do? My solution: in the many contexts, you are implicitly running an IV and you can use this to your advantage.

- Consider the following example:

$Y_t = \alpha + \beta X_t + \varepsilon_t$, where $E(X_t \varepsilon_t) \neq 0$.

$X_t = \gamma + \theta Z_t + u_t$, where Z_t is *independent* of ε_t

Suppose you have only positive values of Z_t, Z_t^+.

Since Z_t exogenous, any function of Z_t is exogenous, including Z_t^+. So you can use Z_t^+ as an instrument for X_t.
Example: Z_t^+ is a narrative of tax increases during fiscal consolidations.
X might be tax revenue or the deficit.

Illustration of Some Pitfalls with an Extended Example from My Own Work

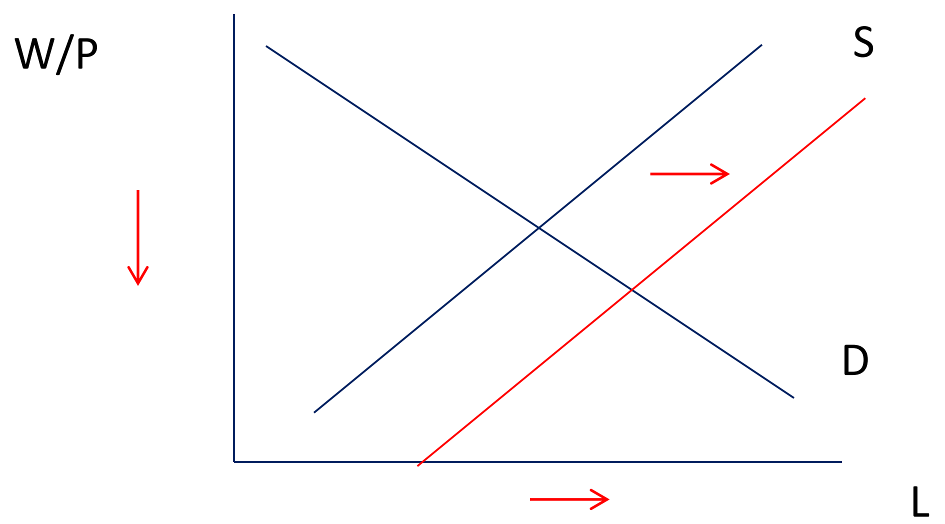
or

“Why I joined the instrument police”

Extended Example

Background: A key controversy is the effect of government spending on real product wages – important for understanding transmission mechanism.

- Neoclassical model predicts that if K doesn't adjust right away, ↑ G → ↓real wages since labor supply increases and there are diminishing returns to labor in the short-run.



Extended Example - continued

- New Keynesian countercyclical markups can overcome this effect.

$$A_t F_L(K_t, L_t) = \mu_t \frac{W_t}{P_t}$$

If $\mu\downarrow$ when L $\uparrow$, then W/P can $\uparrow$ at the same time.

- The empirical results are mixed – Narrative methods find that aggregate real wages fall, Blanchard-Perotti methods find that they rise.

Perotti's Industry Analysis

- In his 2008 NBER Macroeconomics Annual paper, Roberto Perotti asks what happens to real product wages in industries that experienced the greatest increases in military spending during the Vietnam and Carter-Reagan buildups.

- He uses input-output tables to link both direct and indirect government spending to industries (1963 - 1967, 1977 - 1982).

- Perotti ranks industries by the value of $\frac{G_{it}-G_{it-5}}{Y_{it-5}}$, where G_{it} is industry i's shipments to the government in year t and Y_{it-5} is industry i's total shipments in the initial year.

- Perotti examines the top 10 industries in each buildup and notes that real wages rose in 8 of 10 of the industries. He concludes that $\uparrow G \rightarrow \uparrow W/P$.

Assessment

- Perotti's idea of using input-output tables to derive industry-level government spending is terrific. (Extends an older idea by John Shea (1993).)

- However, there are several potential problems with the empirical implementation:

  - What is the counterfactual?

  - Are the instruments relevant?

  - Are the instruments exogenous? In other words, are industry-level government spending shifts uncorrelated with industry technology?

Problem with the Counterfactual

- Perotti's Logic: if $\Delta(W/P) > 0$ in industries with greatest ↑ G from either 1963-1967 or 1977-1982, then neoclassical model is false.

- His implicit counterfactual assumption is that real wages would not have risen if government spending had not risen. Consider the first-order condition again:

$$A_t F_L(K_t, L_t) = \frac{W_t}{P_t}$$

- Perotti was implicitly assuming that *A* and *K* were unchanged over these 4 or 5 year periods. In fact, from 1958 – 1973, average annual growth in economy-wide labor productivity was 3% per year. For 1973-1996, it was 1.5%.

Log Change during Vietnam War
1963-67 (red means rejects neoclassical model)

Industry	Hours	Real wage
Ammunition, excl sm.	116.9	-1.8
Small arms ammun.	101.7	9.6
Oth. Ordnance	41.5	-4.2
Small arms	59.6	6.9
Semiconductor	42.6	44.9
Electronic nec	31.8	25.8
Watches	18.9	9.8
Paving mix	17.6	22.4
Architec metal	19.8	10.0

But what if we compare it to average labor productivity growth from 1958-1973?

Log Change during Vietnam War
1963-67 (green means consistent with neoclassical model)

Industry	Hours	W/P – 12%	K
Ammunition, excl sm.	116.9	-14.2	3.7
Small arms ammun.	101.7	-2.83	11.8
Oth. Ordnance	41.5	-16.6	23.4
Small arms	59.6	-5.5	28.4
Semiconductor	42.6	32.5	45.5
Electronic nec	31.8	13.4	64.2
Watches	18.9	-2.6	12.4
Paving mix	17.6	10.0	16.3
Architec metal	19.8	-2.4	21.0

Nekarda-Ramey Empirical Analysis

- Chris Nekarda and I were intrigued by Perotti's idea so we decided to look into it in more detail ("Industry Evidence on the Effects of Government Spending," AEJ-Macro January 2011)

- We created a full panel data set of 4-digit industries from 1958-2005, merging the NBER productivity database to input-output tables to create government spending by industry.

- In the first version of our paper, we thought that the "semi-conductor" problem was a "fast-growing" industry problem, so we tried to deal with it by modifying Perotti's government variable as follows:

$$\frac{G_{it} - G_{it-5}}{(Y_{it-5} + Y_{it})}$$

Nekarda-Ramey Empirical Analysis

- We controlled for the counterfactual by including both industry and time fixed effects – thus we were comparing the changes in the variables relative to the average in other industries.

- Our modified government demand variable, like Perotti's initial variable, had first-stage F-statistics over 100 for explaining industry output and hours, so both were very relevant.

- However, regressions showed that both our modified variable and Perotti's variable implied industries with greater growth of shipments to the government experienced faster than average labor productivity growth. We thought we had found evidence of increasing returns.

Nekarda-Ramey Empirical Analysis

- Critique during my UC Irvine Seminar: The distribution of government spending across industries is probably correlated with technology.

- They suggested we instead use a Bartik-style instrument. Chris and I studied the suggestion by algebraically decomposing the instrument into a part that could depend on technology and a part that could not. We thus changed our instrument to be:

$\Delta g_{it} = \bar{\theta}_i \Delta \ln G_t$, where $\bar{\theta}_i$ is the time average of $\frac{G_{it}}{Y_{it}}$ and G_t is aggregate government spending.

Nekarda-Ramey Empirical Results

- Our purged instrument was still very relevant

  First-stage F-statistics around 60

- IV regressions with the purged instrument produced estimates suggesting that an increase in output or hours caused by government spending led to:

  - small declines in labor productivity

  - small declines in real product wages

  - rises in the capital stock

  - roughly constant returns to scale

- Thus, the previous findings of increases in productivity and real wages were in part due to the fact that Perotti's instrument wasn't exogenous – it was correlated with technology.

Conclusions from this Exercise

- It is important to get the counterfactual right.

- It is a good idea to look for outliers and influential observations.

- Constructing demand instruments that are correlated with technology will lead to the wrong answers.

Summary General Methods and Pitfalls

1. Good theory and good econometrics are necessary conditions for getting reliable answers to the question: “What are the causal effects of government policy, etc. on macroeconomic variables?”

2. In this section, I have outlined the basic issues and I have surveyed some of the leading methods.

3. I have also highlighted some pitfalls and shown how easy it is to fall into some of these empirical traps.