

Macro Ideas with Micro Data

John Grigsby

September 26, 2017

- 1 Idea Formation
 - Faculty Advice
 - My advice
- 2 Presenting Results
 - Bad Examples
 - Good Examples
- 3 Course Advice
- 4 Datasets
 - Household Datasets
 - Firm Datasets
 - Other datasets

Section 1

Idea Formation

Wise words from the faculty

- Stokey: “Writing a paper is like making waffles... you will have to try a few times before one comes out good”

Wise words from the faculty

- Stokey: “Writing a paper is like making waffles... you will have to try a few times before one comes out good”
- Cochrane: “The WSJ is full of research ideas” (or *The Economist*, *The Atlantic*, NYT etc.)

Wise words from the faculty

- Stokey: “Writing a paper is like making waffles... you will have to try a few times before one comes out good”
- Cochrane: “The WSJ is full of research ideas” (or *The Economist*, *The Atlantic*, NYT etc.
- Sargent: “Seriously read one paper every week”

Wise words from the faculty

- Stokey: “Writing a paper is like making waffles... you will have to try a few times before one comes out good”
- Cochrane: “The WSJ is full of research ideas” (or *The Economist*, *The Atlantic*, NYT etc.
- Sargent: “Seriously read one paper every week”
- Alvarez: “Develop one idea every week, most will be unpublishable”

Wise words from the faculty

- Stokey: “Writing a paper is like making waffles... you will have to try a few times before one comes out good”
- Cochrane: “The WSJ is full of research ideas” (or *The Economist*, *The Atlantic*, NYT etc.
- Sargent: “Seriously read one paper every week”
- Alvarez: “Develop one idea every week, most will be unpublishable”
- “Research is a Poisson process

Wise words from the faculty

- Stokey: “Writing a paper is like making waffles... you will have to try a few times before one comes out good”
- Cochrane: “The WSJ is full of research ideas” (or *The Economist*, *The Atlantic*, NYT etc.
- Sargent: “Seriously read one paper every week”
- Alvarez: “Develop one idea every week, most will be unpublishable”
- “Research is a Poisson process, but the harder I work the faster ideas arrive”

Wise words from the faculty

- Stokey: “Writing a paper is like making waffles... you will have to try a few times before one comes out good”
- Cochrane: “The WSJ is full of research ideas” (or *The Economist*, *The Atlantic*, NYT etc.
- Sargent: “Seriously read one paper every week”
- Alvarez: “Develop one idea every week, most will be unpublishable”
- “Research is a Poisson process, but the harder I work the faster ideas arrive”
- Neal: “Research is a group activity.”

Wise words from the faculty

- Stokey: “Writing a paper is like making waffles... you will have to try a few times before one comes out good”
- Cochrane: “The WSJ is full of research ideas” (or *The Economist*, *The Atlantic*, NYT etc.
- Sargent: “Seriously read one paper every week”
- Alvarez: “Develop one idea every week, most will be unpublishable”
- “Research is a Poisson process, but the harder I work the faster ideas arrive”
- Neal: “Research is a group activity.”
- Hurst: “Ideas that are great are often forgotten. Write them down”

Wise words from the faculty

- Stokey: “Writing a paper is like making waffles... you will have to try a few times before one comes out good”
- Cochrane: “The WSJ is full of research ideas” (or *The Economist*, *The Atlantic*, NYT etc.
- Sargent: “Seriously read one paper every week”
- Alvarez: “Develop one idea every week, most will be unpublishable”
- “Research is a Poisson process, but the harder I work the faster ideas arrive”
- Neal: “Research is a group activity.”
- Hurst: “Ideas that are great are often forgotten. Write them down”
- Bonhomme: “Perserverance is the most underrated attribute of a researcher.”

Wise words from the faculty

- Stokey: “Writing a paper is like making waffles... you will have to try a few times before one comes out good”
- Cochrane: “The WSJ is full of research ideas” (or *The Economist*, *The Atlantic*, NYT etc.)
- Sargent: “Seriously read one paper every week”
- Alvarez: “Develop one idea every week, most will be unpublishable”
- “Research is a Poisson process, but the harder I work the faster ideas arrive”
- Neal: “Research is a group activity.”
- Hurst: “Ideas that are great are often forgotten. Write them down”
- Bonhomme: “Perseverance is the most underrated attribute of a researcher.”
- Davis: “Don’t worry. Be Happy. Work Hard”

Wise words from the faculty

- Stokey: “Writing a paper is like making waffles... you will have to try a few times before one comes out good”
- Cochrane: “The WSJ is full of research ideas” (or *The Economist*, *The Atlantic*, NYT etc.)
- Sargent: “Seriously read one paper every week”
- Alvarez: “Develop one idea every week, most will be unpublishable”
- “Research is a Poisson process, but the harder I work the faster ideas arrive”
- Neal: “Research is a group activity.”
- Hurst: “Ideas that are great are often forgotten. Write them down”
- Bonhomme: “Perseverance is the most underrated attribute of a researcher.”
- Davis: “Don’t worry. Be Happy. Work Hard”
- Lucas: “Write the introduction first.”

In short

Keep your eyes open, find ideas from everywhere, and don't be afraid to fail at first. Just keep trying.

Know the question you're trying to answer and get excited about it.

My Approach: Advice I'm not qualified to give

Step 1 - the big picture

- Think about the big picture question you want to answer.
- Don't immediately worry about how to answer it, and don't get bogged down in details of identification/model at first.
- Non-academic articles, every day conversation, and introspection.

My Approach: Advice I'm not qualified to give

Step 1 - the big picture

- Think about the big picture question you want to answer.
- Don't immediately worry about how to answer it, and don't get bogged down in details of identification/model at first.
- Non-academic articles, every day conversation, and introspection.

Step 2 - the small picture

- Break down the big picture question into bite size chunks.
- Read papers to learn but also to identify gaps in the literature
- Attend seminars. Be constructively critical.
- Recombining ideas is very useful here.
- Look for data everywhere.

Continuing the advice

Step 3 - the Poisson process

- You have a question, how do you begin answer it?
- Give yourself time to just sit and think and think whenever you have bandwidth to do so - on the bus/train, talking with peers.
- Be concrete; fiddle with models/data
- Don't be afraid to approach professors even early on
- Talk in informal working groups
- Sometimes a number of small ideas add up to a great idea. Start with simple ideas, and let them develop over time.

Continuing the advice

Step 3 - the Poisson process

- You have a question, how do you begin answer it?
- Give yourself time to just sit and think and think whenever you have bandwidth to do so - on the bus/train, talking with peers.
- Be concrete; fiddle with models/data
- Don't be afraid to approach professors even early on
- Talk in informal working groups
- Sometimes a number of small ideas add up to a great idea. Start with simple ideas, and let them develop over time.

Step 4 - Refinement: the group activity

- Get feedback as frequently as possible
- Working groups, class presentations, seminars, professors, peers etc.

Section 2

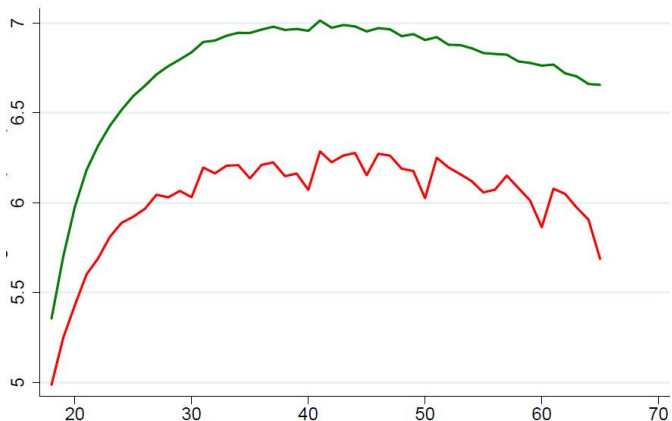
Presenting Results

Successful graphs and charts

- Graphs and charts should stand alone
 - People often read abstract and jump to tables/figures
 - Include sources and every relevant detail
 - Label axes and descriptive titles
- Be printer friendly!
- Present summary statistics clearly
- Graph trends, highlight key periods
- Good ideas can be hidden by poor presentation!
- Know your chart types: line, scatter, bar, map, histogram, CDF plot, Lorenz curve

Bad Example 1: Not enough information

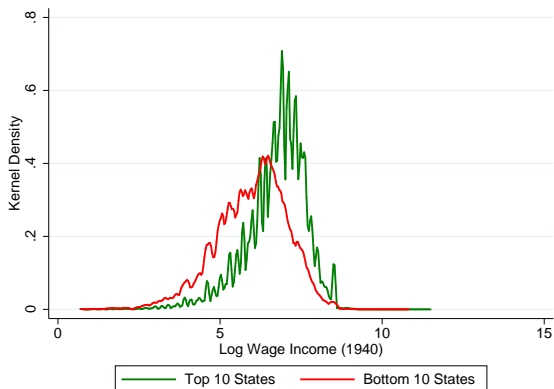
Figure: Life Cycle of Income



Bad Example 2: Colors aren't enough!

It may look fine in full color

Figure: Density plot of income for most and least inventive states

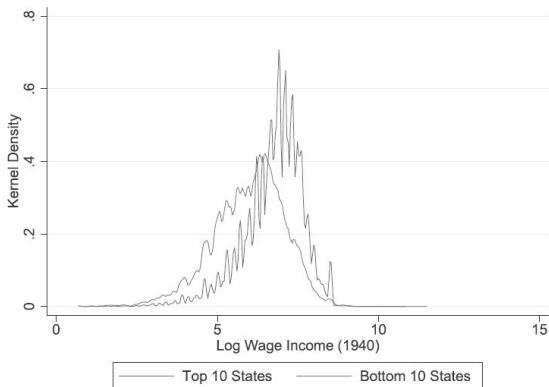


Innovation measured as the average number of patents per capita between 1880 and 1940. The top 10 states are: California, Connecticut, Delaware, Illinois, Massachusetts, Nevada, New Jersey, New York, Ohio, and Rhode Island. The bottom 10 states are: Alabama, Arkansas, Georgia, Mississippi, New Mexico, North Carolina, North Dakota, Oklahoma, South Carolina, and Tennessee.

Bad Example 2: Colors aren't enough!

But unintelligible if printed in grayscale!

Figure: Density plot of income for most and least inventive states



Innovation measured as the average number of patents per capita between 1880 and 1940. The top 10 states are: California, Connecticut, Delaware, Illinois, Massachusetts, Nevada, New Jersey, New York, Ohio, and Rhode Island. The bottom 10 states are: Alabama, Arkansas, Georgia, Mississippi, New Mexico, North Carolina, North Dakota, Oklahoma, South Carolina, and Tennessee.

Good: Figure describes data (Chodorow-Reich 2014)



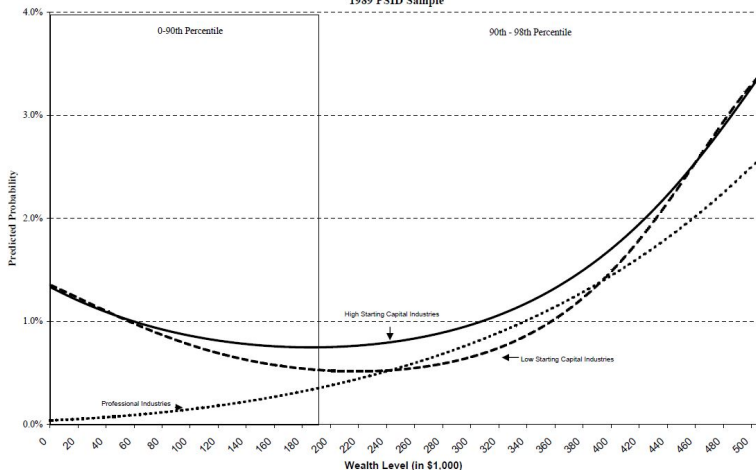
FIGURE I

Stress in the Interbank Lending Market

Source: Federal Reserve Board of Governors (H.15 Release).

Good: Figure tells a story (Hurst & Lusardi 2003)

Figure 2: The Predicted Probability of Entrepreneurship as a Function of Wealth, By Industry:
1989 PSID Sample



Note - This figure displays the predicted probability of starting a business in high starting capital, low starting capital, and professional industries using the non-linear wealth model. Sample was restricted to 1989 non-business owning *PSID* households. High and low starting capital industries are defined in Table A1 of the appendix. For ease of display, the figure only reports wealth levels up through the 98th percentile of the wealth distribution.

Good Example 3: Tables that stand alone (BBD 2013)

Table 2: The Intensity and Composition of Economic Policy Uncertainty in the News Index, by Time Period

Time period	1985:1-1990:6	1990:7-1991:12	1992:1-2001:8	2001:9-2002:12	2003:1-2007:6	2007:7-2008:8	2008:9-2009:12	2010:1-2012:10	1985:1-2012:10
	Mid 1980s to Gulf War I	Gulf War I	1990s boom unil 9/11	9/11 attacks	2000s boom	Beginning of Credit Crunch	Lehman collapse to 'recovery' start	Start of 'recovery' onwards	Overall Average
Overall Economic Uncertainty	217.1	348.0	185.0	325.3	159.0	183.8	369.0	262.8	219.3
Economic Policy Uncertainty	109.0	141.2	87.7	127.8	71.0	83.0	131.5	127.8	100.0
Fiscal Policy	49.4	59.3	35.7	55.1	32.1	32.9	61.2	76.7	45.3
- Fiscal Policy: Taxes	39.7	48.1	31.7	50.9	30.0	31.3	56.6	67.9	39.7
- Fiscal Policy: Spending	22.6	26.7	12.1	17.2	8.5	6.6	17.0	30.6	16.5
Monetary policy	32.5	41.6	25.9	44.9	22.1	31.5	27.6	26.8	28.8
Health care	7.0	15.3	14.9	18.3	13.1	13.4	29.2	39.2	16.3
National security & war	24.9	53.4	17.9	54.5	25.3	15.8	21.2	19.3	24.4
Regulation	15.7	22.9	14.5	19.5	11.1	15.4	29.1	30.4	17.2
- Regulation: financial regulation	3.3	7.0	1.3	5.3	1.7	3.6	10.2	6.8	3.3
Foreign sovereign debt, currency crises	1.4	0.6	2.3	0.5	0.4	0.3	0.4	4.5	1.7
Entitlement programs	7.2	12.5	11.4	18.6	8.8	8.2	15.2	23.4	11.8
Trade policy	3.8	4.0	6.3	2.6	1.7	2.0	1.4	2.3	3.9
Sum of Policy Categories	141.8	209.6	128.8	214.0	114.6	119.4	185.3	222.6	149.4
Ratio of Economic Policy Uncertainty To Overall Economic Uncertainty	0.50	0.41	0.47	0.39	0.45	0.45	0.36	0.49	0.46

Notes: The second row reports average values of our Newsbank Index of Economic Policy Uncertainty in each indicated period (scaling by the total number of articles in a period), expressed as a percentage of the average index value for the entire sample period from 1985:1 to 2012:10. For example, the value of 109 for Economic Policy Uncertainty from 1985:1 to 1990:6 says that the value of the index in that period is 109% of its average value over the full sample period. The top row reports the value of our Newsbank Index of Overall Economic Uncertainty, also expressed as a percentage of the average value of the news-based policy uncertainty index. Entries in Rows 1 to 12 index report analogous values for narrower policy categories based on news article references to specific policy-related terms. For example, the value of 25.8 for "Monetary Policy" from 2010:1 to 2012:10 says that the number of scaled references to monetary policy uncertainty in this period is 25.8 percent of the average number of scaled references to ALL forms of policy-related uncertainty during the 1985:1 to 2012:10 sample period. The categories in Rows 1 through 12 are not mutually exclusive in two respects. First, a given news article may discuss multiple distinct sources of uncertainty such as monetary policy and entitlement reforms. Second, some of the category boundaries overlap. For example, Medicaid is an entitlement program and a major part of the U.S. health care system. News queries run Nov 12, 2012.

Section 3

Course Advice

Advice for the course

- Goes through the whole process of developing, writing, and presenting an idea
- Homeworks are opportunities to develop ideas and get feedback
- Practice writing and presenting results clearly
- Start homeworks early: coming up with ideas is time consuming and difficult at first
- Collaborate!
- Learn how to read papers quickly - extract main idea
- Use that knowledge to structure your papers/proposals

Section 4

Datasets

Gathering Data

- Look for data everywhere; don't be afraid to ask for it!
- Read the data section of papers. If they got the data, it exists and can feasibly be attained.
- Look over public data first. Good centralized places to download include:
 - 1 DataFerrett
 - 2 CensusReporter
 - 3 BLS website
 - 4 NBER Datasets page
 - 5 American FactFinder
 - 6 ATUS-X
 - 7 WRDS (Wharton Research Data Services)
- RA for professors to get access to restricted data while you write your own proposals
- Some colleagues and I maintain a list of data sources available [here](#). Use and update at will!

Cross-Sectional

- **IPUMS** (Integrated Public Use Microdata Series):
 - **CPS** (Current Population Survey): Monthly survey serving as the source for the BLS' official unemployment statistic. Individuals remain in the survey for two 4-month rotations (so partial panel), and has extensive income, demographic, and educational data going back to 1962.
 - **ACS** (American Community Survey): demographic, economic, and migration data going back to 2000. Part of the decennial census.
 - **International**: Harmonized data for 1960 forward, covering 544 million people in 238 censuses from around the world.
- **American Time Use Survey** (ATUS): Representative diaries of how people allocate their time
- **CEX** (Consumer Expenditure Survey): Quarterly survey of consumer expenditures
- **Decennial Census**: Decennial data on population by area, demographics, etc. Complete microdata released with 70 year lag at Library of Congress.

Longitudinal

- **PSID** (Panel Survey of Income Dynamics): Longitudinal income data following households
- **NLSY** (National Longitudinal Study of Youth): Series of panel datasets following young men and women's educational and labor market activities
- **HRS** (Health and Retirement Study): Panel of Americans over 50, following their income, labor force status, and health care
- **Equifax** Consumer Credit Panel - restricted-access credit report data (know someone at the Fed)
- **IRS**: restricted-access tax return data popularized by Chetty et al. Public data available at zipcode level

Additional Labor Market Data

- **JOLTS** (Job Openings and Labor Turnover Survey)
- **QWI** (Quarterly Workforce Indicators)
- **LAUS** (Local Area Unemployment Statistics)
- **ONET**: Information on skill-levels of employees in various occupations
- **CBP** (County Business Patterns): County- and industry-level economic data on establishment counts and payrolls
- **QCEW** (Quarterly Census of Earnings and Wages): County- and industry-level employment and wages going back to 1975

Scanner Data

- IRI Symphony
- Nielsen
 - Homescan: longitudinal consumer panel purchase data across all retail channels
 - Retail Measurement System (RMS): weekly scanner-based sales and causal information gathered from participating grocery, drug, mass merchandiser and convenience stores.
 - MonitorPlus: weekly advertising occurrences, expenditures and impressions across a range of TV media

Firm Datasets - Compustat

- **Compustat**
 - Annual balance sheet and financial data for all publicly-traded companies going back to 1950 (quarterly back to 1962).
 - Accessible through **WRDS** (Wharton Research Data Services)
- Pro: Publicly available
- Pro: Long time frame
- Con: Publicly-traded firms only

Apply for WRDS account **now** if you don't already have one.

LBD and Economic Censuses

LBD (Longitudinal Business Database) is a *restricted-access* census of business establishments and firms in the U.S. with paid employees comprised of survey and administrative records from 1976-present. Covers employment, industry codes, and establishment birth/death.

Plus censuses conducted every 5 years with detailed information on inputs and outputs. Can be linked with LBD

- Census of Manufactures (CMF)
- Census of Services
- Census of Mining
- Census of Retail Trade
- Census of Wholesale Trade
- Census of Transportation, Communications and Utilities

Other firm-level datasets

- Survey of Manufacturing Technology
- Survey of Plant Capacity Utilization
- Capital Expenditure Survey
- Manufacturing Energy Consumption Survey
- Foreign trade data (shipment level data)
- Commodity Flow survey (establishment level data)
- Compustat-LBD Bridge

Employer-Employee Matched Data

In U.S.

- LBD (Longitudinal Business Database)
- IRS
- ADP payroll processing
- Individual state UI claims data (e.g. CA)

Some **foreign countries** with matched data:

- Germany
- France
- Denmark
- Sweden
- Finland
- Colombia

Note: usually quite difficult to get - lengthy applications

Innovation Data

- [NBER Patent Data](#): patent grants, technology class, category, citation network, inventors, assignees
- Check [USPTO](#) (US Patent and Trademark Office) or [EPO](#) (European Patent Office) for longer time frames or information on institutions

Home price indices

- [CoreLogic HPI](#) - Monthly zipcode-level repeat sales home price index, both non-distressed and full U.S.
- [FHFA HPI](#) - Same as above
- [Zillow home price index](#) - Home values by dwelling type

Financial Data

- [SCF](#) (Survey of Consumer Finances): Triannual survey of household balance sheet information run by Federal Reserve Board
- [Call reports](#)
- Publicly-traded security price data
- [Y9-C filings](#): Bank holding company data managed by the Fed. Restricted-access
- [Quarterly Trends for Consolidated Banking Organization](#)
- [SEC EDGAR](#)
- [DealScan](#) - Loan-Level data for syndicated loans
- [SBA](#) (Small Business Administration)

Professional Forecasters

- SPD (Survey of Primary Dealers)
- SPF (Survey of Professional Forecasters)
- Wall Street Journal Economic Forecasting Survey

Other datasets

Macro data

- [FRED](#) (Federal Reserve Economic Data)
- [Groningen Growth and Development Center](#) - Historic International and Sectoral Growth rates
- [BDS](#): provides annual measures of business dynamics (such as job creation and destruction, establishment births and deaths, and firm startups and shutdowns) for the economy and aggregated by establishment and firm characteristics
- [BED](#) - Similar but at establishment level

Geographic Data

- [Census TIGER Line Files](#) - Shape Files for maps
- [NHGIS](#) (National Historical Geographic Information Systems) - Historic shape files and some historic census data
- [NOAA](#) (National Oceanic and Atmospheric Administration) - Climate data

Microdata

- [Survey of Business Owners](#)
- [LEHD](#) (Longitudinal Employer-Household Dynamics): **Synthetic** (i.e. simulated) employer-employee matched dataset
- [Survey of Small Business Finances](#) (SSBF): Owner characteristics, firm size, use of financial services, and the income and balance sheets of the firm collected on small (<500 employees) firms
- [Kauffman Firm Survey](#)
- [SCE](#) (Survey of Consumer Expectations)
- [AHS](#) (American Housing Survey)
- [SIPP](#) (Survey of Income and Program Participation): participation for federal aid programs and welfare.

Health

- [Medicare](#) claims
- Private health records from electronic health reporting (EHR) companies like Athena Health or Epic Systems.
- [HCCI](#) (Health Care Cost Institute) private insurance data
- [Obamacare Health Exchanges](#)
- [FDA](#) (Federal Drug Administration) website has info on [Medical Devices](#) and [Drugs](#)
- [BioMedTracker](#) has clinical trial results data
- [MEPS](#) (Medical Expenditure Panel Survey) - Information on individual medical expenditure and insurance

Education, Advertising, and Geographic Crosswalks

- **FCC** (Federal Communications Commission) has every approved application for advertising air time publicly available.
- **IPEDS** (Integrated Postsecondary Education Data System) has information on every college campus that qualifies for Title IX
- **NSLDS** (National Student Loan Data System) - university level information on student loan amounts and # of recipients.
- **CCD** (Common Core of Data) has K-12 achievement and demographic information for all public school districts in the U.S.

Check out [this site](#) to generate geographic crosswalks for the U.S. (i.e. matching Counties to CBSAs, CBSAs to States, etc.)