ECON 202: Macroeconomics I
Lecture 1 - Overview, Concepts, and Aggregate Data

John Grigsby

January 1, 2017

1. Course Overview
   - Big picture
   - Syllabus
2. Aggregate Data
   - Measuring Output
   - Measuring Prices
   - Labor market indicators
   - Seasonal Adjustment
3. Optimization Review
4. Appendix
   - Household Datasets
   - Firm Datasets
   - Other datasets

Section 1

Course Overview

What is macroeconomics?

Macroeconomics is the study of **aggregate behavior.**

- How do individual decision makers affect the economy as a whole?
- How do we think about "the economy as a whole?"
- How do we measure such a beast?
- How do we think about well-being?

The importance of theory

- Theory helps us understand the world and make predictions
- Empirics, without theory, is the realm of statistics

The importance of theory

- Theory helps us understand the world and make predictions
- Empirics, without theory, is the realm of statistics

Ficciones - Jorge Luis Borges

"To think is to ignore the differences, to generalize, to abstract."

Definitely critique theory, but not simply because it misses out on some special case

The importance of empirics

- We seek to describe the world, not just mathematical constructs
- Data give us a view of the world and discipline our models
- Defense against "mathiness"
- But be wary of crafty empiricists - statistics say *exactly* what they mean, and we often fill in the gaps with falsehood.

Goals of the class

1. Build foundation of economic theory underpinning macroeconomic thought
2. Highlight key trends in macroeconomic data
3. Familiarize you with common data sources
4. Introduce you to (relatively) modern thought on
   - Economic Growth
   - Business Cycles
   - Inflation
   - Asset Pricing
   - Labor Markets
5. Introduce you to key institutions in the U.S. Economy (e.g. Fed)
6. Prepare you for ECON 203 which will cover Fiscal and Monetary Policy

Course mechanics

1. 6 problem sets to practice theory work (20% of grade)
2. 1 empirical problem set in which you write up a macroeconomic fact or event (10% of grade)
3. Midterm in class at beginning of Week 5 (Tuesday Jan 31; 30% of grade)
4. Final during exam week (40% of grade)

Problem sets may be submitted by groups of up to 3 members.

Section 2

Aggregate Data

Aggregation

- How do we go from the price of a coffee in Starbucks of Saieh Hall to price in the economy as a whole?
- How do we measure how well-off a society is?

Aggregation

- How do we go from the price of a coffee in Starbucks of Saieh Hall to price in the economy as a whole?
- How do we measure how well-off a society is?

Must find a way to aggregate micro into macro. For now, we cover:

1. Output measures (GDP and GNP)
2. Price measures (indices, chain weights, CPI, PPI, etc.)
3. Aggregate data sources (FRED, BLS, Census, etc.)
4. Input measures (Labor market, Capital goods)

Tradeoffs of aggregation

Cons:

- Ultimately throwing away information
- Easy to miss important details (e.g. distributions, inequality)
- Can we measure what we want?

Pros:

+ Much easier to look at one number than millions
+ Comparability across nations/times

Important to remember these pitfalls when you do your own work with aggregate data series.

Output measures - measuring well-being?

Two major measures of total output in the economy

1. Gross Domestic Product (GDP)
   - Total value of goods and services produced *in the country* in a certain period of time
2. Gross National Product (GNP)
3. Total value of goods and services produced *by nationals of the country* in a certain period of time

For instance, the cars produced in a Nissan plant in Detroit would count towards the GDP of the USA but the GNP of Japan.

Measuring Output

Suppose we had the following production process

$$\text{Farmer} \xrightarrow{\text{Wheat}} \text{Miller} \xrightarrow{\text{Flour}} \text{Baker} \xrightarrow{\text{Cake}} \text{Consumer}$$

The economy has produced wheat, flour, and cake, which is eventually consumer by the consumer.

But a large part of the dollar value of cake is made up of flour. Counting both would overstate the amount of output people actually consume.

Thus we *only* count the **dollar value of the end use** (cake) in GDP.

Calculating GDP: A simple example

Suppose that the economy only produced pizza and beer in the following way:

	2004	2005
Prices		
Pizza (P)	\$2/slice	\$3/slice
Beer (B)	\$1/can	\$ \$1.10/can
Quantities		
Pizza (P)	50 slices	40 slices
Beer (B)	400 cans	\$ 391 cans

Calculating GDP: A simple example

We can construct nominal GDP as

$$GDP_t^{NOM} = \sum_i p_{it} \times q_{it}$$

where p_{it} =price of good i in period t

q_{it} =quantity of good i in period t

so that

$$GDP_{1997}^{NOM} = (2 \times 50) + (1 \times 400) = \$500$$

$$GDP_{1998}^{NOM} = (3 \times 40) + (1.1 \times 391) = \$550$$

Are people better off?

According to the previous example $GDP^{NOM}_{1997} < GDP^{NOM}_{1998}$. GDP higher so people are better off on average?

No! Notice that the consumption of both pizza and beer actually fell between 1997 and 1998. So people are *worse* off.

Only **prices increased**. Want to make sure that we look at **real activity** not just price adjustments

Real GDP

Solution: Fix prices at a base year:

$$GDP_t^{NOM} = \sum_i p_{ib} \times q_{it}$$

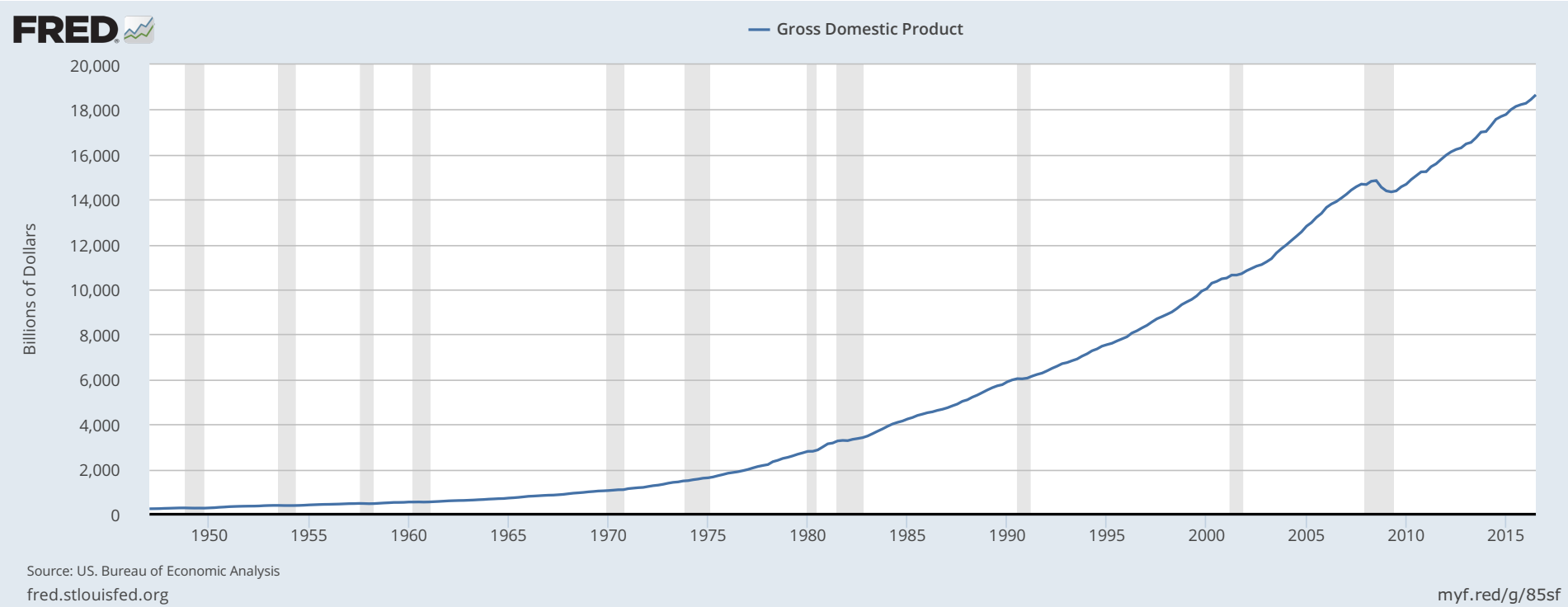
for p_{ib} the price of good i in some fixed base year b. E.g. pick base year = 1997:

$$GDP_{1997}^{REAL} = (2 \times 50) + (1 \times 400) = \$500$$

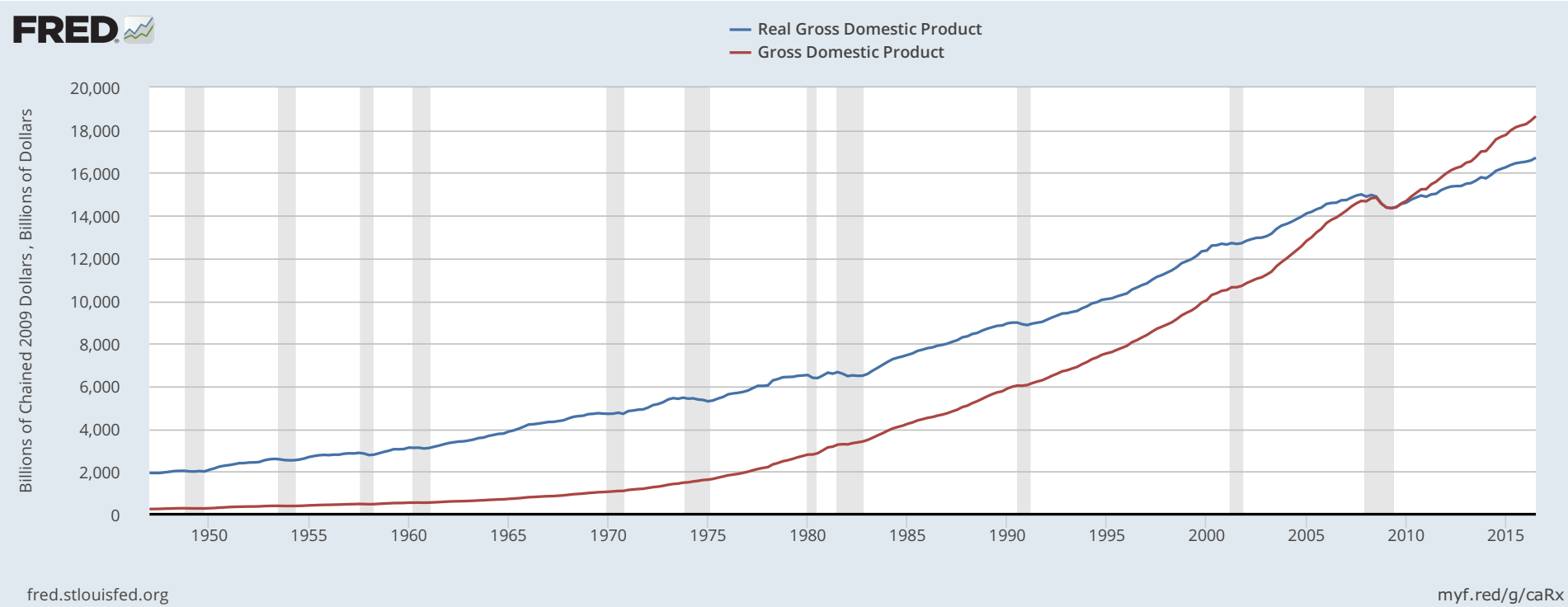
$$GDP_{1998}^{REAL} = (2 \times 20) + (1 \times 391) = \$471$$

So real GDP fell between 1997 and 1998! Adjusting for prices **very** important

Nominal GDP trends



Real (2009 USD) vs Nominal GDP trends



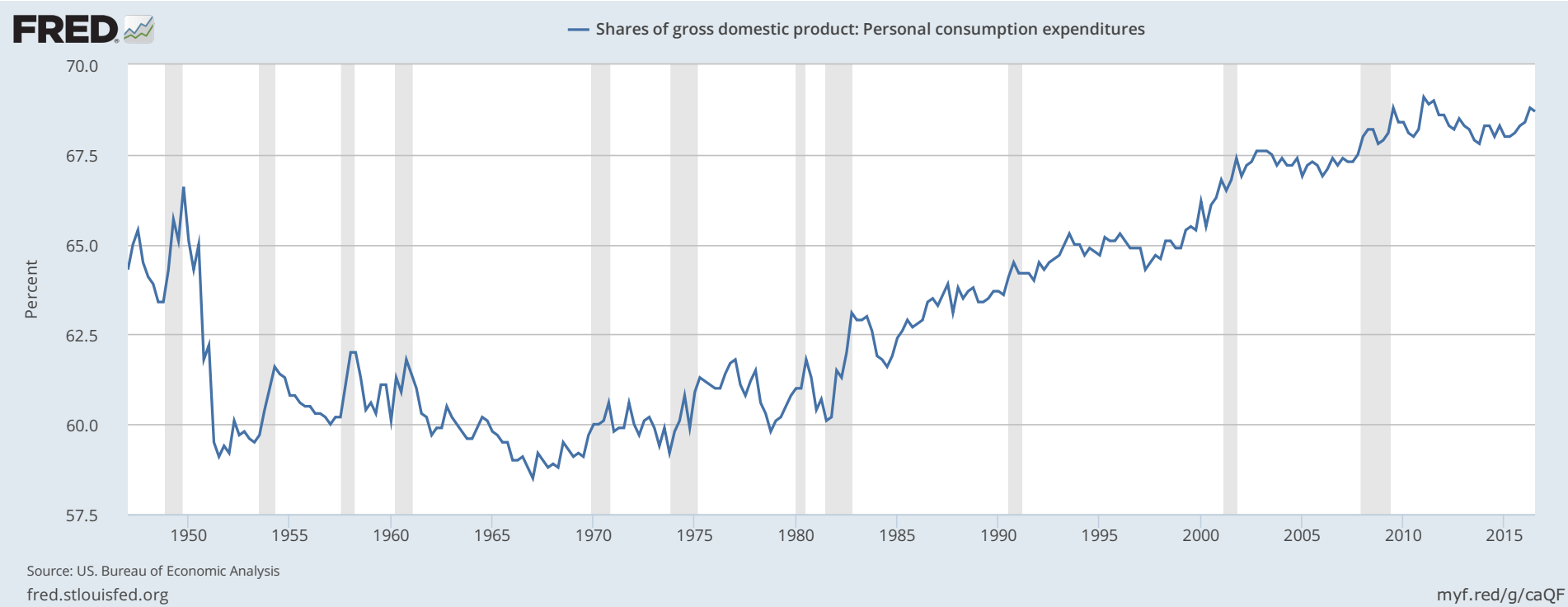
The Components of GDP - Demand Side

Conservation of output: it all must go somewhere!

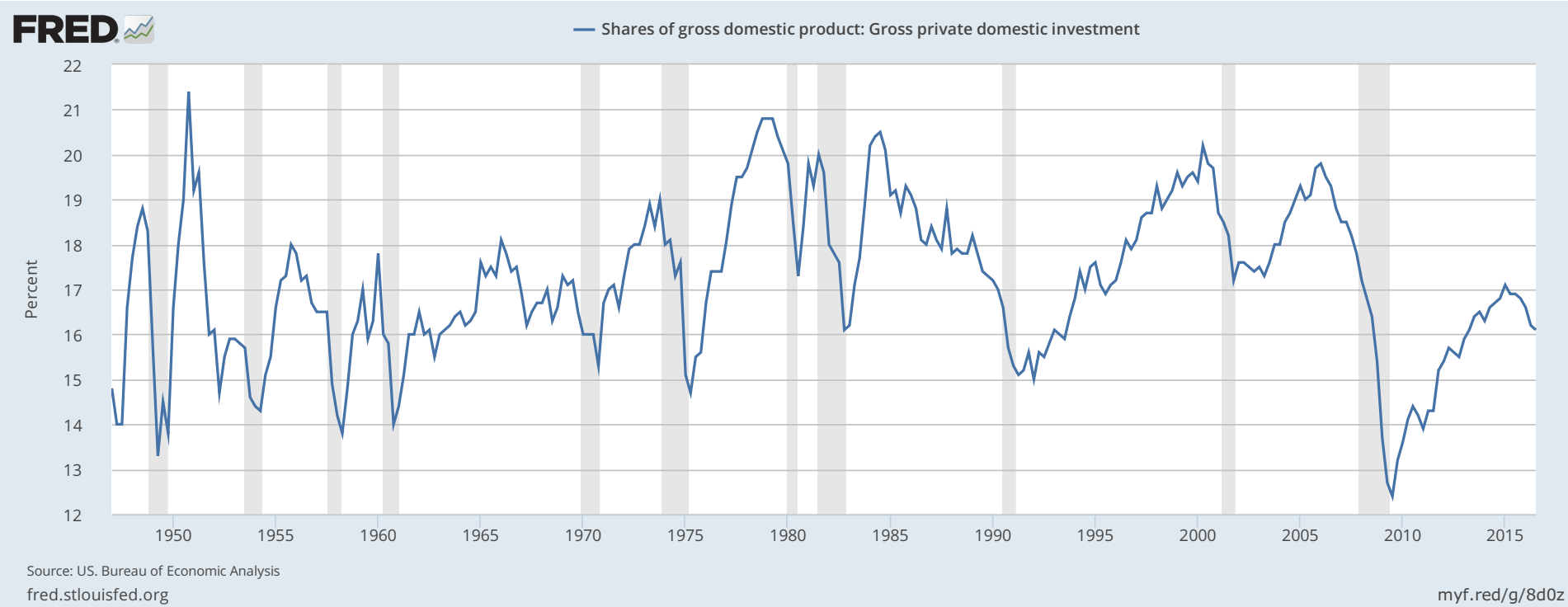
$$\underset{}{\text{GDP}} = \underset{\approx 68\%\\ \text{Stable}}{\text{Consumption}} + \underset{\approx 12\%\\ \text{Volatile}}{\text{Investment}} + \text{Gov. Spend.} + \text{Net Exports}$$

Anything you don't consume is saved. Everything that is saved is invested, either in capital, bonds, under the mattress.

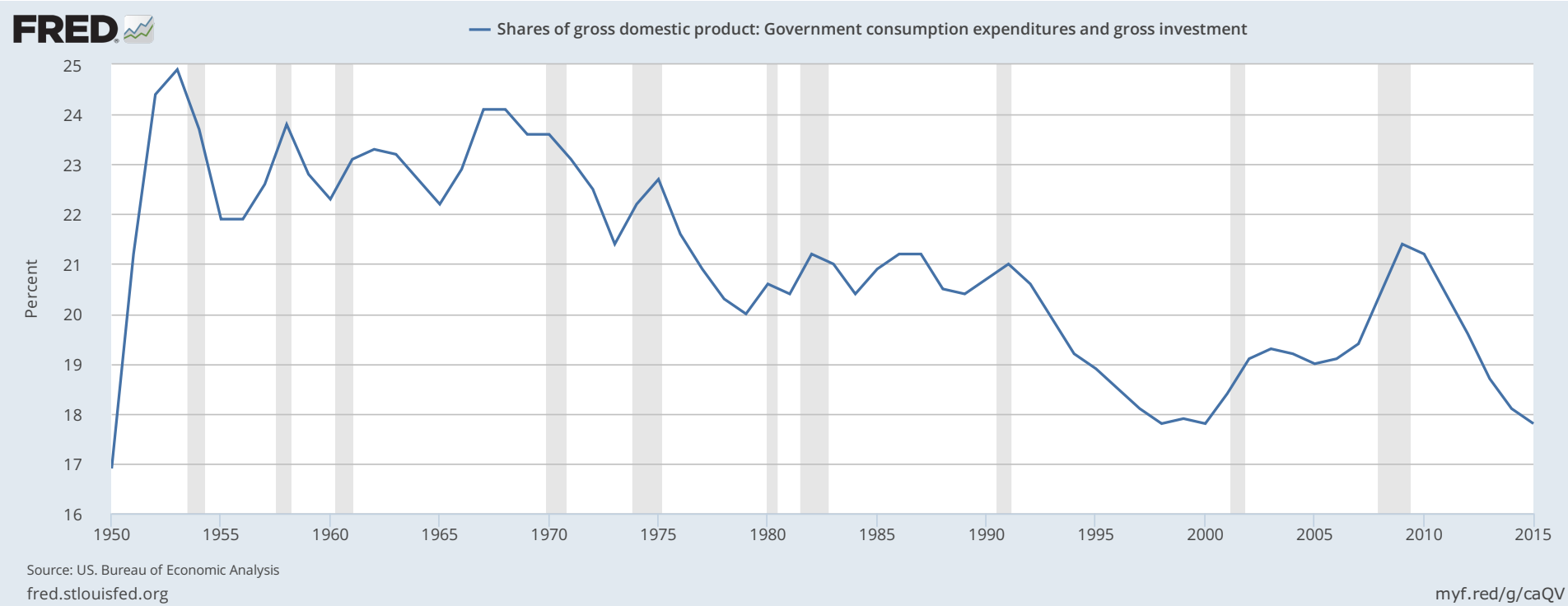
Consumption Share



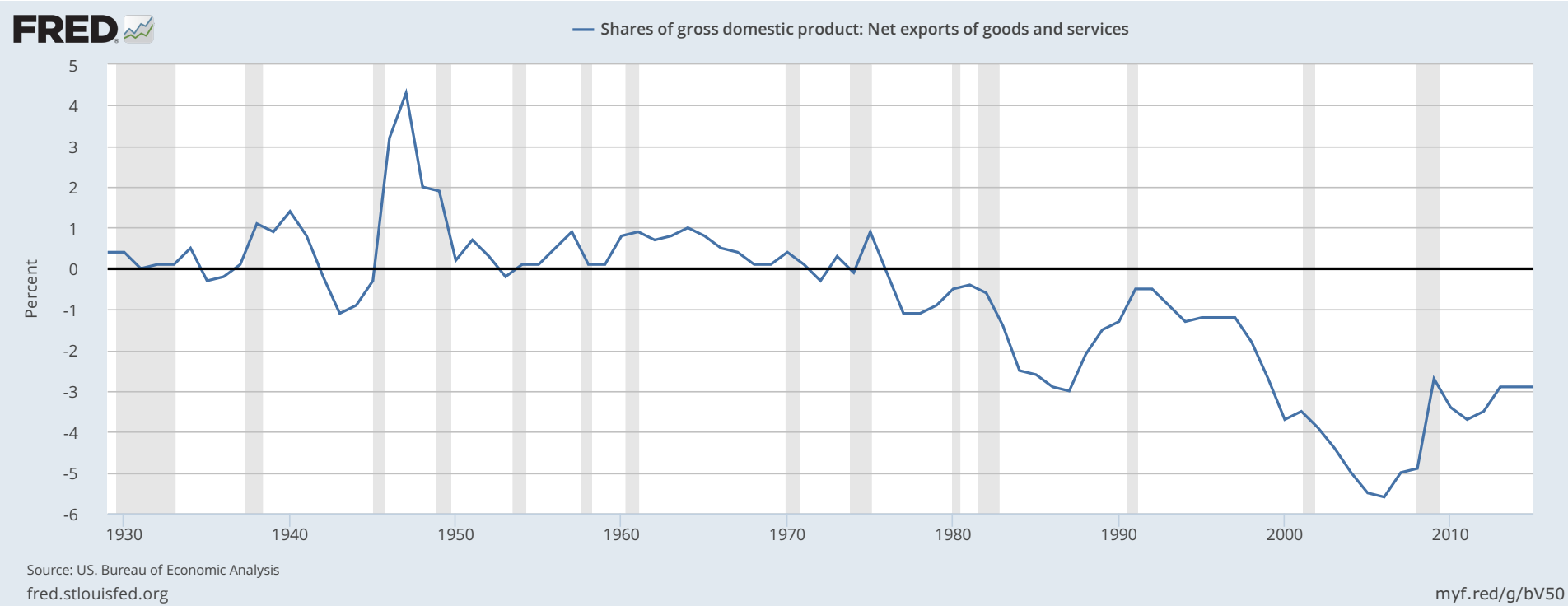
Investment Share



Government Spending Share



Net Exports Share



Problems with real GDP

Suppose we wanted to compare output in 1960 to output in 2010. We could construct real GDP using 1960 as a base year:

$$GDP_{1960} = \sum_i p_{i,1960} \times q_{i,1960}$$

$$GDP_{2010} = \sum_i p_{i,1960} \times q_{i,2010}$$

Several problems with this:

1. Goods that exist in 2010 may not have existed in 1960. What's their price?
2. Ignores quality changes over time, which may show up in prices
3. Highly sensitive to choice of base year
4. People substitute away from relatively high priced goods

See, for example, Steindel for a nice summary of chain weights

The problem restated

Cahill 2002

When relative prices change, households substitute towards purchasing relatively cheaper goods, causing relative production to rise for those products. As a result, goods whose relative prices fall over time tend to see relative increases in production, and vice versa. With a fixed-weighting scheme, this means that for years after the base year, the index weight (i.e., the base years price) for goods with the largest increases in production tend to be relatively large compared to contemporaneous weights, and vice versa. This means that GDP growth is overstated for years after the base year. For years before the base year, goods with the largest increases in production have relatively low base-year price weights, and so GDP growth is understated. This in turn creates a further problem: when the base year is periodically changed, historical real GDP growth figures are adjusted downward.

(Partial) Solution: Chain-weighting

Before we were throwing out data! Instead, use geometric average of current and previous year to construct indices:

$$Y_t = \sqrt{\underbrace{\left(\frac{\sum_i p_{it-1} \times q_{it}}{\sum_i p_{it-1} \times q_{it-1}}\right)}_{\text{Lespeyres Quantity Index}} \underbrace{\left(\frac{\sum_i p_{it} \times q_{it}}{\sum_i p_{it} \times q_{it-1}}\right)}_{\text{Paasche Quantity Index}}} \times Y_{t-1}$$

for Y_t real GDP in period t.

Chain-weighting: comparing 1960 and 2010

Fix 1960 to be base year so that 1960 nominal GDP = 1960 real GDP.

$$
\begin{aligned}
Y_{2010} &= \sqrt{\left(\frac{\sum p_{2009} \times q_{2010}}{\sum p_{2009} \times q_{2009}}\right)\left(\frac{\sum p_{2010} \times q_{2010}}{\sum p_{2010} \times q_{2009}}\right)} \times Y_{2009} \\
&= \sqrt{\left(\frac{\sum p_{2009} \times q_{2010}}{\sum p_{2009} \times q_{2009}}\right)\left(\frac{\sum p_{2010} \times q_{2010}}{\sum p_{2010} \times q_{2009}}\right)} \\
&\times \sqrt{\left(\frac{\sum p_{2008} \times q_{2009}}{\sum p_{2008} \times q_{2008}}\right)\left(\frac{\sum p_{2009} \times q_{2009}}{\sum p_{2009} \times q_{2008}}\right)} \times Y_{2008} \\
&= \prod_{t=1960}^{2010} \sqrt{\left(\frac{\sum p_{t-1} \times q_t}{\sum p_{t-1} \times q_{t-1}}\right)\left(\frac{\sum p_t \times q_t}{\sum p_t \times q_{t-1}}\right)} \times Y_{1960}
\end{aligned}
$$

Aggregate Price Level

- *Relative* prices of good drive consumption and substitution patterns across goods
- But the actual *level* of the price **does not matter**
- Consider two economies: one where everything costs P and one where everything costs $2P$, and people have twice as much income. Intuitively nothing changes: everyone can purchase just as much as before
- This is why we deflate nominal GDP to real levels: we care about consumption not prices

Nevertheless, there are reasons why we care about **inflation** (i.e. increases in the price level), which is usually $\approx$2% per year

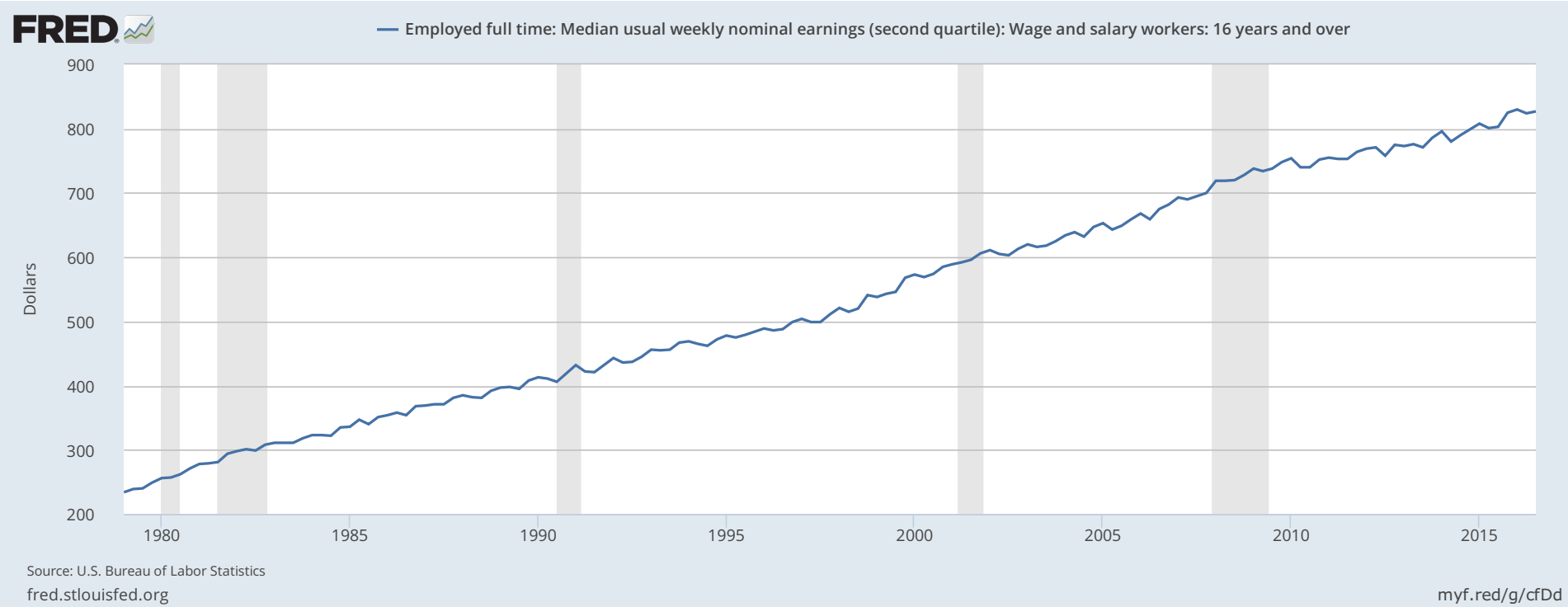
Inflation - why do we care?

We care about inflation because the world is not perfectly flexible. For example:

- Bond holders get paid fixed coupons (e.g. 1% on your savings account). Whether inflation is 2% or 1% matters for the investor's valuation of the bond.
- Workers held to annual contracts/salaries, which may be hard to adjust. Inflation of 2% per year means that their wages will be worth $\approx$ 2% less at the end of the year
- Historical periods of hyperinflation/deflation (Great Depression, Japanese credit crunch, German hyperinflation leading to rise of Nazi party) dramatically affect incentives to save/consume/produce
- How do we index government programs?

Median Wages: Real vs Nominal

Median Nominal Wages Rising Steadily



Median Wages: Real vs Nominal

But Median Real Wages much more cyclical

How do we measure aggregate prices?

- The Bureau of Labor Statistics (BLS) produces various price indices.
- Most commonly used is Consumer Price Index (CPI) which measures prices that households face
  - Calculated by calculating weighted average prices that household faces, weighted by the consumption share of each item
  - Build "basket of goods" that households consume (e.g. healthcare, food, gas, video games, etc.) and track price of that basket over time

$$\text{Inflation}_t = \frac{\sum_i p_{i,t} q_{i,t-1}}{\sum_i p_{i,t-1} q_{i,t-1}}$$

- Also consider GDP Deflator, Producer Price Index (PPI) etc.

Limitations of aggregate inflation measures

1. Subject to same substitution concerns as real GDP; also exist chain-weighted indices
2. Recent research (Argente & Lee (2016), Jaravel (2016)) shows that rich and poor people face different inflation rates
3. Generally misses sectoral differences: healthcare prices skyrocketing, while food & gas volatile up and down

Be wary of these limitations when evaluating policy/investment vehicles.

The labor market - reasons to care and be cautious

Reasons to care

1. Labor important input to production function (about 2/3 of firms' expenses are on labor)
2. Labor income by far largest part of median person's income

Reasons to be cautious

1. How do we measure "labor?"
   - Input to production? Hours and human capital matter
   - Worker well-being? Hours and wage matter
2. Aggregate measures can miss important distributional considerations (inequality)

Employment

MEASURE 1: **Employment-to-population ratio**

$$\frac{\#\text{ people employed}}{\text{Total }\#\text{ of people}}$$

- Usually condition on working age (18-65) or prime age (25-55) people
- $\approx$ 70% in U.S.

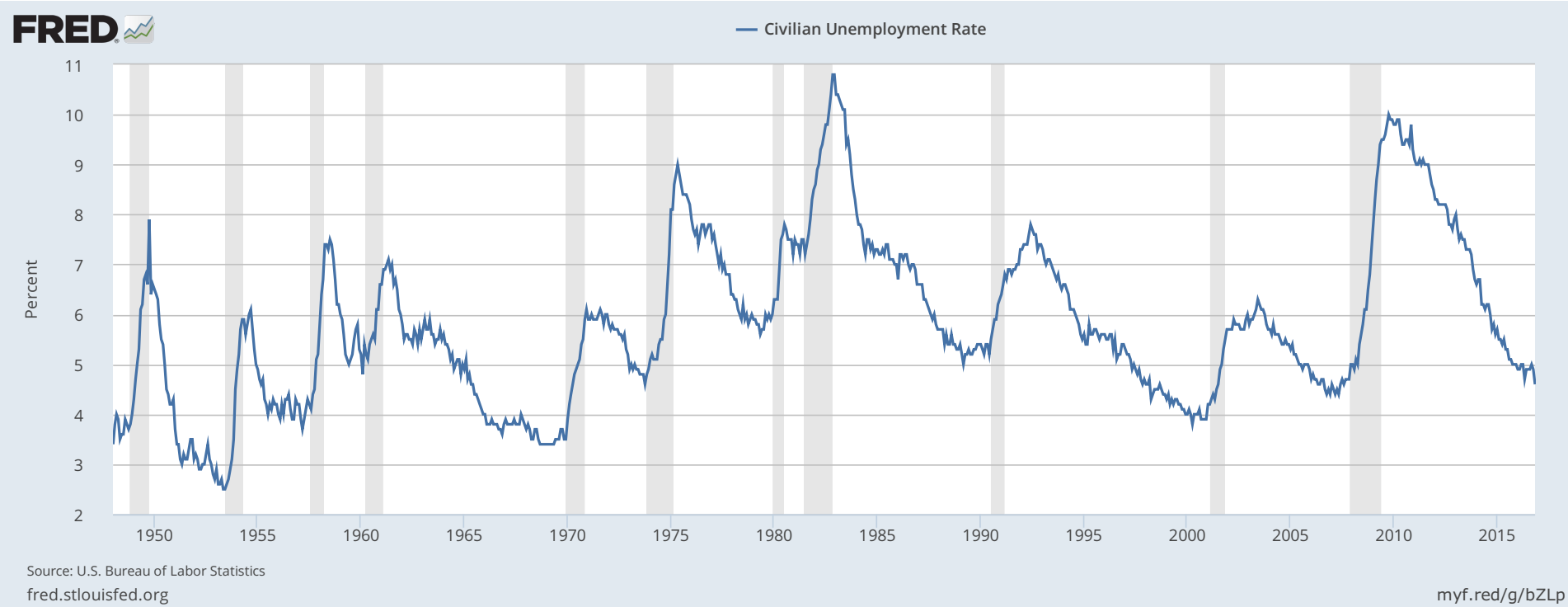
Employment

MEASURE 2: **Unemployment rate**

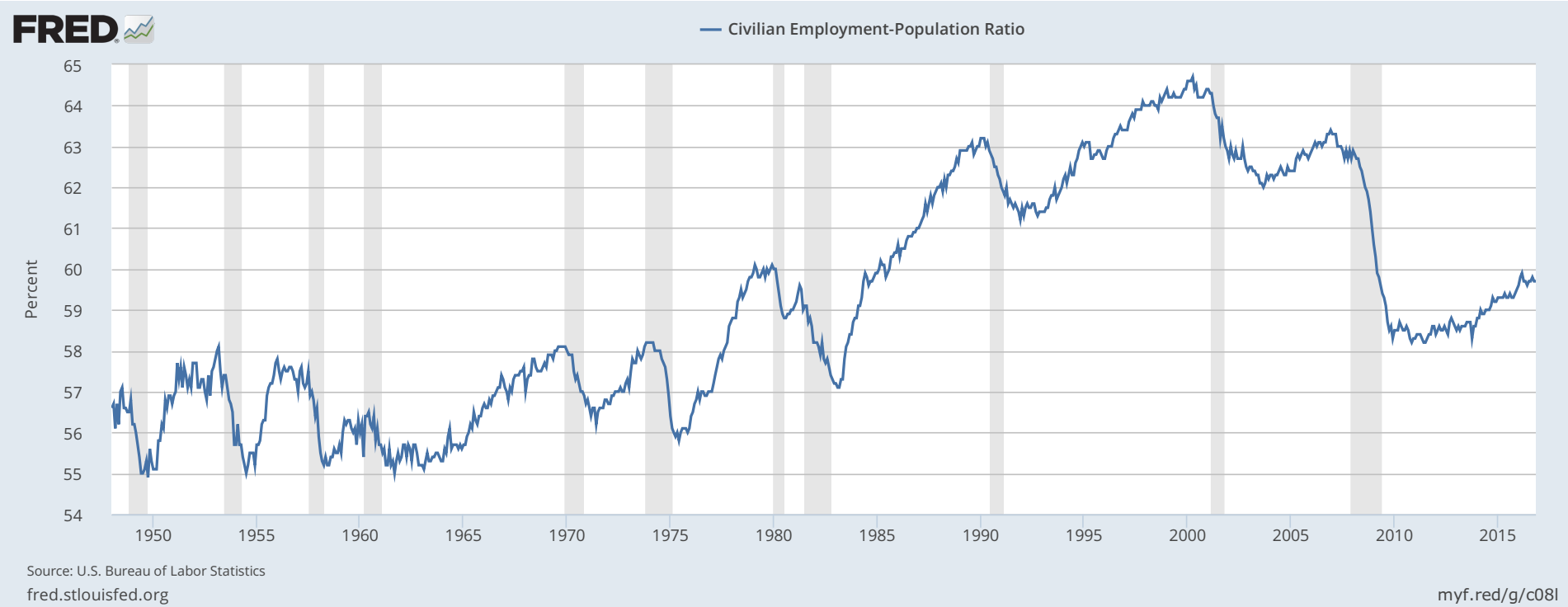
$$\frac{\#\text{ people unemployed}}{\text{Total }\#\text{ of people in labor force}}$$

- To be unemployed, you must not be working currently, and have searched for a job at least once in the last 2 weeks
- To be in the labor force, you must be employed or unemployed according to previous definition
- If you're retired, or a student, or a stay-at-home parent, or just playing video games not looking for a job, you are not unemployed.
- Usually 4-6% in U.S.; got as high as 10% during recession.

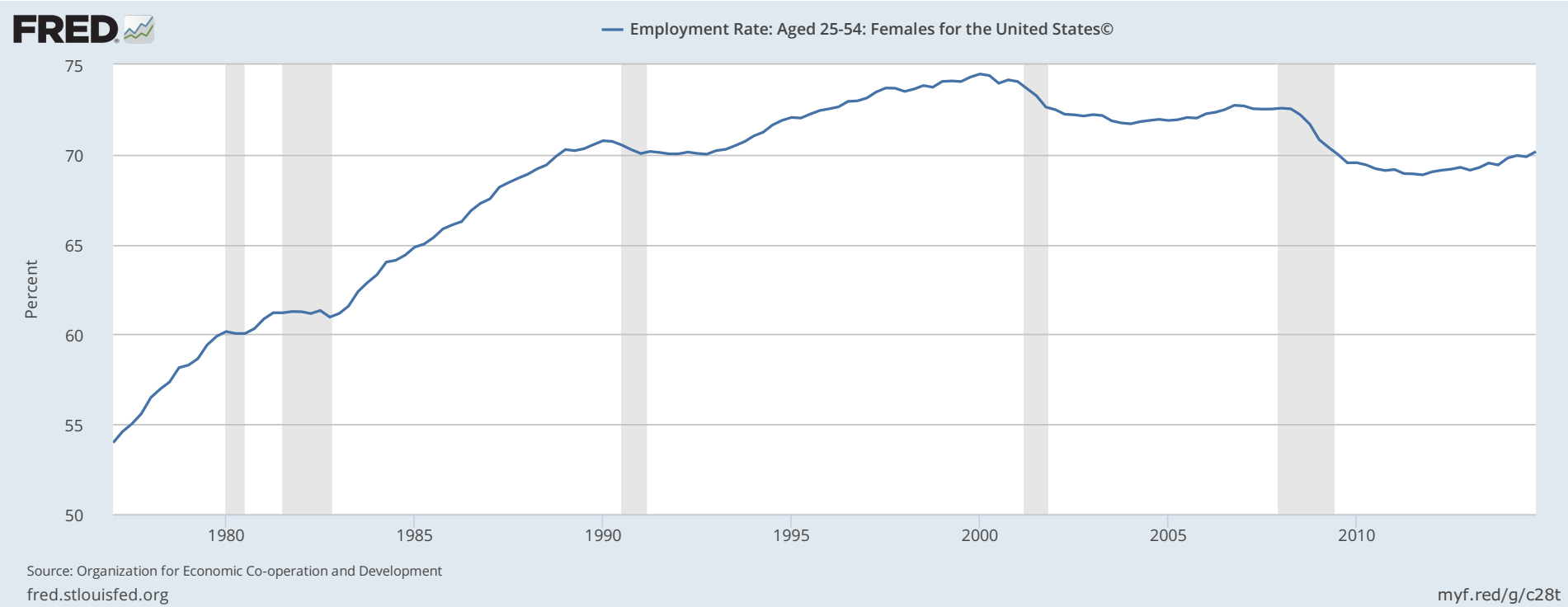
Unemployment Rate



Employment-to-population ratio



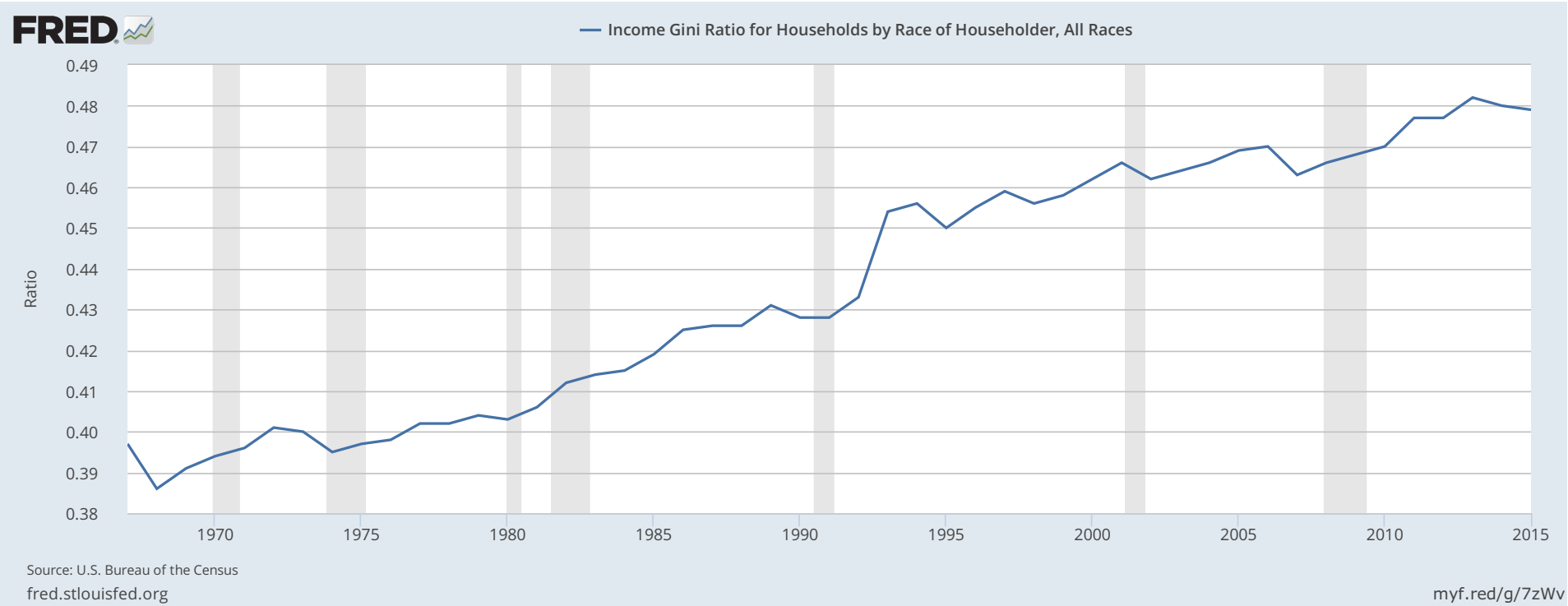
Female Employment-to-population ratio



Inequality

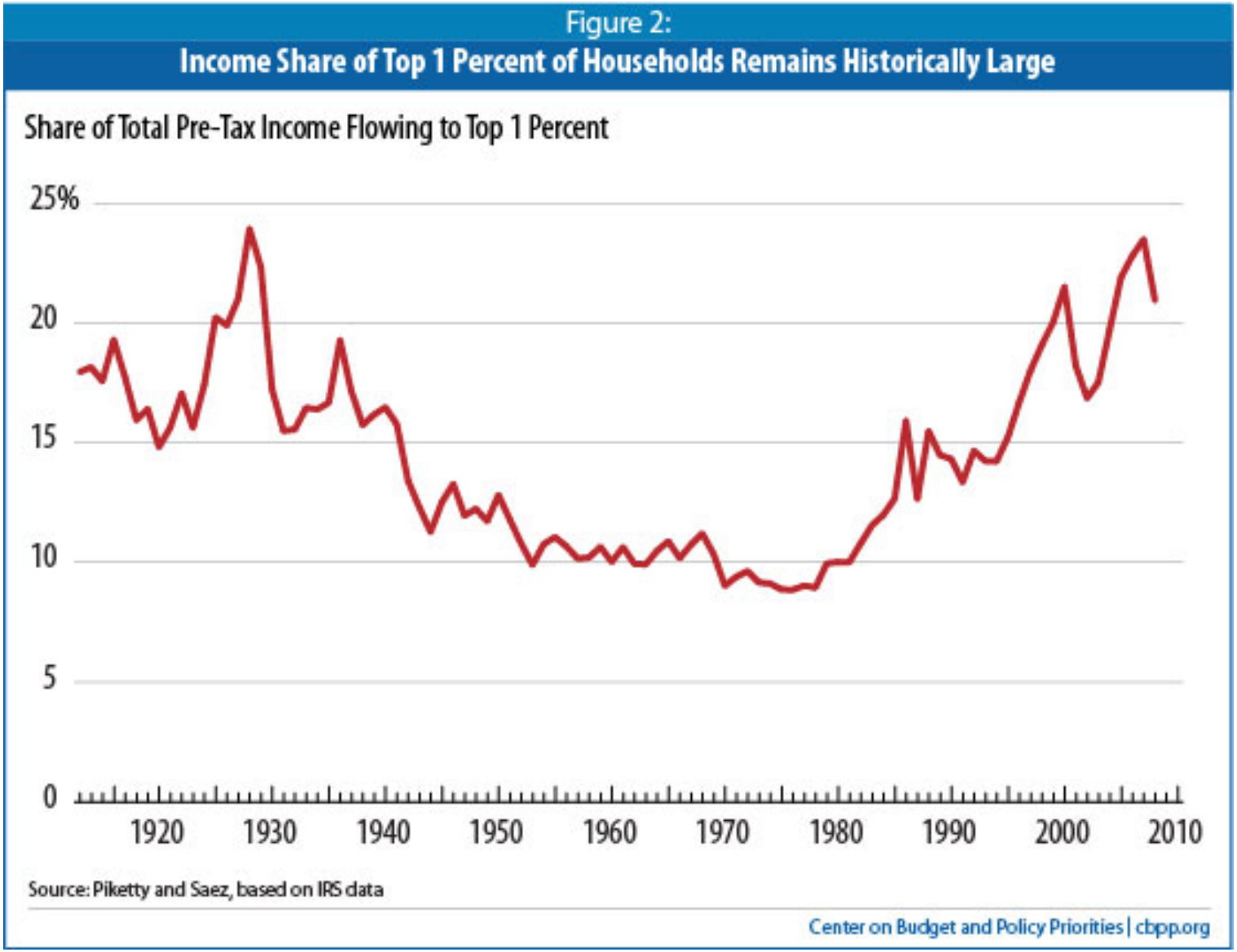
$$\text{Gini Coefficient} \qquad G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n\sum_{j} x_j}$$

Ranges from 0 (complete equality: no difference between any i and j) to 1 (complete inequality: everything owned by one person)



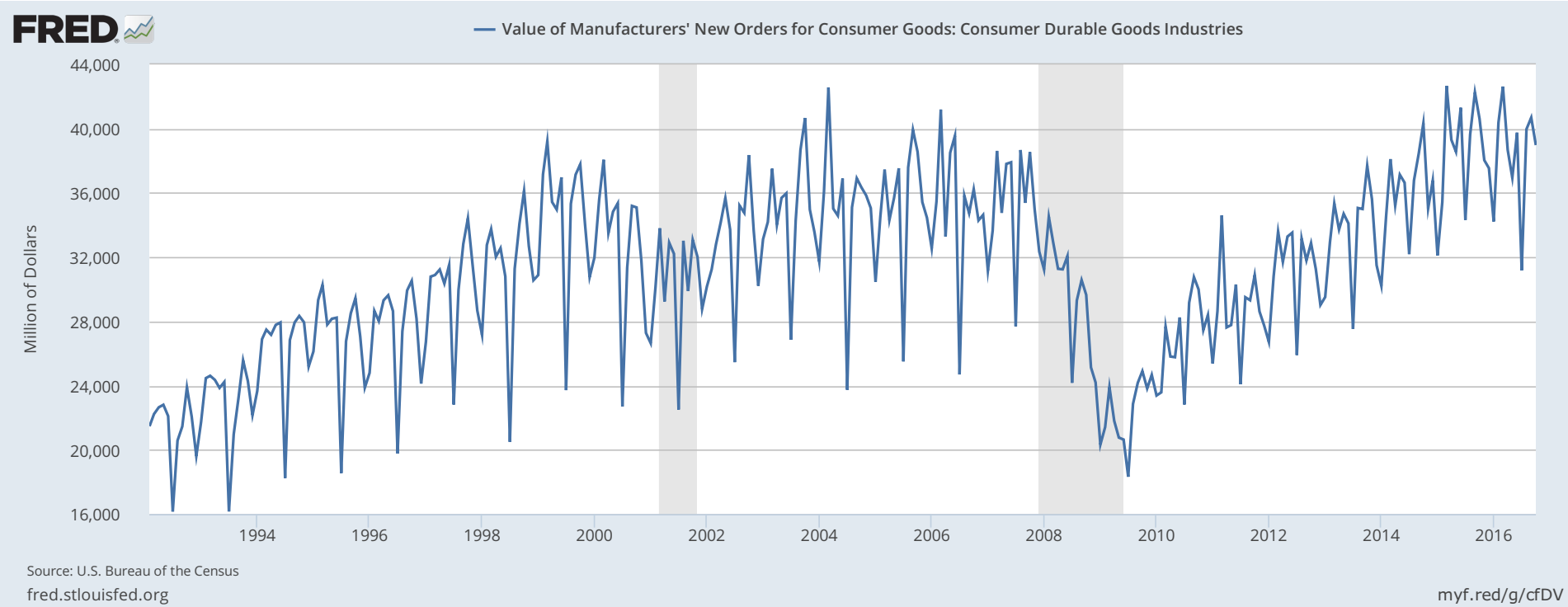
Inequality

Top 1% Share of Income



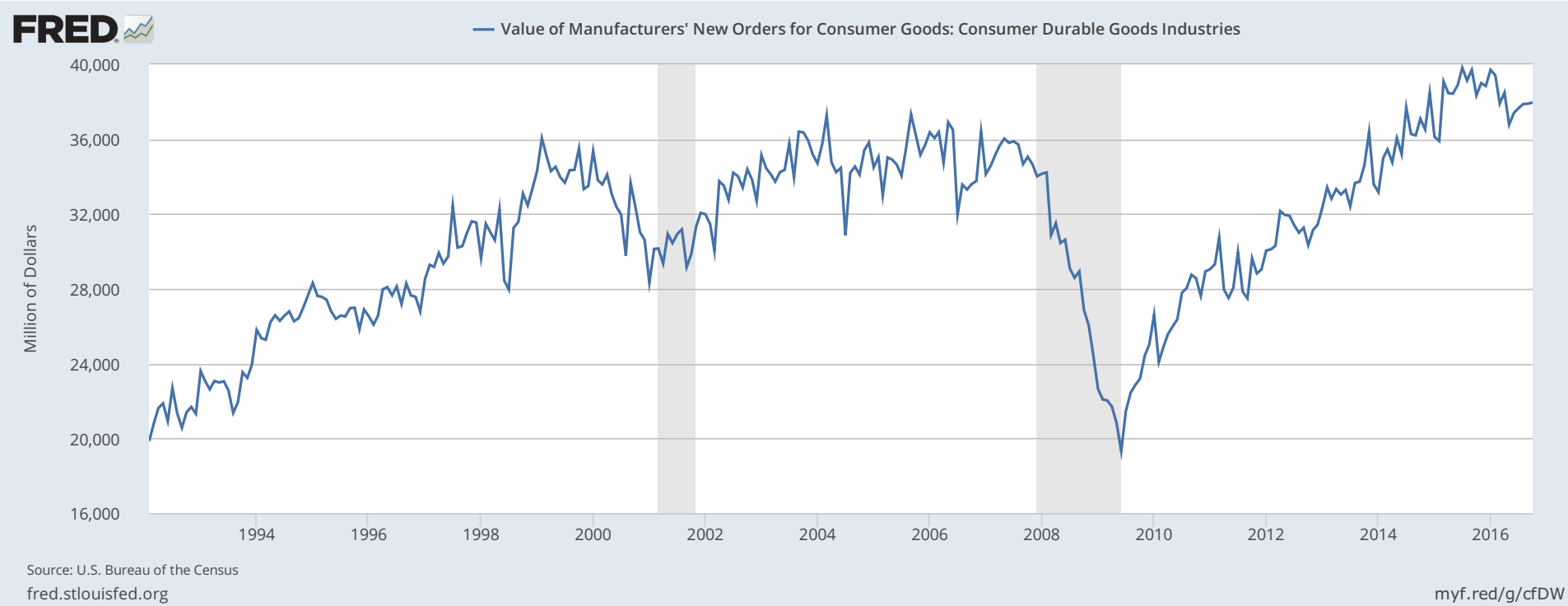
Seasonality in data

Durable Goods Orders; **Not Seasonally Adjusted**: Raw Data



Seasonality in data

Durable Goods Orders; **Seasonally Adjusted**



So what?

Seasonal fluctuations may not reflect meaningful economic conditions

- Always a spike in consumption spending around Christmas
- Spikes in tourism around summer time

Many, many ways to seasonally adjust, but most statistical bodies (BLS, BEA, Census, etc.) do it for you. Simplest way is to run a regression, for m month and t year

$$Y_{mt} = \alpha_m + \epsilon_{mt}$$

and just show the ϵ_{mt}. Essentially take out the average monthly effect.

Don't worry too much about mechanics, but be aware of seasonality.

Aggregate data sources

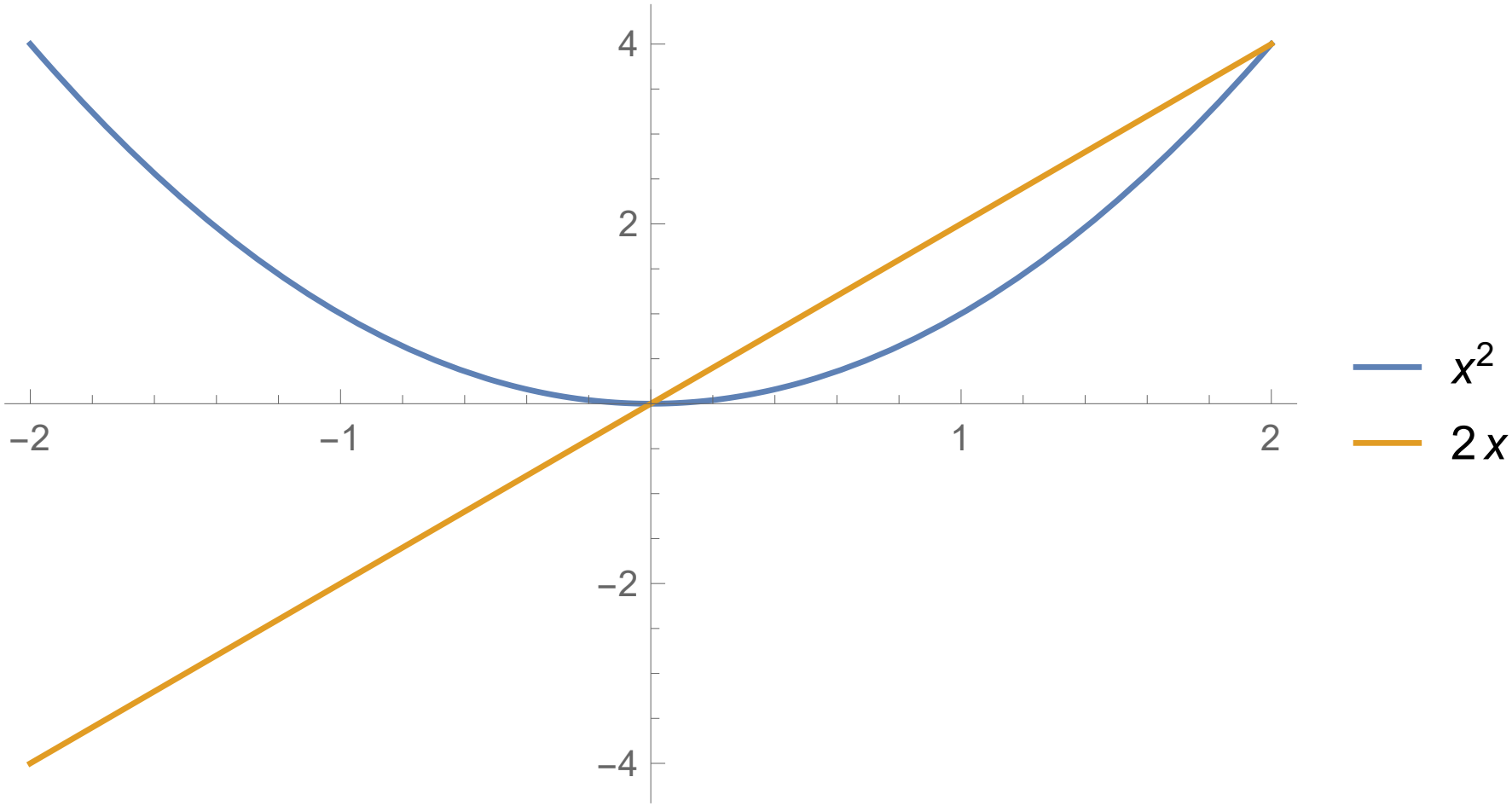
1. FRED (Federal Reserve Economic Data)
2. BLS (Bureau of Labor Statistics)
3. Census Bureau
4. BEA (Bureau of Economic Analysis)

There are also a host of publicly-available microdata sources. Will post slides detailing them.

Section 3

Optimization Review

Unconstrained Optimization: Set Derivative = 0



Optimization with Equality Constraints

$$\max_{x_1,x_2} u(x_1,x_2) \qquad \text{s.t.} \quad p_1x_1 + p_2x_2 = m$$

Two Options:

1. **Substitute constraint into problem, proceed unconstrained**
2. **Use a Lagrangian**

For now, suppose $u(x_1,x_2) = \ln x_1 + \ln x_2$.

N.B. These preferences are Cobb-Douglas because utility is invariant to monotonic transformations. Thus $u(x_1,x_2) = \ln x_1 + \ln x_2$ represent the same preferences as $u(x_1,x_2) = \frac{1}{2}\ln x_1 + \frac{1}{2}\ln x_2 = \ln\left(x_1^{1/2}x_2^{1/2}\right)$, which are the same as $u(x_1,x_2) = x_1^{1/2}x_2^{1/2}$.

Option 1: Substitute Constraint Into Optimization

Constraint implies

$$x_1 = \frac{m - p_2 x_2}{p_1}$$

Substituting in, problem becomes

$$\max_{x_2} \ln\left(\frac{m - p_2 x_2}{p_1}\right) + \ln x_2$$

Taking first order conditions:

$$-\frac{p_2}{p_1} \cdot \frac{p_1}{m - p_2 x_2^*} + \frac{1}{x_2^*} = 0$$

$$\Rightarrow -p_2 + \frac{m}{x_2^*} - p_2 = 0$$

$$\Rightarrow p_2 x_2 = \frac{m}{2} \qquad p_1 x_1 = m - \frac{m}{2} = \frac{m}{2}$$

Option 2: Use a Lagrangian

Write Lagrangian

$$\mathscr{L} = \ln x_1 + \ln x_2 + \lambda (m - p_1 x_1 + p_2 x_2)$$

Take first order conditions with respect to x_1, x_2, λ

$$[x_1] : \frac{1}{x_1} - \lambda p_1 = 0$$
$$[x_2] : \frac{1}{x_2} - \lambda p_2 = 0$$
$$[\lambda] : m - p_1 x_1 + p_2 x_2 = 0$$

Optimization with Inequality Constraints

$$\max_{x_1,x_2} u(x_1, x_2) \qquad \text{s.t.} p_1 x_1 + p_2 x_2 = m$$

$$x_1 \geq 0, \quad x_2 \geq 0$$

Four places to check:

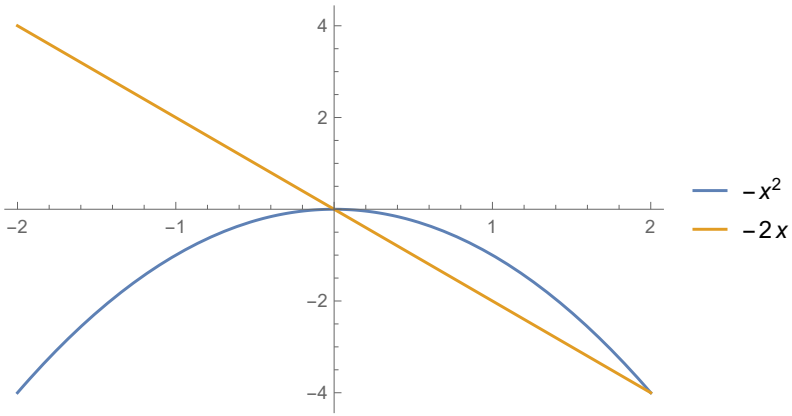
1. Internal solution (Inequality constraints do not bind)
   - Solve optimization as above, ignoring inequality constraint
2. $x_1 = 0, x_2 = 0$ (Both ineq. constraints bind)
   - Solve optimization as above, ignoring inequality constraint

   plug into objective function and check
3. $x_1 = 0, x_2 > 0$ (Only x_1 constraint binds)
4. $x_1 > 0, x_2 = 0$ (Only x_2 constraint binds)
   - Use Kuhn-Tucker conditions, solving problem with Lagrangian

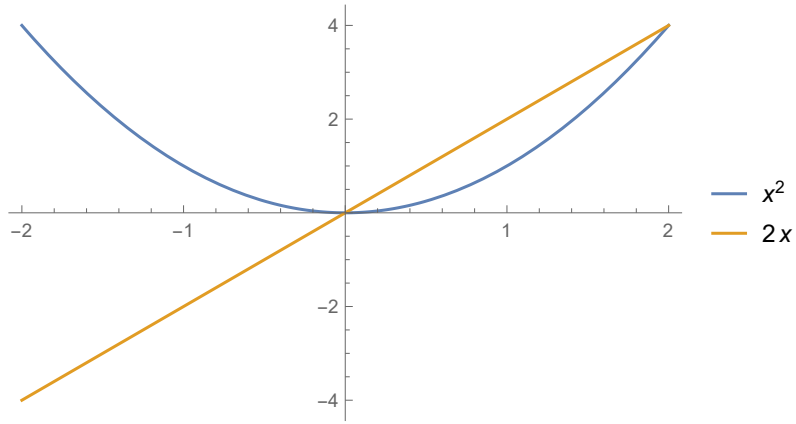
Second Order Conditions

- The first order conditions above guarantee that we've found a critical point. But is it a max/min?
- Check second order conditions!

Think of second order conditions as checking the second derivative. If $f''(x) > 0$, then min, if $f''(x) < 0$, then max. In constrained case, check the definiteness of the Bordered Hessian matrix (**Not in this class**).



SOC = -2 $\Rightarrow$ MAX



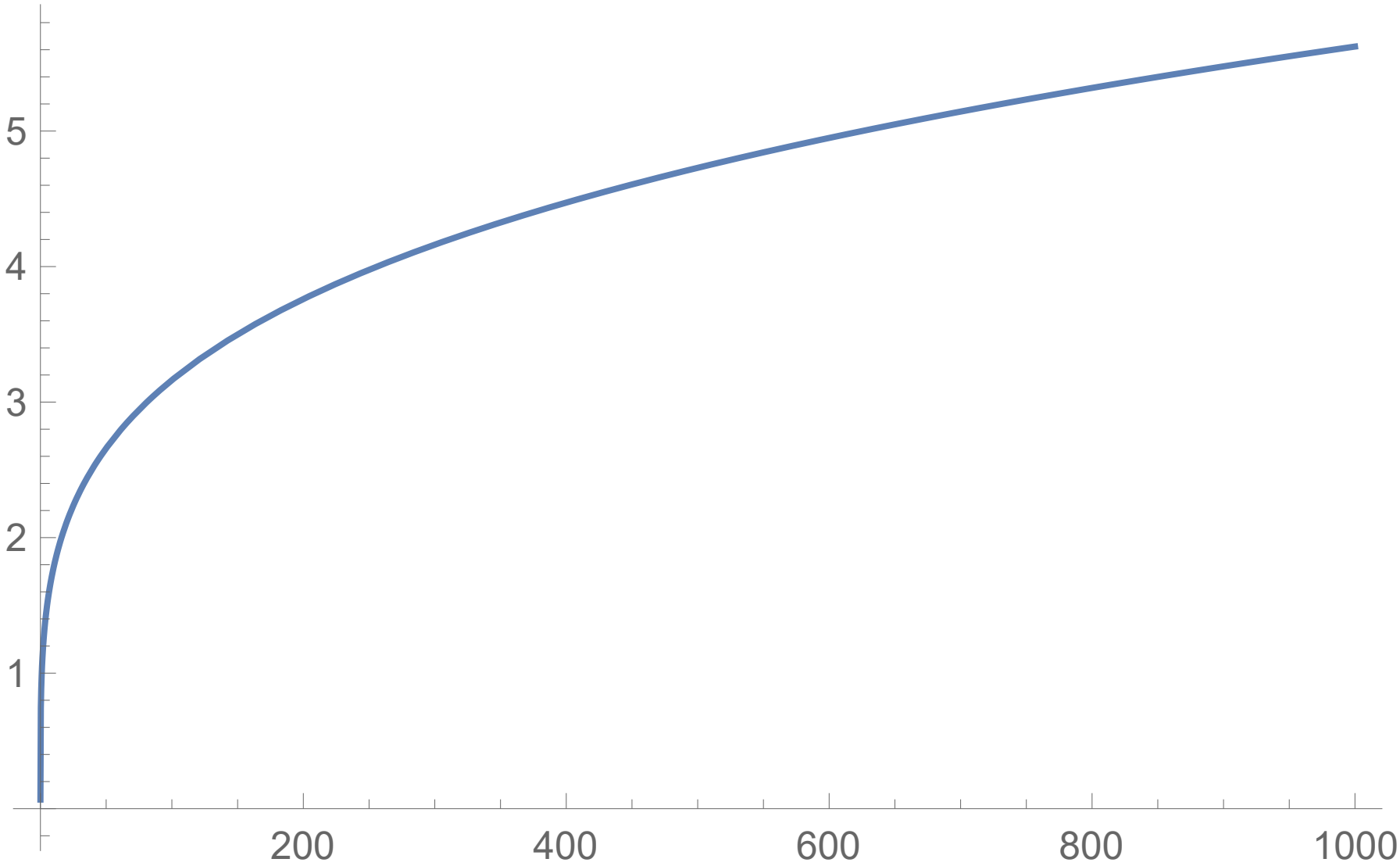
SOC = +2 $\Rightarrow$ MIN

Inada Conditions (e.g. Cobb-Douglas satisfies them)

In macro, we often make assumptions on preferences/production:

1. $f(0) = 0$ (only relevant for production function)
   - No inputs $\Rightarrow$ no output
2. $f(x)$ is continuously differentiable
   - Can take FOC without any problem
3. $f(x)$ is strictly increasing in all of its arguments ($f'(x) > 0 \forall x$).
   - Budget constraint doesn't bind; always want more inputs
4. $f''(x) < 0$ so function is concave
   - Don't have to worry about SOC: every critical point is max
5. $\lim_{x \to 0} f'(x) = +\infty$: Limit of first derivative is infinity at 0
   - First unit is infinitely good: never at corner solution with $x = 0$
6. $\lim_{x \to \infty} f'(x) = 0$: Limit of first derivative is 0 at infinity
   - Eventually run into diminishing returns; guarantees existence

An example of a function satisfying Inada conditions:



An example of a function satisfying Inada conditions:

$$f(x) = x^{0.25}$$

1. $f(0) = 0^{0.25} = 0$
2. $f(x)$ polynomial so continuously differentiable
3. $f'(x) = \frac{0.25}{x^{0.75}} > 0 \quad \forall x > 0$
4. $f''(x) = -0.25 \cdot 0.75 \frac{1}{x^{1.75}} < 0 \quad \forall x > 0$
5. $f'(x) \xrightarrow{x \to 0} \frac{0.25}{0} = \infty$ (abuse of notation)
6. $f'(x) \xrightarrow{x \to \infty} \frac{0.25}{\infty} = 0$

Section 4

Appendix

Gathering Data

- Look for data everywhere; don't be afraid to ask for it!
- 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)
- Description of various micro data sources in following slides. **Far from exhaustive list**

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): Represetative 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

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

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
- Publically-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.
- MEPS (Medical Expenditure Panel Survey) - Information on individual medical expenditure and insurance