

The Perils of Identification in Macroeconomics:

Potential Pitfalls and an Extended Example

By Valerie A. Ramey

Applied Microeconomics Point of View

- Angrist and Pischke (2010) – econometric revolution in applied micro
- Macroeconomics is lagging behind on econometric revolution.
- In Macroeconomists' defense, these issues tend to be much harder because of (i) general equilibrium considerations; (ii) the importance of dynamics; and (iii) the importance of expectations.
- However, I think we can still learn a lot from the econometric revolution in micro.

I will use examples from estimating
government spending multipliers

How Do We Estimate the Government Spending Multiplier Relevant for a Short-Run Stimulus Package?

Ideally, the IMF would run a randomized 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 would be financed by deficit spending. The government would commit to a future increase in tax rates to finance the deficit.
- After 2 years, the IMF economists could use **Diff-in-Diff methods** to analyze the data.

How can we reproduce a randomized trial in aggregate time series data?

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

Control Group – all other quarterly or country-year observations (we can also 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 given the dynamics of the economy.

Empirical Challenges in Estimating Effects in Nonexperimental Data

A. Identification

1. Exogeneity
2. Relevance
3. Surprises

B. Counterfactuals

C. Robustness and sensitivity

A. Identification

Identification is at the heart of every empirical effort that is not simple data description.

Identification:

- turns correlations into causal relationships
- is achieved by applying theoretical assumptions to data

A. Identification (cont.)

Consider the classic case of trying to identify demand parameters using data on price and quantity of strawberries.

2 ways to identify the demand parameters:

1. **Identification using additional data:** Find additional variables that **theory** tells us enters the supply curve but not the demand curve, such as rainfall.
2. **Structural Identification:** Use outside estimates or assume the form of the supply curve, impose these parameters on the data and then estimate the demand curve parameters.

Some Issues in Identification

1. Instruments must be exogenous.

Note that exogeneity must always be defined relative to the model.

Is the instrument uncorrelated with the error term in the structural equation?

Typical Ways that Macroeconomists Achieve Identification

1. **Structural identification**

Specify a full DSGE model, make assumptions on the types of shocks and their serial correlation properties, and then estimate the parameters and simulate (Cogan et al (2009)).

2. **Standard VARs achieve identification through timing assumptions.**

Blanchard-Perotti (2002) identify government spending shocks using a Choleski decomposition, ordering government spending first.

3. **Structural VARs (SVARS) bring in outside information.**

Blanchard-Perotti (2002) identify tax shocks using institutional information on the endogenous part of taxes and transfers.

Typical Ways that Macroeconomists Achieve Identification

4. Intuitive arguments

Hall (1980, 1986) and Barro (1981) argued that military spending represents an exogenous shocks to the economy.

5. Narrative Methods

Romer-Romer (1989), Ramey-Shapiro (1998), Ramey (2011) use narrative methods to identify news about increases in military spending.

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

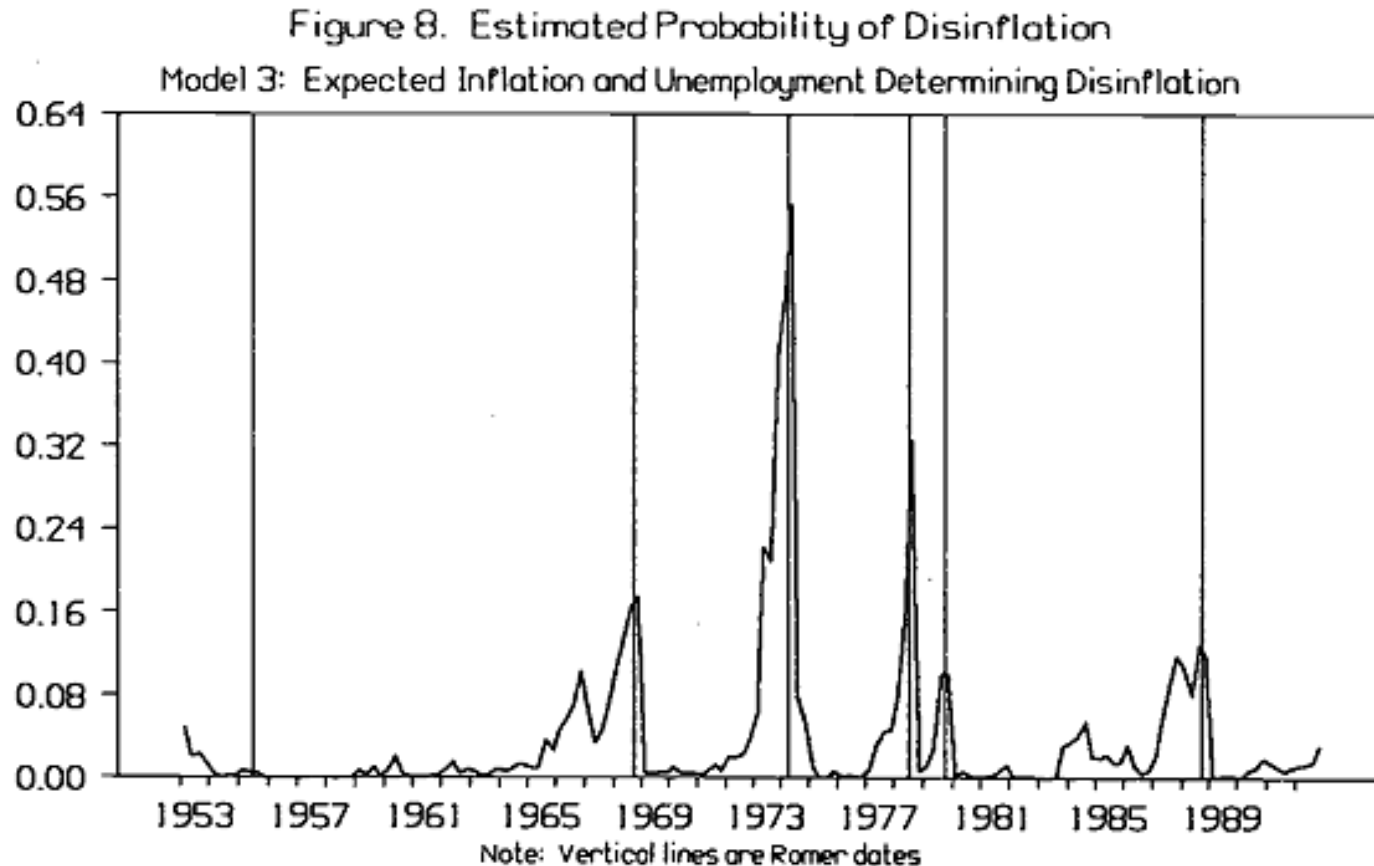
- Romer and Romer (1989) used the narrative approach to identify dates at which Fed decided to reduce inflation.
- They took this as an exogenous shock to policy and then studied the effects.
- We now know that they were estimating the reaction part of policy, not an exogenous shock.

$$i_t = .04 + 1.5(\pi_t - .02) + 0.5(y_t - \bar{y}_t)$$

- 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”

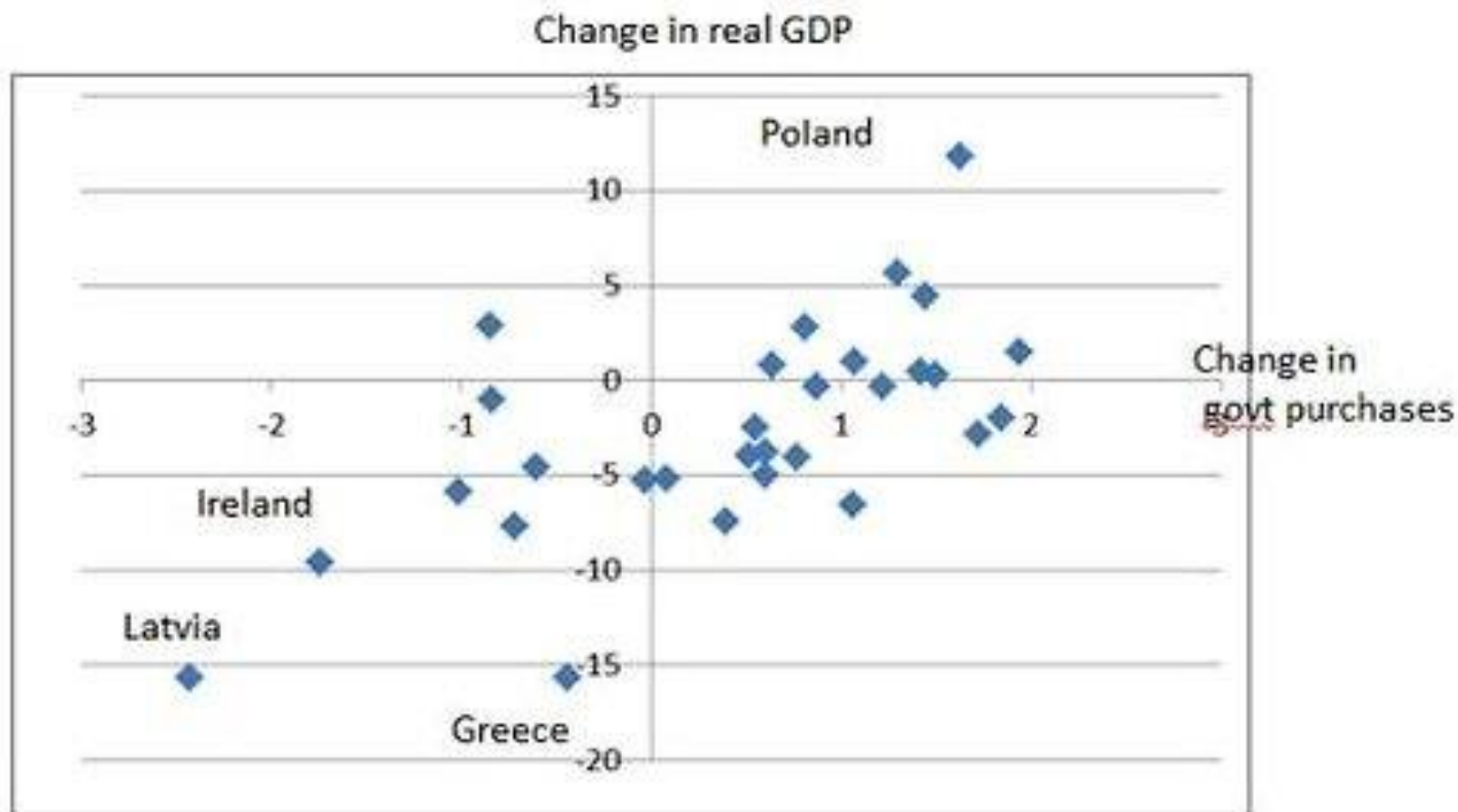


Thus, these dates can't be used to answer the question: What is the independent effect of the Federal Reserve raising interest rates?

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.
- A recent **IMF study on fiscal consolidations** uses a narrative approach, but doesn't take into account the endogeneity of the decision to undertake a fiscal consolidation.
- Consider the following evidence offered by Krugman that fiscal consolidations lower GDP growth:

Krugman's evidence for high multipliers



Reasons Why Fiscal Consolidations May be Correlated with Slow Growth

- **Bad leadership:** Past legislation lines the pockets of cronies, distorts economic incentives, raises the deficit, and leads to decreased productivity.
- **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; (3) causes resources to shift to one of the most distorted and inefficient sectors of the economy (health care).
- **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.

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

More Issues in Identification

2. Instruments must be relevant.

- Most macroeconomists understand the exogeneity requirement, but few seem to be aware of the importance of **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.
- **Rule-of-thumb**: if the first-stage (marginal) F-statistic on your instruments is less than 10, you may have instrument relevance problems.

Examples from Fiscal Empirical Work

- Virtually any shock identified by a standard VAR will have a high first-stage F-statistic, since the shock is defined as the difference between the variable and its projection on lagged values of itself and other variables.
- Other types of instruments:

Variable that is instrumented	Instrument and sample	First-stage F-statistic
government spending	Ramey news, 1947:1 – 2010:4	19.75
government spending	Ramey news, 1958:1 – 2010:4	1.40
government spending	Fisher-Peters stock returns, 1958:1 – 2008:4	2.51
government tax receipts	Romer-Romer exogenous tax variable	7.64

All regressions include 4 lagged values of endog variable, GDP, employment, treasury bill rate, Barro-Redlick avg. marginal tax rate.

Some Issues in Identification

3. The identified shocks must be unanticipated.

This was the whole point of my recent 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.

Also, the work by Leeper, Walker, and Yang works out the econometrics when there is foresight about tax policy.

B. Counterfactuals - getting the control group right

- We want to compare the path of the economy after government spending has increased to what would have happened if government spending hadn't increased.
- 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.*

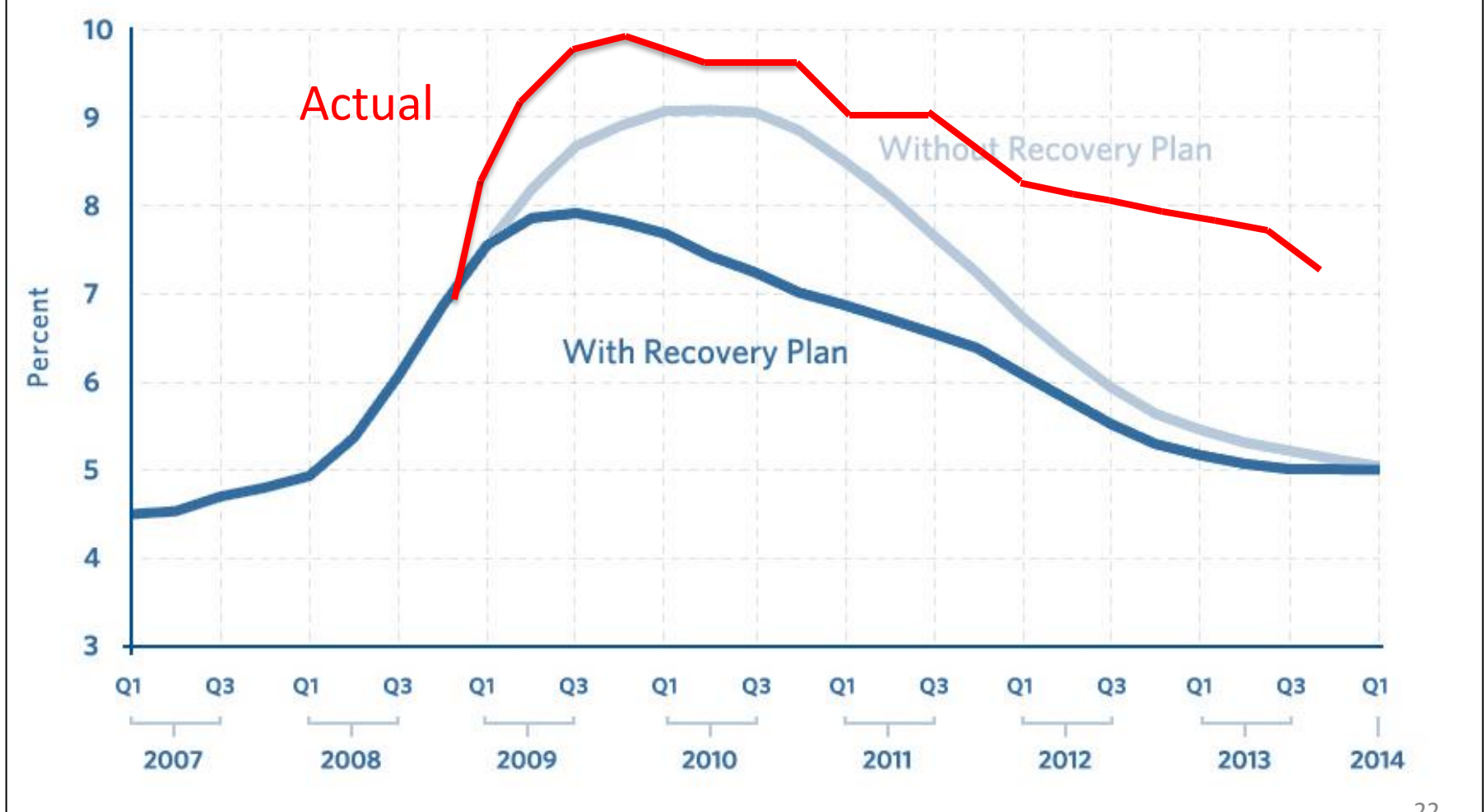
Think about counterfactuals in a less subtle setting: Measuring the Effect of Going to the Hospital

- **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?**

Let's consider a similar recent instance of counterfactual problems:

From Christina Romer and Jared Bernstein, January 2009

Figure 1
Unemployment Rate With and Without the Recovery Plan



C. Robustness and Sensitivity

- Theory can only guide us so far, so we often must make decisions on which theory doesn't provide enough guidance – detrending, number of lags, which control variables to include, how to compute the multiplier. We need to check sensitivity of results to these elements.
- Always plot both the raw data and the partial correlations so you can detect influential observations and outliers.

But do all these issues really matter in practice?

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

COMMENT ON DE LONG AND SUMMERS

791

(1) for the 61-country sample are¹

$$DYL = -0.017 + 0.223 i_E + 0.096 i_S + 0.020 GAP - 0.023 DL$$

(0.010) (0.069) (0.039) (0.009) (0.194)

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

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 .

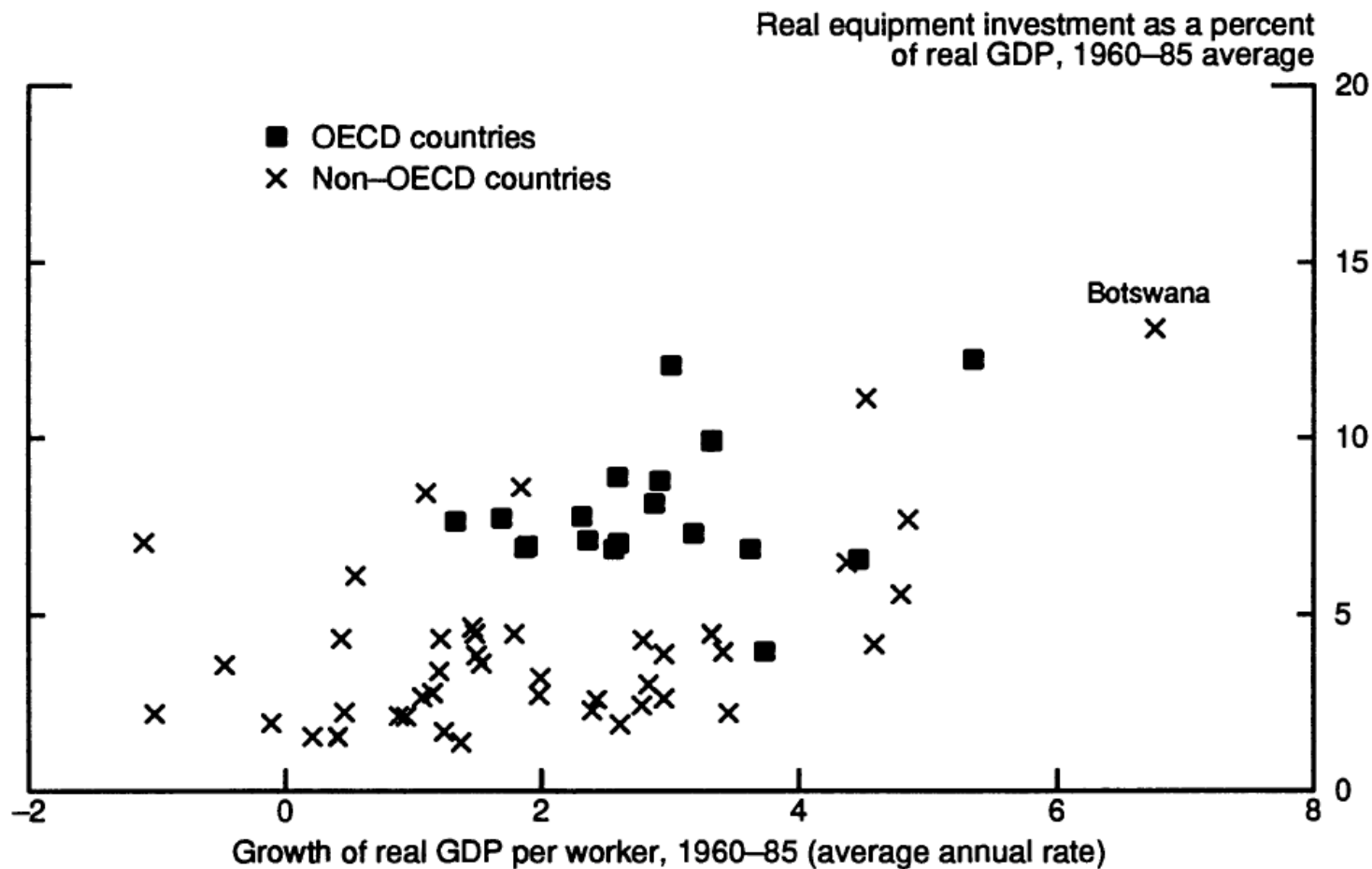


FIGURE I
 Real Equipment Investment and the Growth of Real GDP per Worker
 Note that data are from the appendix to De Long and Summers [1991].

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)	Non-OECD (6)
Constant	-0.017 (0.010)	-0.016 (0.009)	-0.001 (0.008)	-0.020 (0.021)	-0.012 (0.010)	-0.020 (0.020)
Equipment share (β_E)	0.223 (0.069)	0.314 (0.071)	0.028 (0.083)	0.240 (0.099)	0.157 (0.075)	0.114 (0.115)
Structures share (β_S)	0.096 (0.039)	0.020 (0.052)	0.057 (0.045)	0.094 (0.053)	0.107 (0.039)	0.099 (0.051)
GDP Gap (Θ)	0.020 (0.009)	0.030 (0.009)	0.035 (0.008)	0.017 (0.020)	0.017 (0.009)	0.015 (0.019)
Growth of (γ) labor force	-0.023 (0.194)	0.032 (0.148)	-0.025 (0.274)	0.203 (0.369)	-0.046 (0.189)	0.378 (0.368)
N	61	25	18	43	60	42
\bar{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

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, $\uparrow G \rightarrow \downarrow$ real wages since labor supply increases and there are diminishing returns to labor in the short-run.

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

- New Keynesian countercyclical markups can overcome this effect.

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

- 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 $\uparrow G \rightarrow \uparrow W/P$, contrary to neoclassical.

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 **questions we should ask about 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 $\Delta L > 0$ & $\Delta(W/P) > 0$ in industries with greatest $\uparrow G$ from either 1963-1967 or 1977-1982, then neoclassical model is false.

Let G_i denote all defense and civilian purchases by the general government in sector i .³³ Let $\Delta G_{i,67/63}/Y_{i,63}$ and $\Delta G_{i,82,77}/Y_{i,77}$ denote the changes in G_i during the Vietnam War and the Carter-Reagan buildup, as shares of the initial year's industry output. Column 3 of table 3.4 lists the first ten industries in the Vietnam War and the Carter-Reagan buildup by the value of this variable; for each of these industries, column 4 shows the share of real government spending in output in the initial year of each episode, $G_{i,63}/Y_{i,63}$ and $G_{i,77}/Y_{i,77}$. This list appears to make intuitive sense: most industries in it are clearly defense related. The next columns of the table display the percentage changes of real output, of hours, and of the real hourly product wage of production workers. The percentage changes are calculated between the averages during the last two years of the episodes (1966–1967 or 1981–1982, respectively) and the averages during the first two years (1963–1964 or 1977–1978).

Not surprisingly, virtually all these industries experienced a large increase in output and hours. More interestingly, in both episodes the real product wage increased in eight industries out of ten.

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?

Problem with the Counterfactual

- Perotti's implicit **counterfactual assumption** is that real wages would not have risen if government spending had not risen since he is comparing $\Delta(W/P)$ to a 0 threshold. 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 (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:

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

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

TABLE 2—COMPARISON OF GOVERNMENT DEMAND INSTRUMENTS

Instrument	Formula	Sample	Coefficient on instrument for indicated dependent variable	
			Real gross output	Labor productivity
<i>Five-year changes</i>				
1. Perotti (2008)	$\Delta_5 G_{it}/S_{i(t-5)}$	1963–1992, 1,630 obs.	1.294*** (0.122)	0.196*** (0.069)

The real gross output column results show the high first stage F–statistic (112).

But the last column results imply that a demand shock that raises output by 10% raises labor productivity by 1.5%.

This suggests increasing returns!

Nekarda-Ramey Empirical Analysis

- Critique during my UC Irvine Seminar: Perotti's instrument is valid as a demand instrument only if the distribution of government spending across industries is **uncorrelated with technology**.
 - Gary Richardson examples
 - Min Ouyang suggestion (also known as "Bartik instruments")
- Chris and I studied Min's suggestion by algebraically decomposing the instrument into a part that could depend on technology and a part that could not.

Nekarda-Ramey Empirical Analysis

To see this, first define an industry's share of all shipments to the government as $\phi_{it} = G_{it}/G_t$, where G_t is aggregate real shipments to the government. Rearranging this expression relates an industry's shipments to the government to total government spending:

$$(3) \quad G_{it} = \phi_{it}G_t.$$

Differentiating this expression with respect to time yields

$$(4) \quad \dot{G}_{it} = \phi_{it}\dot{G}_t + G_t\dot{\phi}_{it},$$

where a dot over a variable indicates its time derivative. Using equation (4), we can decompose the numerator of Perotti's (2008) measure as

$$(5) \quad \Delta_5 G_{it} \simeq \bar{\phi}_i \times \Delta_5 G_t + \bar{G} \times \Delta_5 \phi_{it},$$

where Δ_5 denotes the five-year difference and $\bar{\phi}_i$ and \bar{G} indicate averages over time.

Consider using $\Delta_5 G_{it}$ as an instrument in a panel data estimation. Including industry and year fixed effects captures any long-run differences in technology across industries and any aggregate changes in technology. The first term in equation (5)

Perotti's instrument also includes lagged industry total shipments in the denominator. Thus, even if one used only the first term of the numerator, there is still a possibility of correlation with technology. Therefore, our measure purges the demand instrument further. To derive our instrument, we divide both sides of equation (4) by S_{it} to obtain

$$(6) \quad \frac{\dot{G}_{it}}{S_{it}} = \frac{\phi_{it}\dot{G}_t}{S_{it}} + \frac{G_t\dot{\phi}_{it}}{S_{it}}.$$

The first term on the right-hand side can be rewritten as

$$(7) \quad \frac{\phi_{it}\dot{G}_t}{S_{it}} = \frac{G_{it}}{S_{it}} \frac{\dot{G}_t}{G_t} = \theta_{it} \frac{\dot{G}_t}{G_t},$$

where $\theta_{it} \equiv G_{it}/S_{it}$ is the fraction of an industry's total shipments that are sent to the government. Approximating the time derivative, we define our government demand instrument as

$$(8) \quad \Delta g_{it} = \bar{\theta}_i \times \Delta \ln G_t,$$

where $\bar{\theta}_i$ is the time average of θ_{it} . In order to construct our instrument at an annual frequency to match the MID, we use aggregate real federal purchases from the national income and product accounts (NIPA).

Because we have substituted the long-run average of θ_{it} , this measure should be uncorrelated with industry-specific technological change for the same reasons given above. It also has intuitive appeal: It weights the percent change in aggregate government spending by the long-run importance of government spending to the industry. We do not include the second term on the right-hand side of equation (6) in our instrument because it is likely to be correlated with technology. As we show next, our instrument remains highly relevant despite discarding this source of variation in G_{it} .

Nekarda-Ramey Empirical Results

TABLE 2—COMPARISON OF GOVERNMENT DEMAND INSTRUMENTS

Instrument	Formula	Sample	Coefficient on instrument for indicated dependent variable	
			Real gross output	Labor productivity
<i>Five-year changes</i>				
1. Perotti (2008)	$\Delta_5 G_{it}/S_{i(t-5)}$	1963–1992, 1,630 obs.	1.294*** (0.122)	0.196*** (0.069)
2. Purged numerator	$\bar{\phi}_i \times \Delta_5 G_{it}/S_{i(t-5)}$	1963–1992, 1,630 obs.	0.897*** (0.121)	–0.001 (0.067)
3. Purged numerator and denominator	$\bar{\phi}_i \times \Delta_5 G_t/\bar{S}_i$	1963–1992, 1,643 obs.	1.541*** (0.128)	0.056 (0.073)
4. NR w/average share	$\bar{\theta}_i \times \Delta_5 G_t$	1963–1992, 1,643 obs.	1.516*** (0.129)	0.067 (0.073)
5. NR w/average share	$\bar{\theta}_i \times \Delta_5 \ln G_t^N$	1963–1992, 1,643 obs.	1.546*** (0.138)	–0.003 (0.078)
6. NR w/1963 share	$\theta_{i1963} \times \Delta_5 \ln G_t^N$	1963–1992, 1,643 obs.	1.490*** (0.139)	–0.017 (0.078)
<i>Annual changes</i>				
7. NR w/average share	$\bar{\theta}_i \times \Delta \ln G_t^N$	1960–2005, 12,536 obs.	1.475*** (0.115)	–0.130 (0.087)
8. NR w/1963 share	$\theta_{i1963} \times \Delta \ln G_t^N$	1960–2005, 12,536 obs.	1.776*** (0.139)	–0.156 (0.104)

Notes: G_{it} is real shipments by industry i to government (IO); S_{it} is real total shipments by industry i ; G_t^N is real federal purchases (NIPA). $\phi_i \equiv G_{it}/G_t$ and $\theta_i \equiv G_{it}/S_{it}$. An overbar indicates a time average. Specification is $\Delta \ln(\text{Dependent variable}_{it}) = \alpha_i + \alpha_t + \beta \text{Instrument}_{it} + \omega_{it}$. All instruments are standardized to have unit standard deviation. All regressions include industry (α_i) and year (α_t) fixed effects. Standard errors are reported in parentheses.

Nekarda-Ramey Empirical Results

- Our purged instrument was still very **relevant**

First-stage F-statistics above 100.

- 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

1. It is important to get the counterfactual right.
2. It is a good idea to look for outliers and influential observations.
3. Constructing demand instruments that are correlated with technology will lead to the wrong answers.