The Future of Econometrics with Big Data: Modeling Deeper into Sources of Economic Shocks and Volatility

Robert J. Shiller

Yale University

Tenth Annual Society for Financial Econometrics Conference

New York University

June 21, 2017

# The Term "Big Data" First Caught on around 2010, from Proquest News and Newspapers

1987 - 2017 (years)

Survey Data Outside of Government Tends Often to Be Episodic, Not Time Series

- One time only surveys are reflections of concerns at a point of time
- Consumer sentiment and consumer confidence indices are based on questions that were written a half century or more ago
- One of the Michigan Consumer Sentiment survey questions is a clear reference to the Great Depression

"Stock prices in the United States, when compared with measures of true fundamental value or sensible investment value, are too high, too low, or about right?" (Yale, Shiller, surveys)

Stock Market Confidence Indexes (c) 2001 - 2017
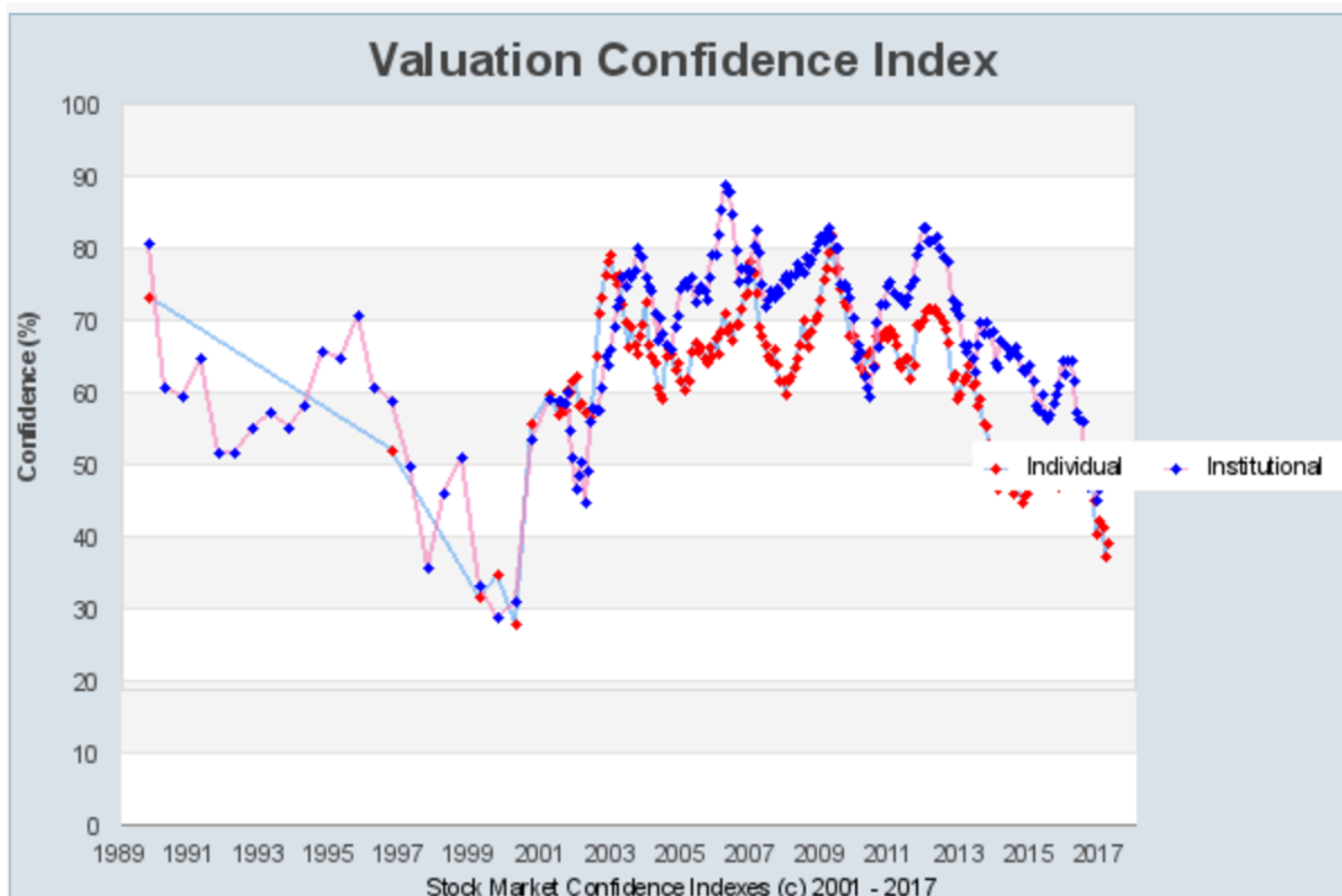
Center for Research in Securities Prices

The center is best known for its initial report on rates of return on investments in all common stocks on the New York Stock Exchange from 1926 through 1960. The $200,000 study, which took 3½ years and millions of computations, was well publicized by the press and by Merrill Lynch, Pierce, Fenner & Smith, which had both suggested the research and given about half the money for it.

Computer-Oriented Study Traces Movements, Yields for 1926-60

By Ed Morse

NEW YORK (P) — During a 35-year period all stocks on the New York Stock Exchange had a combined rate of return of 9.01% before taxes, according to a computerized study made public Sunday by the University of Chicago.

The study said the return on stock was greater than that for bank savings, mortgages and municipal, government and corporate bonds.

The 3½-year study produced "the most comprehensive research to date of the movement of stocks on the New York Stock Exchange," said Dean George P. Shultz of the university's graduate school of business.

The stretch from 1926 to 1960 included periods of depression and recession, war booms and postwar inflation, as well as the Great Crash of 1929.

In a nutshell, the first results of the study were as follows:

Over the 35-year period, the rates of return, compounded annually, on all stocks listed on the "Big Board" were, for example,

9.01% for tax-exempt institutions, 8.2% for taxpayers in the $10,000 income class and 6.8% for those in the $50,000 income bracket.

The rate of return included dividends and other corporate payouts, the change in market value of the stocks, the cost of brokers' commissions and taxes paid. Dividends were reinvested.

Some Comparisons

Comparable after-tax information in other forms of investment was not available, but some before-tax information was found, as follows:

CRSP Tape 1960s

- First serious effort to document returns

Vast Stock Analysis a Tale of Many Woes

BY WILLIAM CLARK
[Financial Editor]

(Second of Two Articles)

The center for research in security prices at the University of Chicago graduate school of business is well along in the first phase of a project designed to produce more information on stock price fluctuations than has heretofore been available. The project, with which Merrill Lynch, Pierce, Fenner & Smith, Inc., is linked as sponsor, has had its rough moments—by the thousands.

Lawrence Fisher, assistant professor of finance at the school, was placed in direct charge of operations at the center when it was established a couple of years ago. Almost immediately he ran into troubles.

Data was to be collected on all common stocks listed on the New York Stock exchange beginning in January, 1926, at intervals of one month. The exchange was asked to supply a copy of stock listings at the end of each month during the period. To the astonishment of Fisher and his associates, no such lists existed.

Takes Weeks of Effort

There followed weeks of clerical effort to get the information from the records of the exchange, record it on forms designed by Fisher, and transfer it to punch cards which later would be used to place it on tape. The clerical staff

dend or two in whisky warehouse receipts. Some companies have paid dividends in stock of other companies, as Standard Oil company [Indiana] which pays an occasional supplementary dividend in stock of Standard Oil company [New Jersey].

Dividends Are Classified

Moreoever, each dividend was classified as to tax status. It might be capital gain, return of capital, ordinary income, or one of three or four others. And one dividend might fall into as many as three different categories.

"This kind of detail had never been recorded before," said Lorie, who was telling the story of the whole project. "It's all Fisher's fault. He insisted on it."

The recording of capital changes, without going into the gory details, was an equally back-breaking labor.

To figure what the investor earned on his stock, the cost of commissions also had to be taken into account. It was necessary to compute earnings if [1] dividends were reinvested in additional stock, or [2] dividends payments simply were retained as cash. Allowances had to be made for payment of taxes appropriate to the tax status of the investor, and for keeping or not keeping stock made available by mergers or

Meaning of Words Changes Through Time

- Dipsomania vs alcoholism

- Contagious ideas (Hume 1742), extraordinary popular delusions (Mackay 1841) idea microbes (Le Bon 1895), memes (Dawkins, 1975), thought viruses

- What is a stock?

- What is a bond?

- What is a bank?

Semantic Search

- Wikipedia definition: "*Semantic search* seeks to improve *search* accuracy by understanding the searcher's intent and the contextual meaning of terms as they appear in the searchable dataspace, whether on the Web or within a closed system, to generate more relevant results."

- Latent Dirichlet Allocation

U.S. Price Level (CPI) since 1927
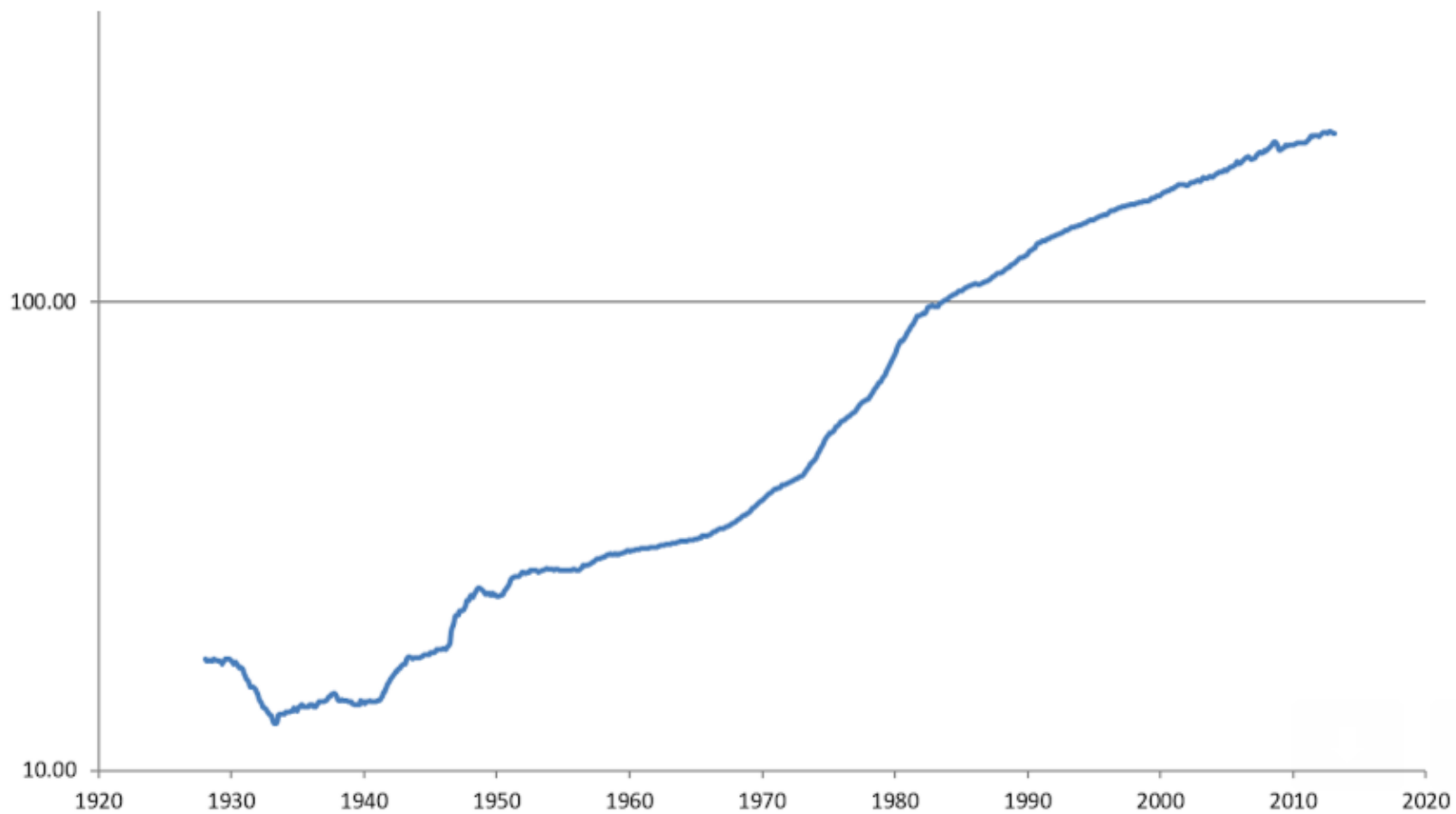
Origins of the "Out of Control" Inflation 1960-80 as Discovered with Google Ngrams
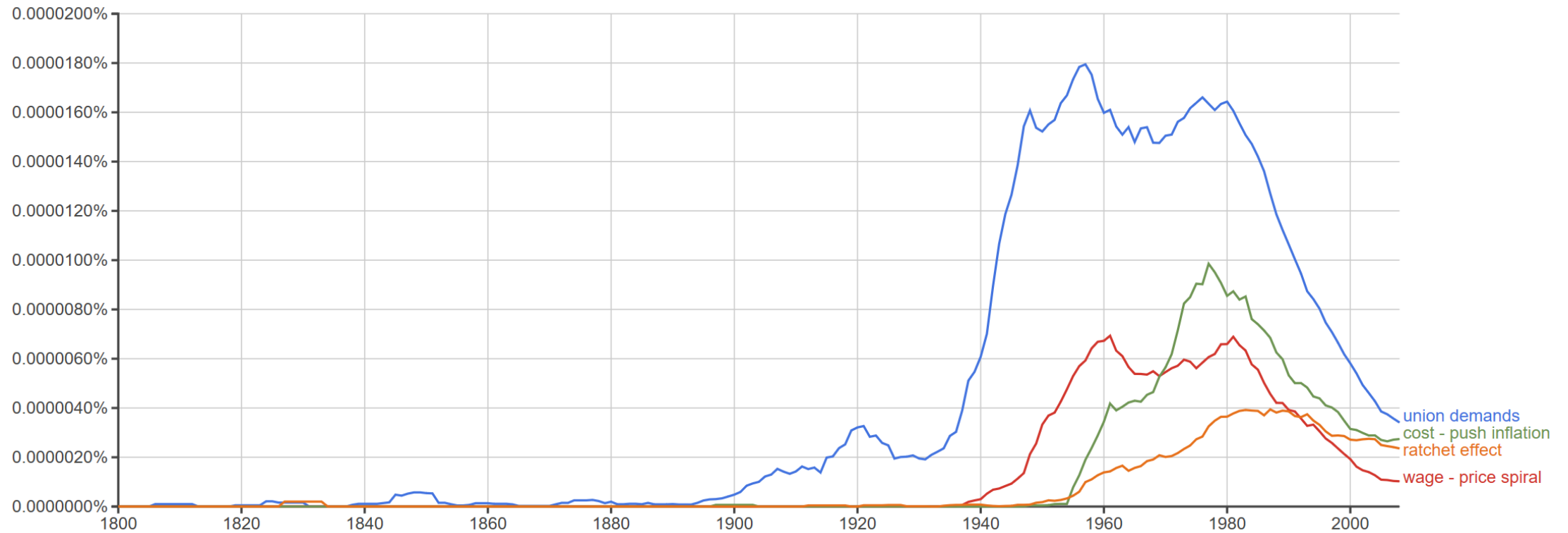
Google Books Ngram Viewer

Graph these comma-separated phrases: | labor union * | 🔽 | ☐ case-insens

between | 1800 | and | 2008 | from the corpus | English ▼ | with smoothing of | 3 ▼ | . | **Search lots of bo**

labor union and
labor union leaders
labor union in
labor union members
labor union movement
labor union or
labor union is
labor union to
labor union was
labor union officials
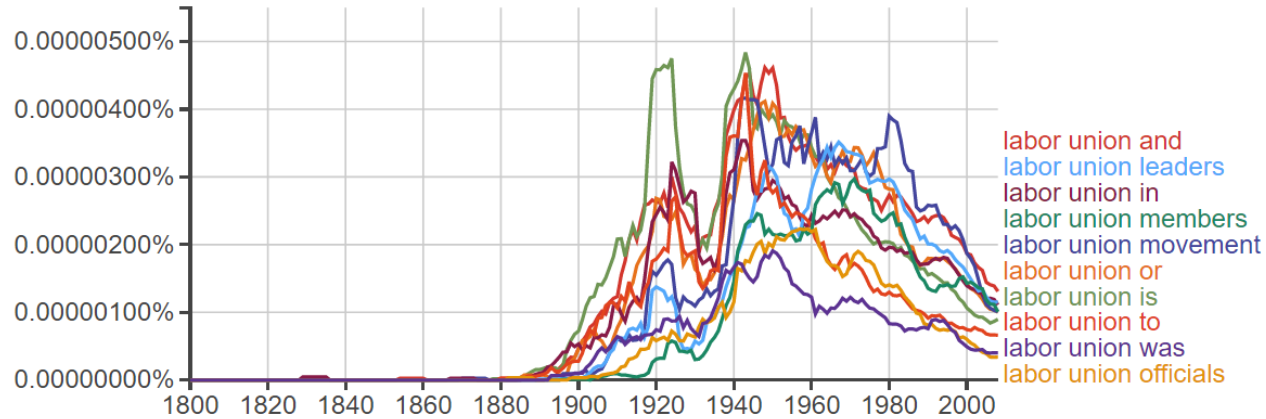
(click on line/label for focus, right click to expand/contract wildcards)

Search in Google Books:

1800 - 1918	1919 - 1946	1947 - 1952	1953 - 1990	1991 - 2008	labor union and
1800 - 1914	1915 - 1940	1941 - 1946	1947 - 1985	1986 - 2008	labor union is
1800 - 1920	1921 - 1943	1944 - 1949	1950 - 1991	1992 - 2008	labor union or
1800 - 1929	1930 - 1978	1979 - 1984	1985 - 1994	1995 - 2008	labor union movemen
1800 - 1919	1920 - 1939	1940 - 1945	1946 - 1994	1995 - 2008	labor union in
1800 - 1929	1930 - 1966	1967 - 1972	1973 - 1996	1997 - 2008	labor union leaders
1800 - 1916	1917 - 1947	1948 - 1953	1954 - 1985	1986 - 2008	labor union to
1800 - 1941	1942 - 1967	1968 - 1972	1973 - 1996	1997 - 2008	labor union members
1800 - 1930	1931 - 1953	1954 - 1958	1959 - 1989	1990 - 2008	labor union officials
1800 - 1919	1920 - 1945	1946 - 1951	1952 - 1989	1990 - 2008	labor union was

Focus Groups an Continuing Trend since 1980s but Not in Economics or Finance
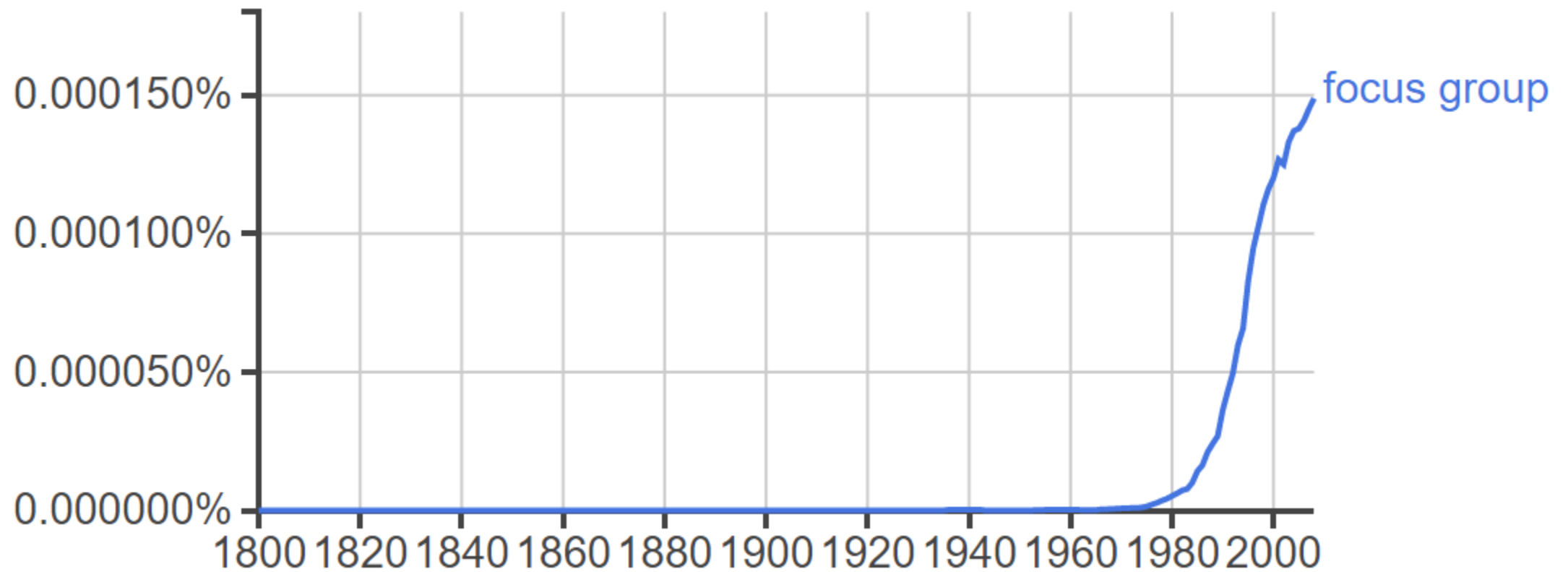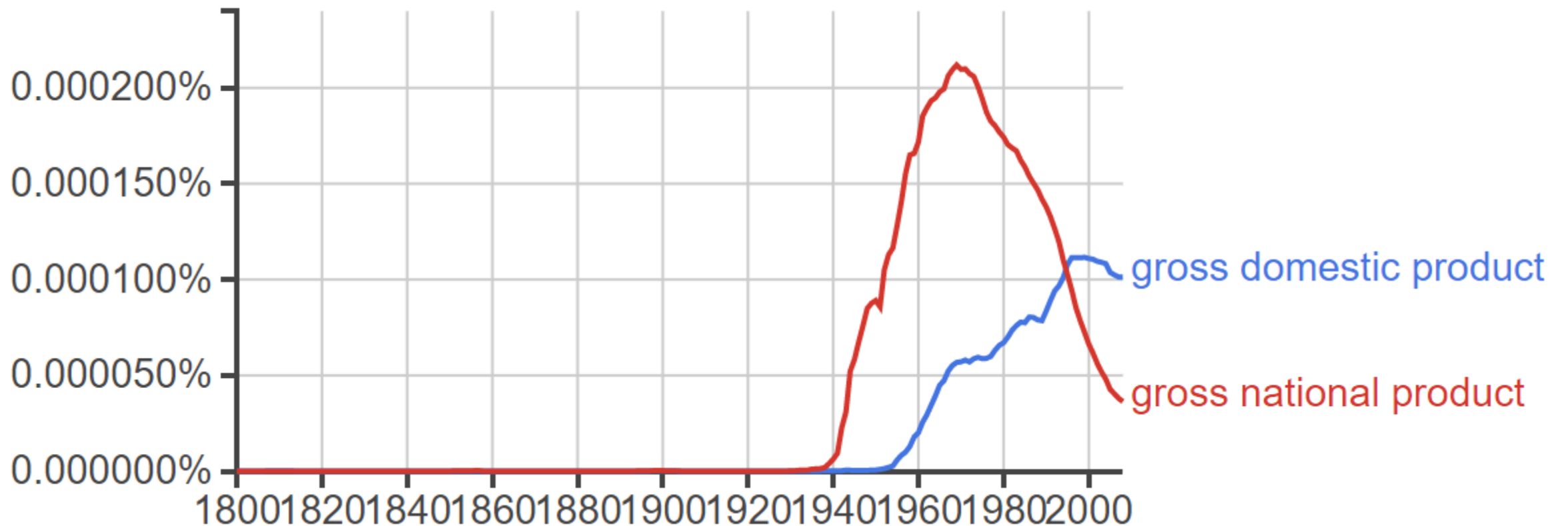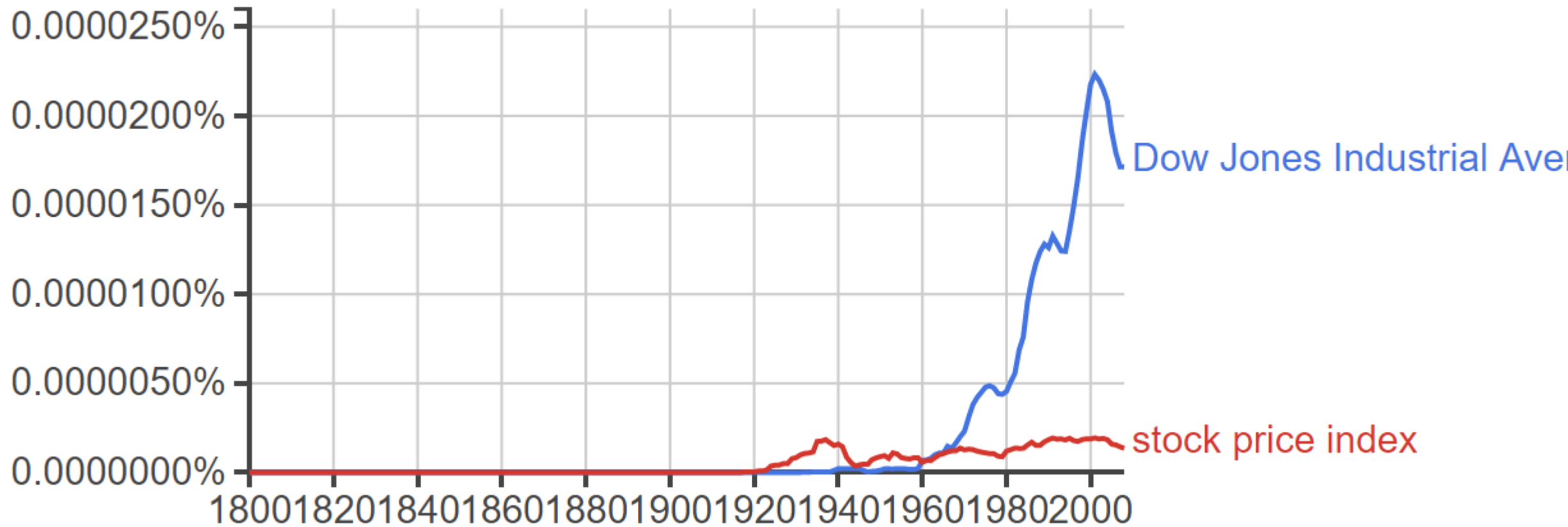
Econometrics Is a Relatively Young Science
Google Ngrams
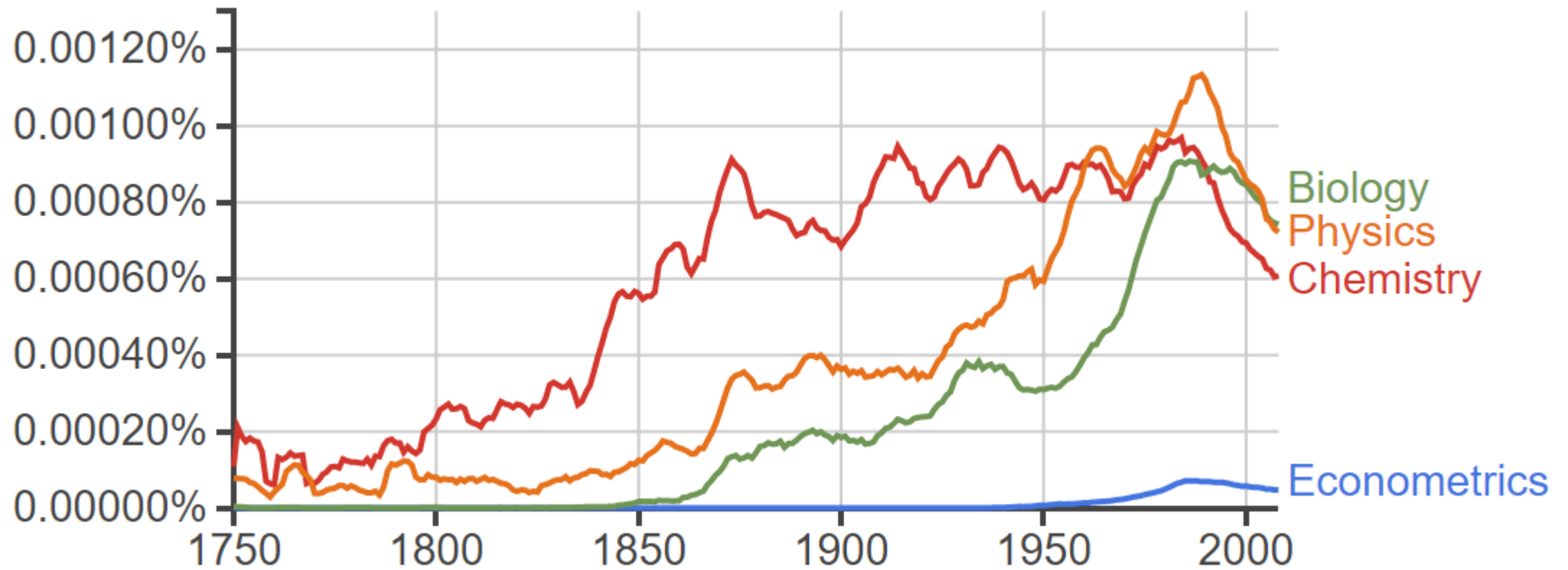
Rise of Econometrics Coincides with Rise of National Income and Product Accounts, after Great Depression, Google Ngrams

Rise of Econometrics Coincides with Rise of Financial Time Series Data, Google Ngrams

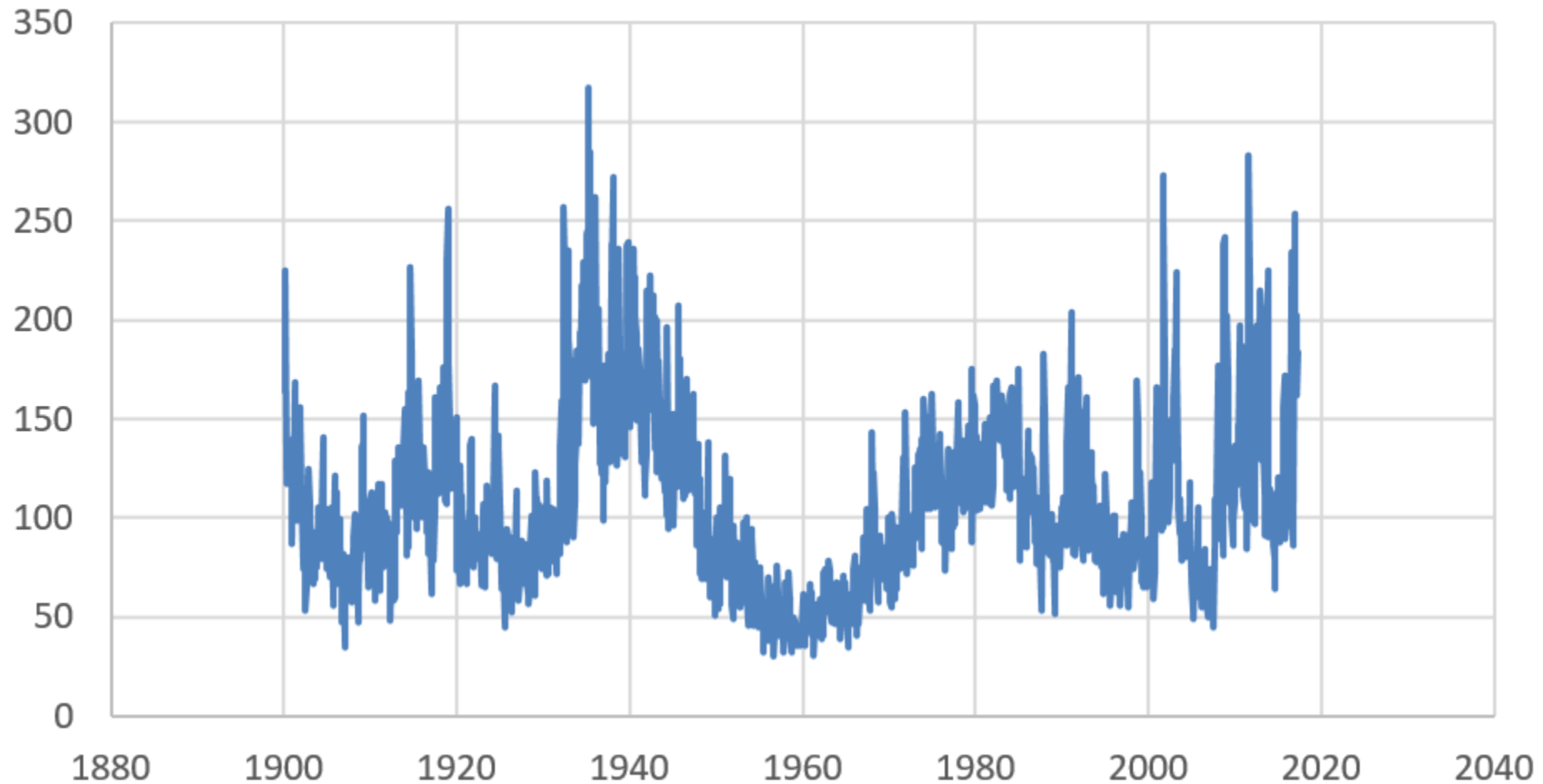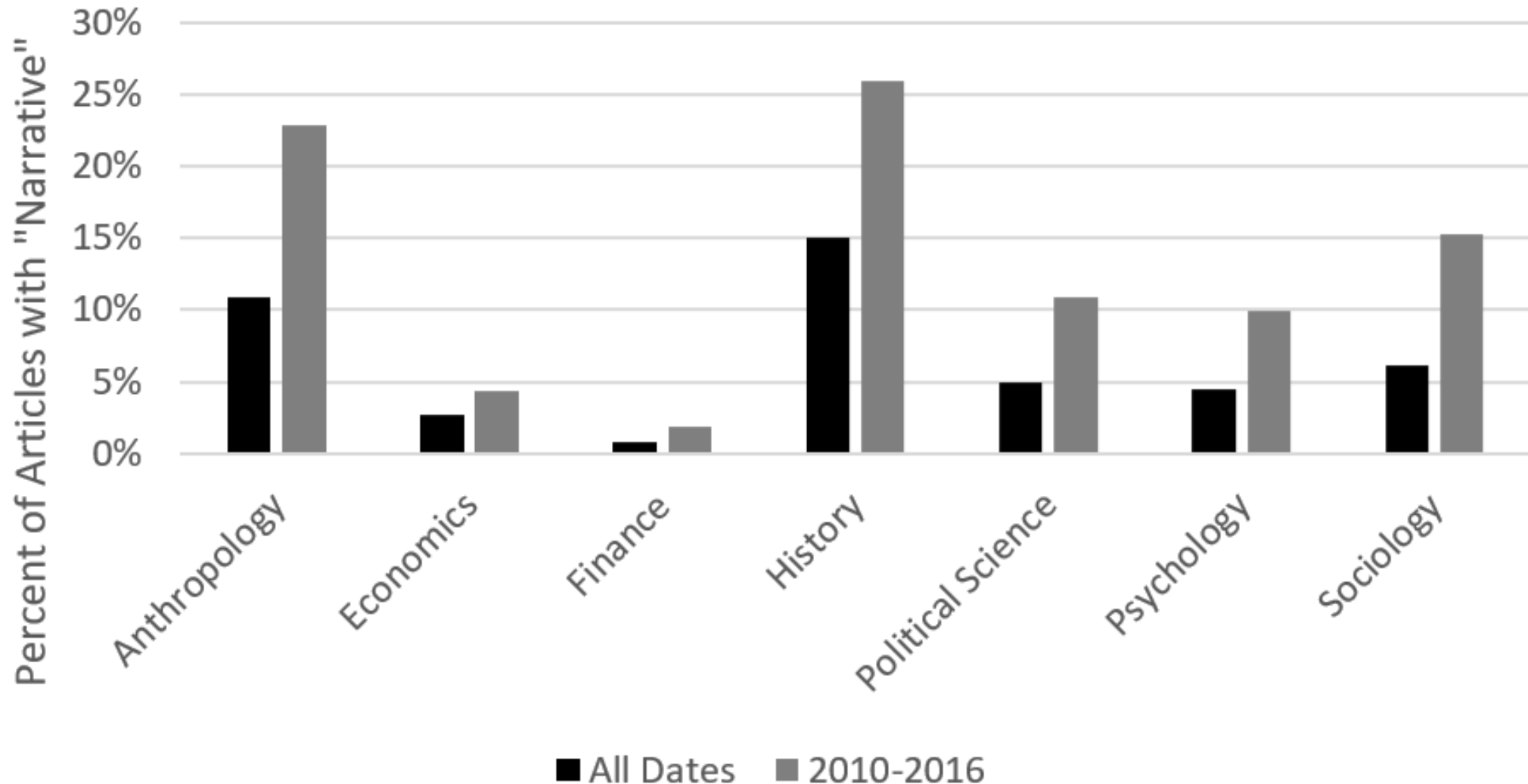# Baker Bloom & Davis News-Based Economic Policy Uncertainty Index Jan 1900-April 2017

Figure 1: JSTOR Counts of Word "Narrative" as Percent of All Articles, by Discipline

Kermack-McKendrick SIR Disease Epidemic Model 1927

- S=fraction of population susceptible, I=fraction of population infected and now contagious, R=fraction of population recovered and now immune, S+I+R=N, *c*=contagion rate. *r*=infection rate

$$\frac{dS}{dt} = -cSI$$
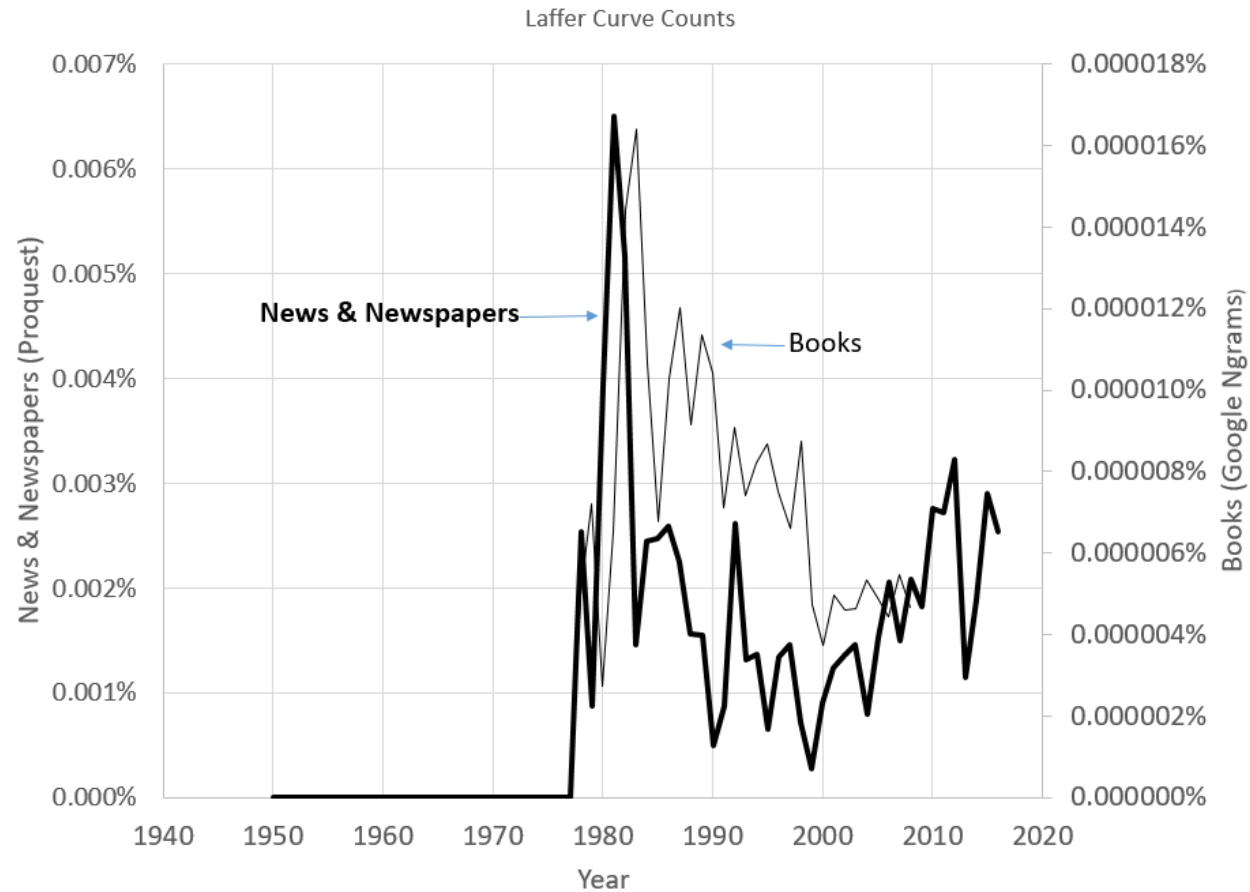
$$\frac{dI}{dt} = cSI - rI$$

$$\frac{dR}{dt} = rI$$

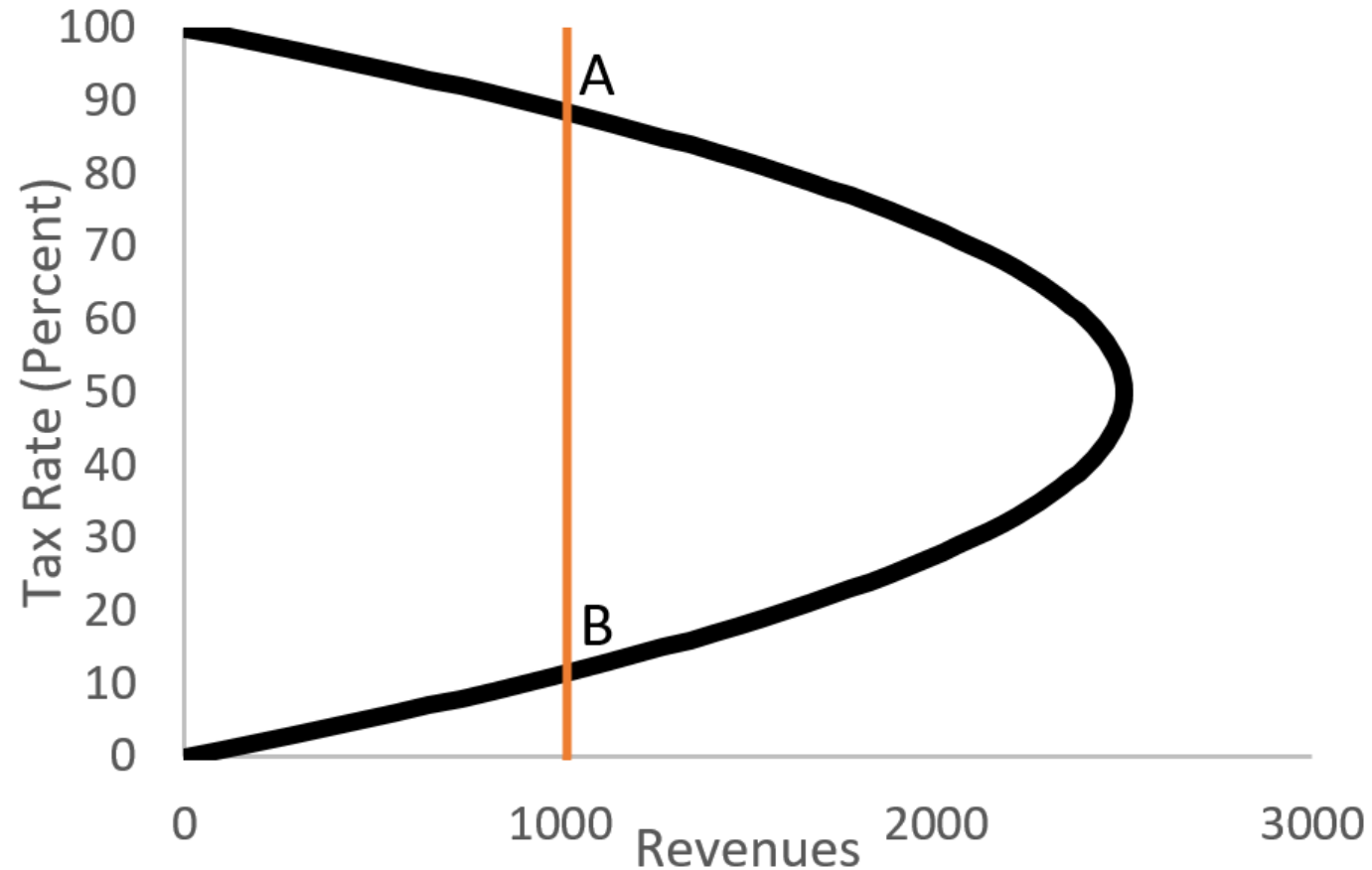# Figure 2: Time Paths of S, I, and R in Kermack-McKendrick Model
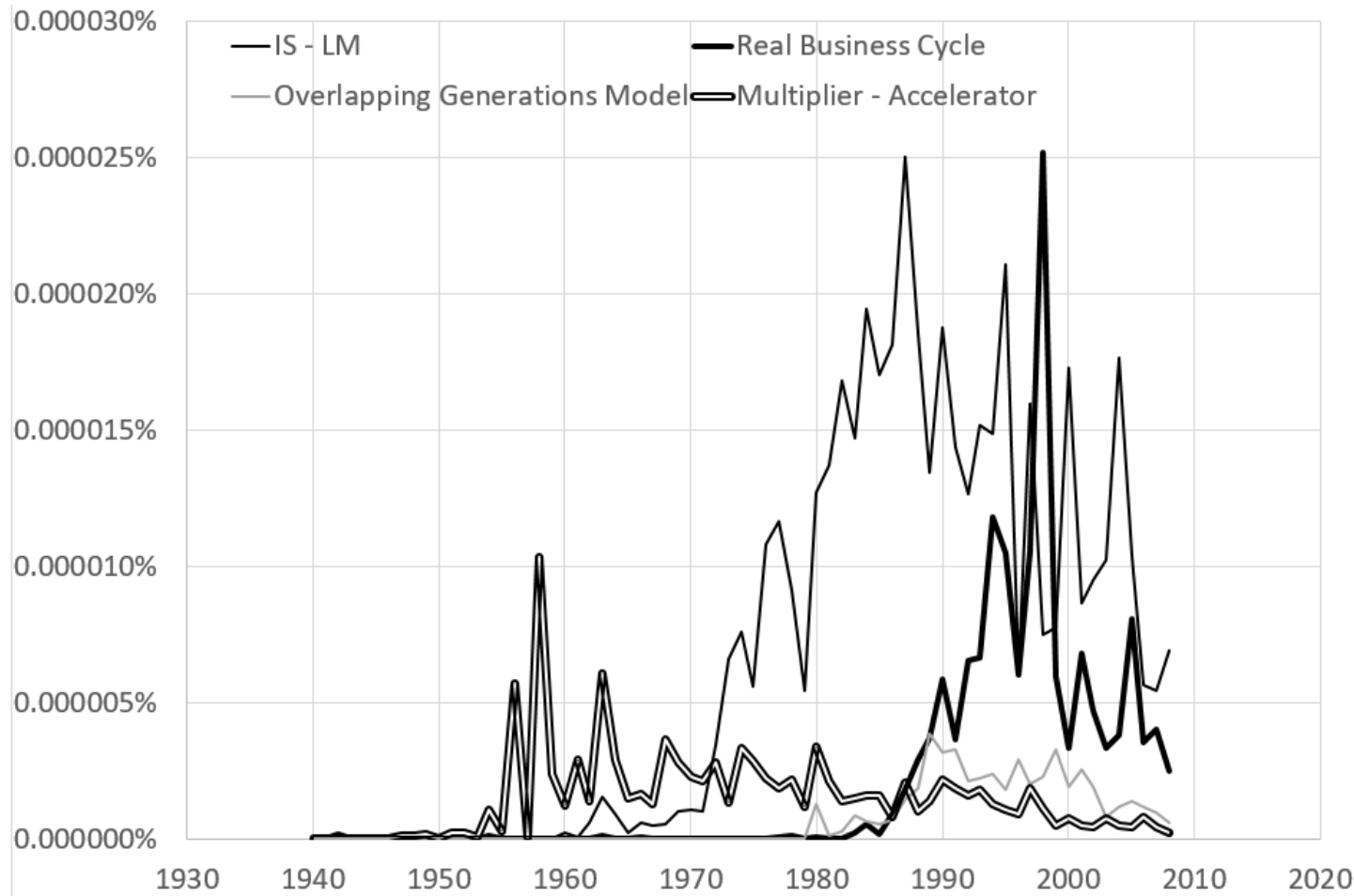## $N=100$, $I(0)=1$, $c=0.005$, $r=0.05$
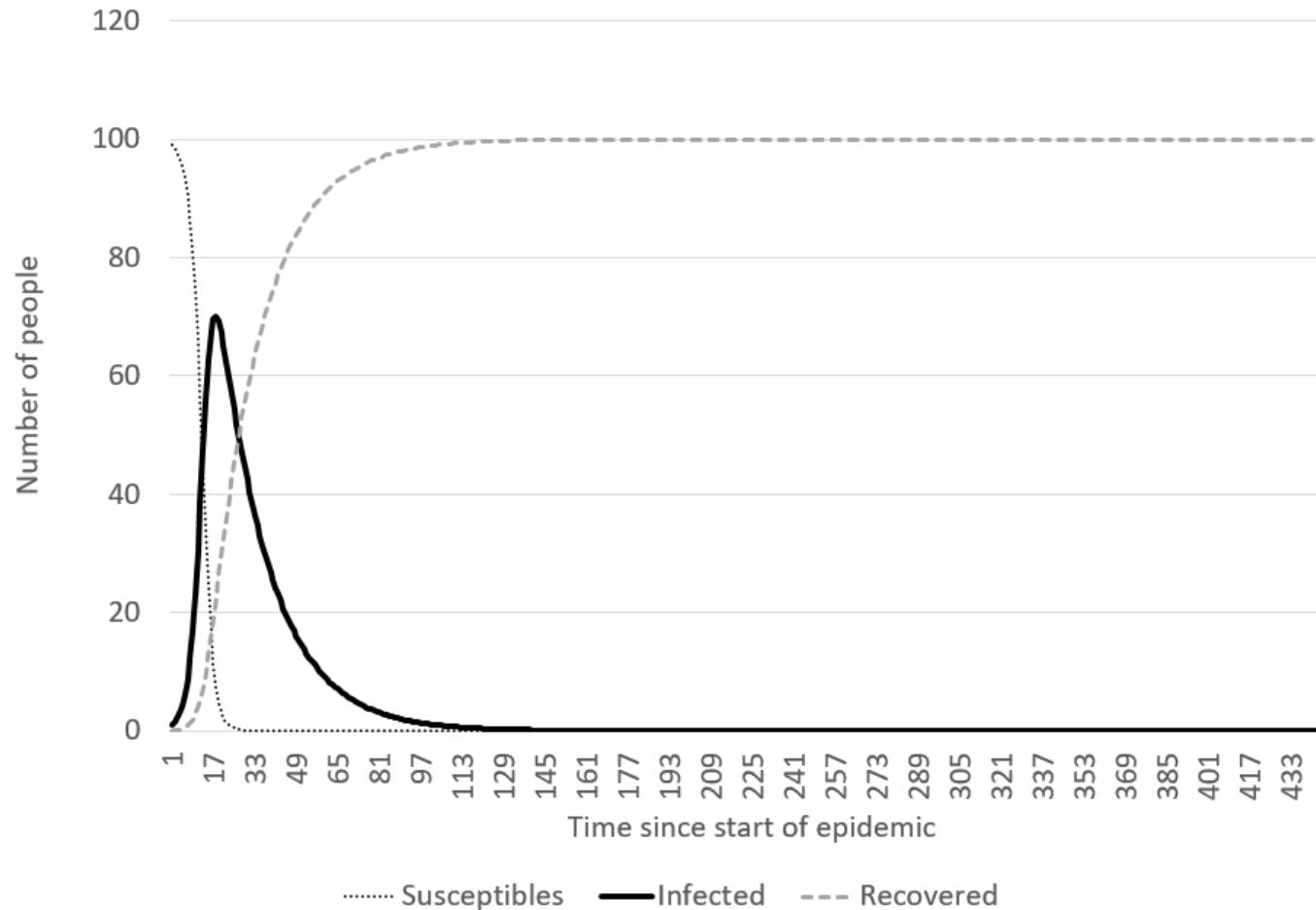
Laffer Curve Counts in New & Newspapers 1950-2016 & Books 1950-2008

The Laffer Curve, Arthur Laffer via Jude Wanniski, *The Way the World Works*, 1978

Google Ngrams (Books) Counts for Some Major Macroeconomic Models 1940-2008

Size of Epidemic Determined by *c/r*

- *dS/dR=-(c/r)S*

- $S = (N - I_0)e^{-\left(\frac{c}{r}\right)R}$

- $I_\infty = 0$

- $\frac{c}{r} = R_\infty^{-1}\log\frac{N-I_0}{N-R_\infty}$

- Size of epidemic depends only on *ratio* of contagion rate to removal rate

- Speed of epidemic holding *c/r* constant depends on their *levels*

Feedback Variations: Multipliers and Bubbles

Compare with Multiplier-Accelerator Model (Samuelson 1939)

The national income at time t, Y_t, can be written as the sum of three components: (1) governmental expenditure, g_t, (2) consumption expenditure, C_t, and (3) induced private investment, I_t.

$$Y_t = g_t + C_t + I_t.$$

But according to the Hansen assumptions

$$C_t = \alpha Y_{t-1}$$
$$I_t = \beta [C_t - C_{t-1}] = \alpha\beta Y_{t-1} - \alpha\beta Y_{t-2}$$

and
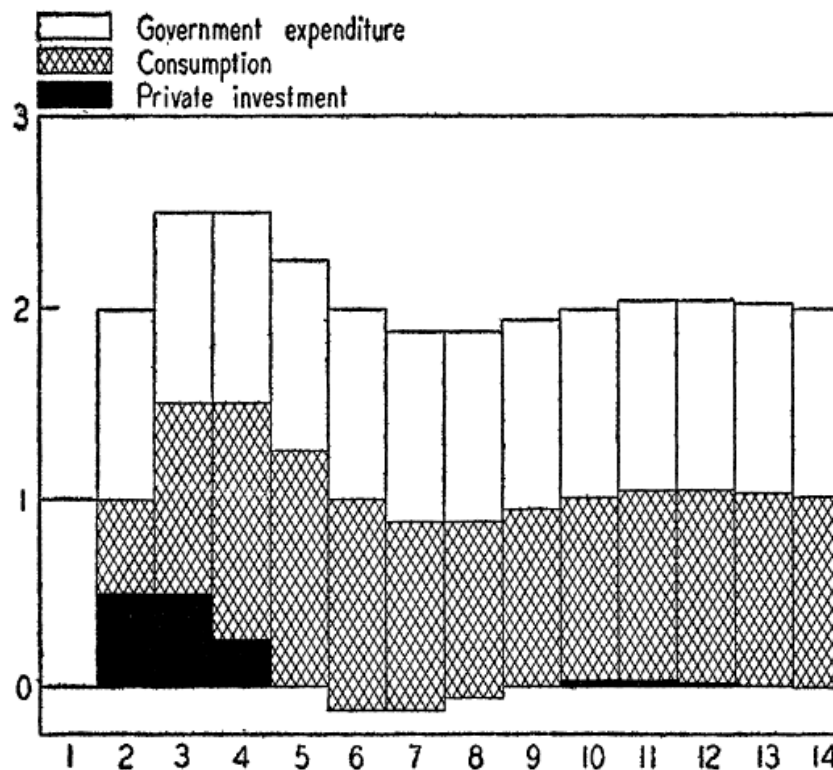
$$g_t = 1.$$

Therefore, our national income can be rewritten

$$Y_t = 1 + \alpha[1 + \beta] Y_{t-1} - \alpha\beta Y_{t-2}.$$

From Paul Samuelson . 1939. "Interactions between the Multiplier Analysis and the Principle of Acceleration" *Review of Economics and Statistics*

CHART I.—GRAPHIC REPRESENTATION OF DATA IN TABLE I

(*Unit: one dollar*)

# Uniqueness of Human Species in its Reliance on Narratives

- Homo narrans – Walter Fisher 1984
- Homo narrator – Stephen Jay Gould 1994
- Homo narrativus – Farrand and Weil 2001

Our Surveys of Stock Market Expectations since 1989

• "What do you think is the probability of a catastrophic stock market crash in the U. S., like that of October 28, 1929 or October 19, 1987, in the next six months, including the case that a crash occurred in the other countries and spreads to the U. S.? (An answer of 0% means that it cannot happen, an answer of 100% means it is sure to happen.)
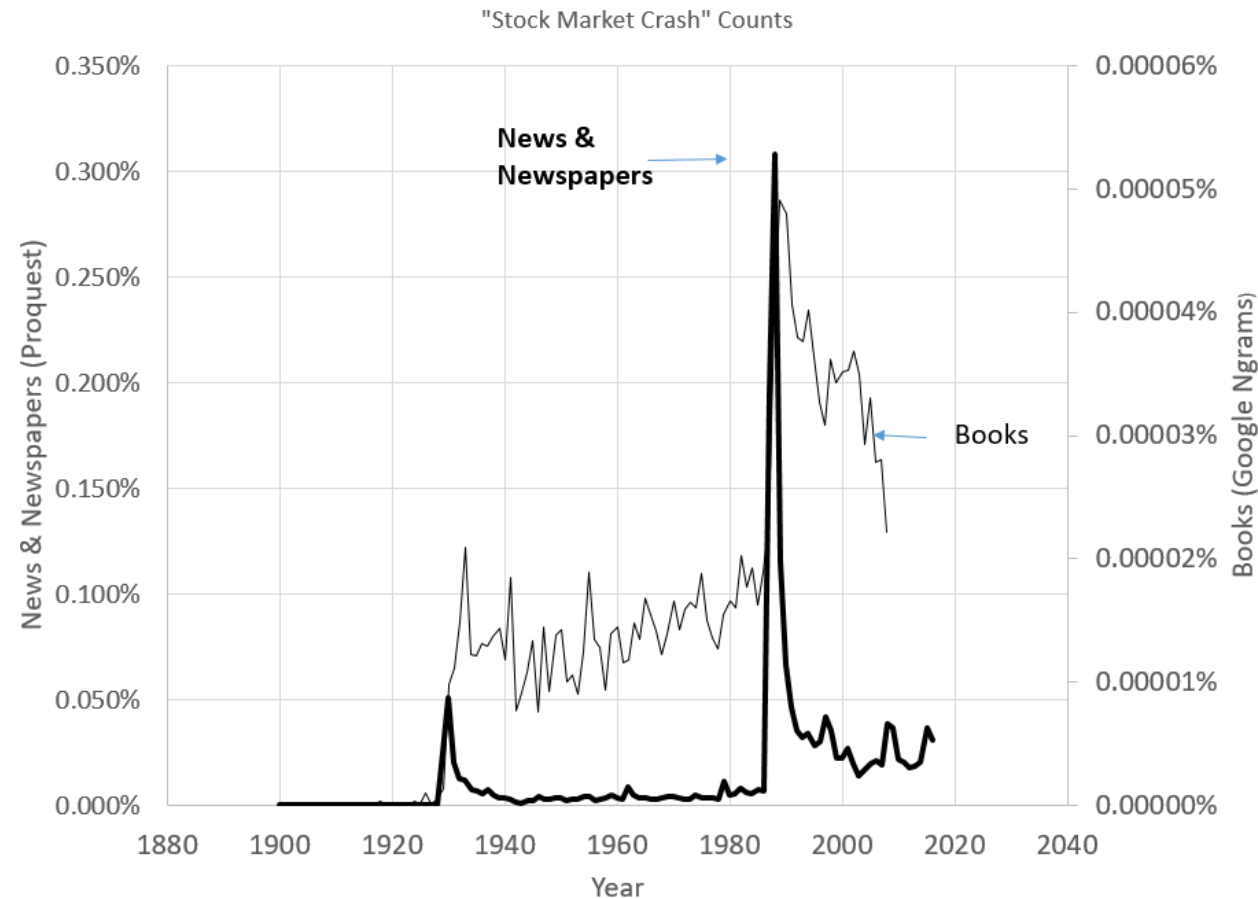
Probability in U. S.:_____%"

From William Goetzmann, Dasol Kim and Robert Shiller "Crash Beliefs from Investor Surveys" 2016
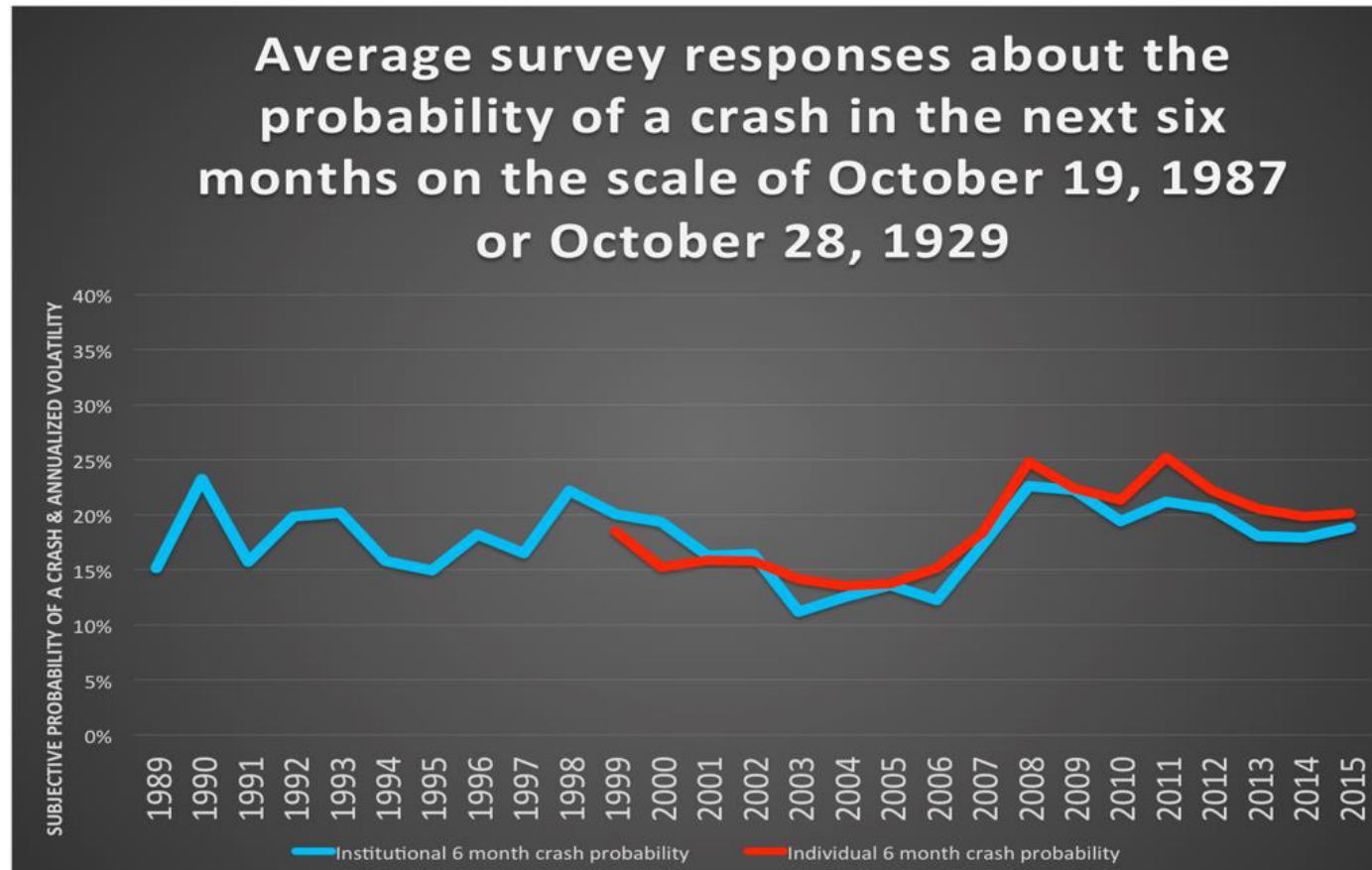
Profiteer Counts as Percent of Database each Year in New & Newspapers, Books, 1900-2008-16

Profiteer Counts
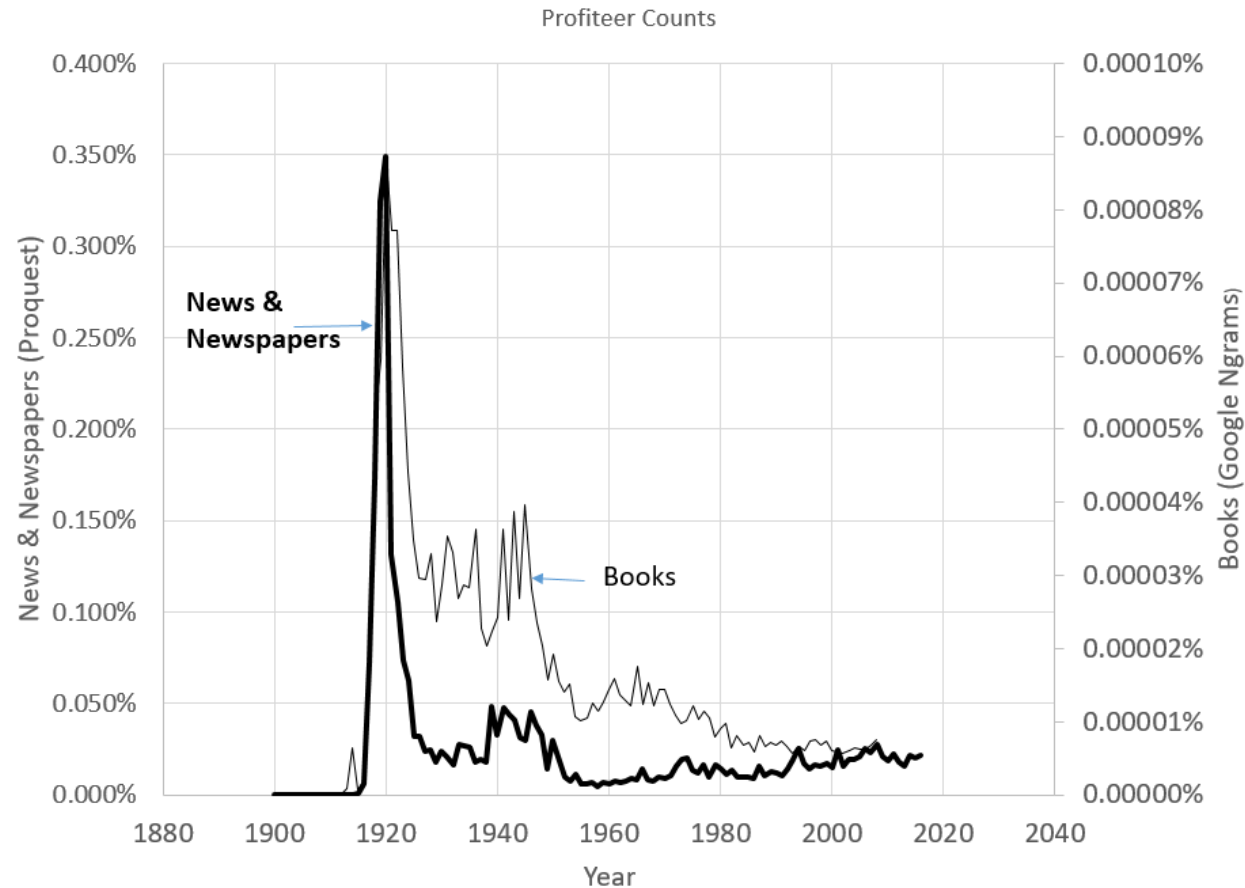
"Stock Market Crash" Counts as Percent of Database each Year

"Stock Market Crash" Counts

"Great Depression" Counts as Percent of Database each Year