

A bibliography of economics of the covid era

**Peter B. Meyer, Dan Gillman, Brandon Kopp,
Michael Schultz, Leo Sveikauskas, and Samantha Tyner**
U.S. Bureau of Labor Statistics
Aug 27, 2020

For CDC Health Economics Research Group

Findings and views expressed in this work do not represent the BLS.



Outline

- What we are doing
- Our sources
 - Working papers mostly
 - Data sources and findings from that literature
- Bibliography collection: a pilot project, on a wiki
- Compare to other covid-related bibliographies (CORD-19)
- Machine learning techniques and inferences
- Goals for today
 - Get feedback
 - Invite cooperation



What we are doing

- A team surveys the fast-moving literature on “covid economics”
 - Focused on the U.S. economy and the recession
- We summarize findings relevant to our agency
 - Employment, compensation, productivity, prices, and consumer expenditures
- BLS publishes data and does not recommend or implement economic policy.



Sources of literature

- The works of interest are working papers, mainly from:
 - NBER – National Bureau of Economic Research
 - The *Covid Economics* series at CEPR
 - IZA – Institute of Labor Economics
 - Becker-Friedman Institute (Univ of Chicago)
 - SSRN – Social Science Research Network
 - Think tanks – Brookings and Mercatus
- We monitor new papers in these series
 - Manually evaluating each abstract
 - Web scraping on occasion
 - Including many works not about covid or the recession explicitly
- We are not regularly watching for new papers at ArXiv, the Federal Reserve, World Bank, or other international sources
- Cooperation with other Feds such as those at CDC would be welcome.

It's a dynamic literature

- These are working papers, usually not peer reviewed
- Multiple outlets – a paper can appear in 4 places in just the past six months
- Some are not academic-style papers
 - Blog posts, news releases, policy briefs, and slide or video presentations can be relevant
- Revisions can appear within our time span also
 - June version versus May version
 - Can change title or authors quickly
 - In principle, inferences about relevance apply as of each time

Data sources used in this literature

Classic labor survey data

- Especially Current Population Survey, Current Establishment Statistics
- Occupation information: O*NET, OES
- Unemployment insurance claims
- SIPP

Payrolls

- ADP's National Employment Report

Financial

- Credit card transactions
- Consumer Expenditure Survey
- 401(k) investments
- Asset prices

Mobility, from cell-phone locations

- Google searches (Trends)
- Opportunity Insights project

Output: GDP and proxies

- Weekly Economic Index, ADS Business Conditions Index, HIS Markit

Work-at-home hours and job postings

- Homebase
- Burning Glass job
- Indeed's Hiring Lab

1918 pandemic comparisons

- Census and WWI records, internationally

And more. Many data sources → more complete picture

Substantive issues in the literature (1)

- Certain occupations and industries hit hardest
 - Retail, travel, restaurants, in-person services, child care
 - Superstar firms seem to benefit relatively, small firms more likely to shut down
- Employment effects by demographics
 - Worse for women in this recession; last one was worse for men
 - Worse for young, poor, minorities
- Very sharp downturn in March-April, partial recovery in May-June
 - Recoveries from recent recessions have been slow
- Shifts in consumer expenditures
 - Toward online purchases and delivery
 - Some online services and suppliers boomed
 - Wealthy are spending less, e.g. travel and restaurants

Substantive issues (2)

- Effects of formal lockdowns per se, versus public response to the virus
- Effects of social distancing on production/output
- Who does (or can do) remote work, or essential work
- ➔ Product demand, labor demand, and labor supply responses to covid

- Effects of support efforts, and their timing
 - Unemployment insurance
 - Economic Impact payments, and timing
 - Effects of the small-firm support by the Paycheck Protection Plan (Chetty et al, Autor et al)

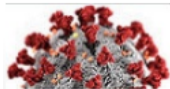
- International comparisons

- Possible follow-up with us

Bibliography is collected on a wiki

- A wiki is a Web site which can be edited from the browser and remembers past versions of pages
 - Here, in a relational database with Wikipedia's software
 - We are using an extension that lets us design our own tables
- There are four tables for the bibliography
- Each of these objects has a wiki page and a row in a table
 - ▶ 1752 **source works** about economics and/or the covid era
 - ▶ 254 **keywords** categorizing these works
 - ▶ 192 **authors** of works
 - ▶ 29 **publishers** of working paper series and journals





Mongey, Pilossoph, and Weinberg, 2020

- Simon Mongey, Laura Pilossoph, Alex Weinberg. Apr 26, 2020. Which Workers Bear the Burden of Social Distancing Policies? BFI Working Paper.
- Simon Mongey; Laura Pilossoph; Alex Weinberg. 2020. Which Workers Bear the Burden of Social Distancing Policies? NBER Working Paper 27085. May 2020
- Simon Mongey, Laura Pilossoph and Alex Weinberg. 2020. Which Workers Bear the Burden of Social Distancing Policies? *Covid Economics* 12: May 1 2020

Our comments [\[edit\]](#)

Comments – This provides a very focused look at which workers are most vulnerable because they perform non-essential work that cannot be performed with social distance. It documents tight correlation of this vulnerability with general economic vulnerability, especially in the tail of the distribution, and notes that a substantial portion of workers in that correlated tail are not particularly young. These workers are at risk of both unemployment, and if they work, to contracting the virus. The paper recommends that policy makers consider targeting assistance at the tail, where the social cost of assistance is low but the private benefit to recipients is high.

Original title	Which Workers Bear the Burden of Social Distancing Policies?
Simple title	Which Workers Bear the Burden of Social Distancing Policies?
Authors	Simon Mongey, Laura Pilossoph , Alex Weinberg
Year	2020
Publisher	NBER, BFI, Covid Economics
Keywords	NBER.EFG, NBER.LS, distancing, employment, workers, O*NET, CPS, PSID, MSA, ATUS, SafeGraph
Reviewed by	J
Relevance	
Relevance2	high
Memo sent	2020-05-06
AbstractURL	https://www.nber.org/papers/w27085
PaperURL	https://www.nber.org/papers/w27085.pdf
Abstract	What are the characteristics of workers in jobs likely to be initially affected by broad social distancing and later by narrower policy tailored to jobs with low risk of disease transmission? We use O NET to construct a measure of the likelihood that jobs can be conducted from home (a variant of Dingel and Neiman, 2020) and a measure of low physical proximity to others at work. We validate the measures by showing how they relate to similar measures constructed using time use data from ATUS. Our main finding is that workers in low-work-from-home or high-physical- proximity jobs are more economically vulnerable across various measures constructed from the CPS and PSID: they are less educated, of lower income, have fewer liquid assets relative to income, and are more likely renters. We further substantiate the measures with behavior during the epidemic. First, we show that MSAs with less pre-virus employment in work-from-home jobs experienced smaller declines in the incidence of 'staying-at-home', as measured using SafeGraph cell phone data. Second, we show that both occupations and types of workers predicted to be employed in low work-from-home jobs experienced greater declines in employment according to the March 2020 CPS. For example, non-college educated workers experienced a 4ppt larger decline in employment relative to those with a college degree.

Category: Document

Wiki page about a paper

➤ Hyperlinks, formatting, footnotes, work as on Wikipedia

➤ The structured info at the bottom is a row in a table which can be searched/queried

➤ Includes the abstract

Author page lists author's works

- The list of the author's works is dynamic; if one is added it will be listed
- Not a biography of the author
- This is useful for finding pages that are about the same paper

Simon Mongey

Publications by author Simon Mongey

- Akbarpour, Cook, Marzuoli, Mongey, Nagaraj, Saccarola, Tebaldi, Vasserman, and Yang, 2020 (Simple title: Socioeconomic Network Heterogeneity and Pandemic Policy Response, Keywords: [NBER.HC](#) • [NBER.HE](#) • [NBER.PE](#) • [NBER.TWP](#))
- [Berger, Herkenhoff, and Mongey, 2020](#) (Simple title: An SEIR infectious disease model with testing and conditional quarantine, Keywords: [NBER.EFG](#) • [NBER.HC](#) • [NBER.HE](#) • [SIR model](#) • [Testing](#) • [Quarantine](#) • [Model](#) • [Simulation](#) • [Transmission models](#))
- Mongey and Weinberg, 2020 (Simple title: Characteristics of Workers in Low Work-From-Home and High Personal-Proximity Occupations, Keywords: [Employment](#) • [Wages](#) • [Occupations](#) • [Gender](#))
- Mongey, Pilosoph, and Weinberg, 2020 (Simple title: Which Workers Bear the Burden of Social Distancing Policies?, Keywords: [NBER.EFG](#) • [NBER.LS](#) • [Distancing](#) • [Employment](#) • [Workers](#) • [O*NET](#) • [CPS](#) • [PSID](#) • [MSA](#) • [ATUS](#) • [SafeGraph](#))

Other names	Simon Mongey
Institutional affiliations	Department of Economics, University of Chicago; NBER
Keywords	

Category: Author

Keywords

- We can tag a work with keywords
- Keywords can have pages which explain them
- An automatic report on a keyword page shows which works were tagged with that keyword
- Subject classifications, like Journal of Economic Literature (JEL) codes, or NBER codes, are just keywords here

SIR model

An **SIR model** is an infectious disease **epidemiology** model has a population divided into three groups: S is a measure of the **susceptible** population, I measures the **infectious** population, and R is a measure of the **recovered**, immune or deceased individuals.^[1] With assumptions about transmission and recovery rates of the virus, one can model time series of the three categories.

Population categories can be added: An Exposed category makes it a Susceptible-Exposed-Infectious-Recovered (SEIR) model. SEAIRD stands for Susceptible, Exposed, Asymptomatic, Infected, Recovered, Deceased.^[2]

Mechanically, this model may be similar to economic models of the employed (E), unemployed (U), and the not in labor force (N). Often the transition rates from state to state are simple in such models, or they can be predicted by a variety of variables at each time.

References [\[edit\]](#)

- ↑ [Wikipedia:SIR model](#)
- ↑ [Aspri, Beretta, Gandolfi, and Wasmer, 2020](#)

Works tagged with keyword "SIR model"

- [Acemoglu, Chernozhukov, Werning, and Whinston, 2020](#) (Simple title: A Multi-Risk SIR Model with Optimally Targeted Lockdown, Keywords: [NBER.EFG](#) • [NBER.HE](#) • [NBER.PE](#) • [Macro](#) • [SIR model](#) • [Targeted policies](#))
- [Alfaro, Faia, Lamersdorf, and Saidi, 2020](#) (Simple title: Social Interactions in Pandemics: Fear, Altruism,

JEL J24

JEL code J24 covers "**Human Capital; Skills; Occupational Choice; Labor Productivity.**"^[1]

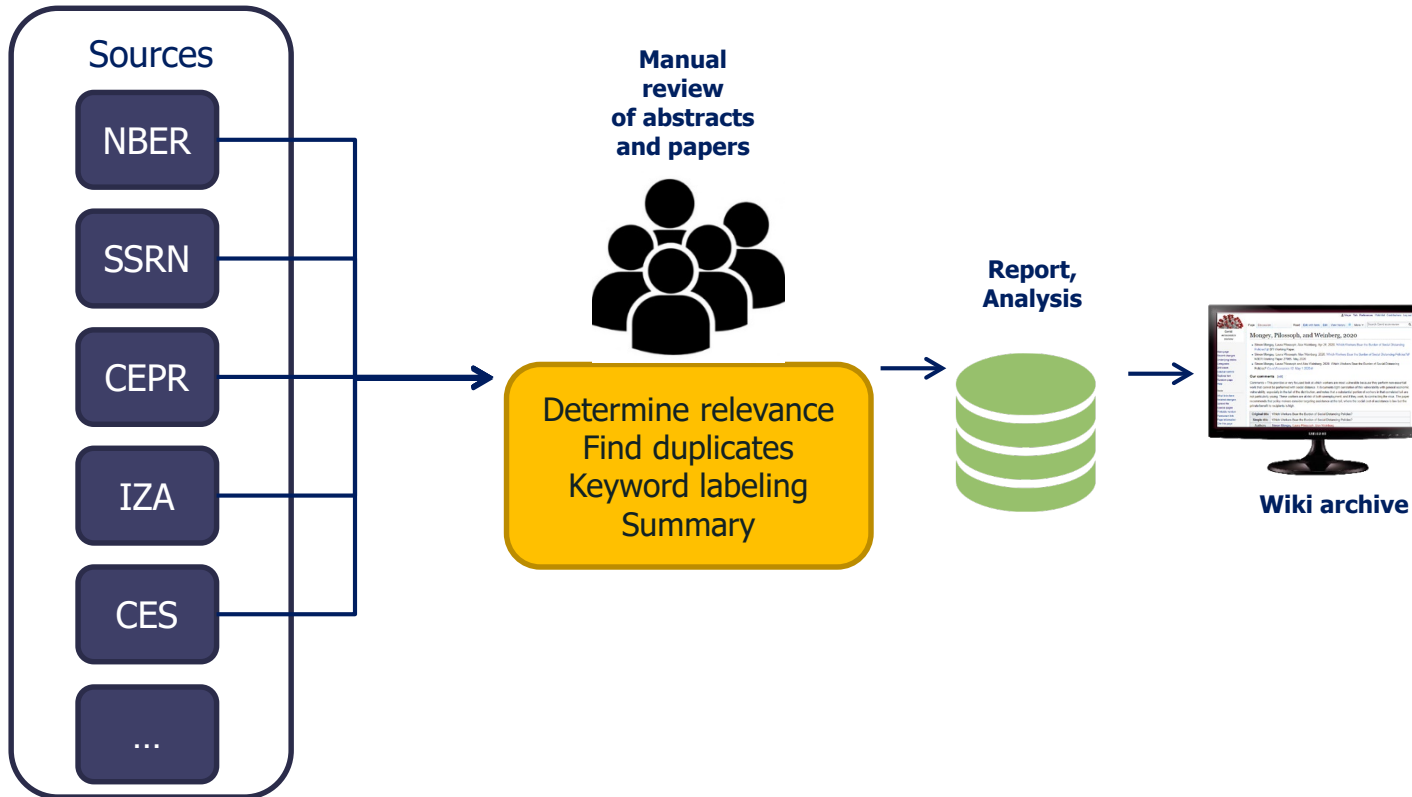
References [\[edit\]](#)

- ↑ [JEL codes at AEA site](#)[↗](#)

Works tagged with keyword "JEL J24"

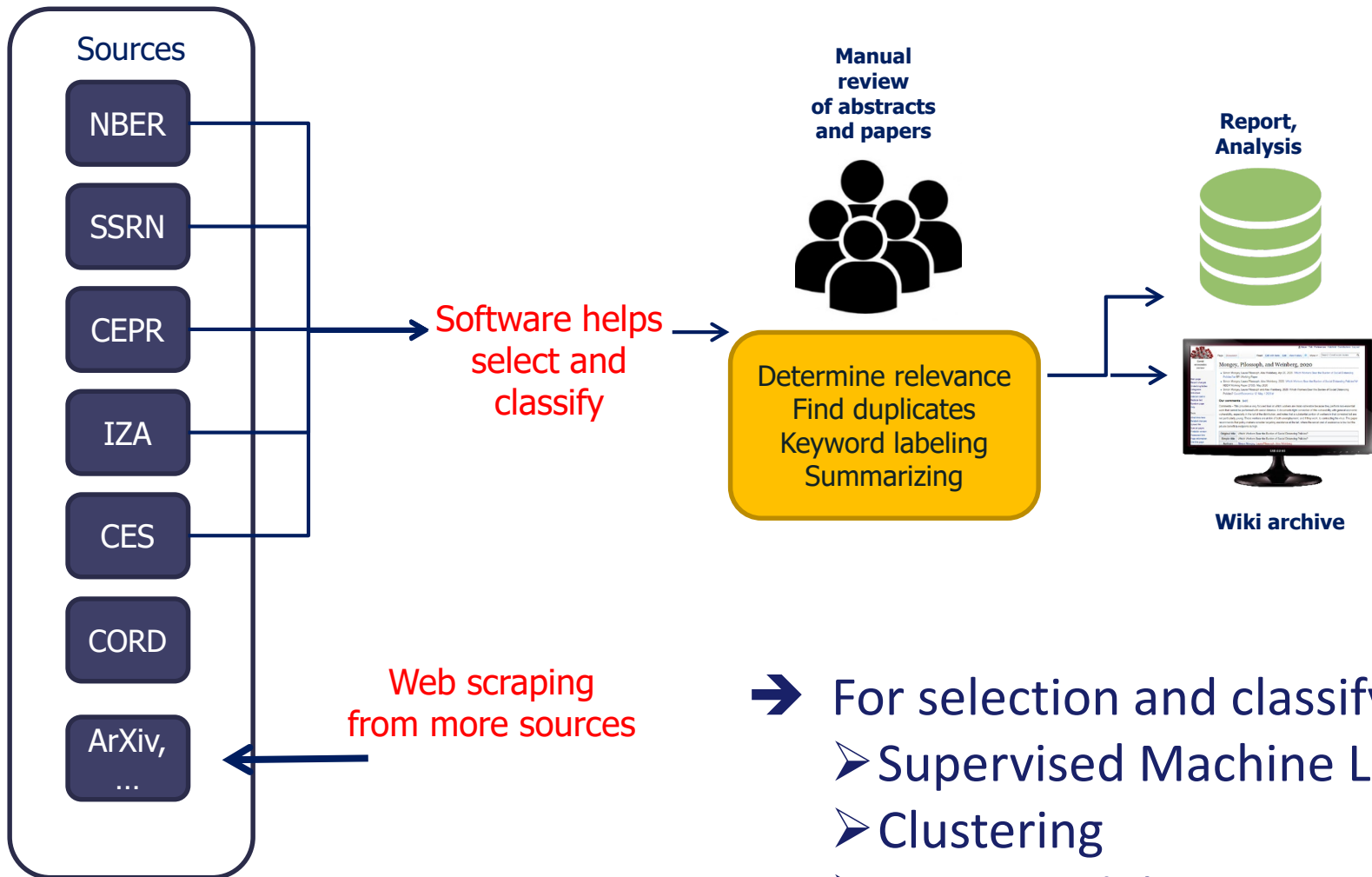
- [Adams-Prassl, Boneva, Golin, and Rauh, 2020a](#) (Simple title: Inequality in the Impact of the Coronavirus Shock: Evidence from Real Time Surveys, Keywords: [Inequality](#) • [JEL J21](#) • [JEL J22](#) • [JEL J24](#) • [JEL J33](#) • [JEL J63](#) • [COVID-19](#) • [Inequality](#) • [Labor market](#) • [Recessions](#) • [Unemployment](#))
- [Alipour, Falck, and Schüller, 2020](#) (Simple title: Germany's Capacities to Work from Home, Keywords: [Work from home](#) • [Germany](#) • [COVID-19](#) • [JEL D24](#) • [JEL J22](#) • [JEL J24](#) • [JEL O33](#) • [JEL R12](#))

Evaluation is all by hand now



Humans download papers, find duplicates of papers already downloaded, make a determination of relevance, assign keywords, and upload or enter data into a database

Possible help from data science



- ➔ For selection and classifying:
- Supervised Machine Learning
 - Clustering
 - Topic Modeling

Applying data science methods

- If we automate steps of this process, human reviewers can focus on determining **quality** and **novelty** in context
- We are exploring:
 - **Webscraping** to automate downloading, deduplication, and data entry
 - Prioritizing papers with **supervised machine learning** for review – new papers or from new sources such as CORD-19, SEAN, or ArXiv
 - **Classifying, clustering, topic modeling**: to categorize and provide keywords/labels.

Ideally this would speed up our reviews, enable us to cover more, or discover relevant works we would have missed.

The techniques would be relevant beyond this topic.



CORD-19 database

- Covid Open Research Dataset, sponsored by many partners
 - White House OSTP (Office of Science and Technology Policy), Kaggle, Paul Allen Institute, National Library of Medicine, Chan Zuckerberg Initiative, Microsoft Research, Georgetown U's Center for Security and Emerging Technology
- More than 224,000 scholarly papers, published and preprints
 - Papers refer to covid, SARS, MERS, or corona virus
 - Medicine and Biology are most common fields
 - Sources include PubMed, WHO, ArXiv, bioRxiv, and medArXiv
 - Some publishers license this kind of download for covid in particular
- Abstracts usually with full text, gathered by [semanticscholar.org](https://www.semanticscholar.org)
 - Duplicates checked, metadata standards applied, meant for data science
- Tools are available for the CORD-19 database, e.g. text mining
 - Wang et al (2020), Nance et al (2020)
 - People apply data science methods to it
 - There are Jupyter notebooks on Kaggle for use on CORD-19

Keyword search in CORD-19

- BLS program words: employment, unemployment, industry, occupation, compensation, productivity, prices, inflation, CPI, PPI, consumer expenditures
- Some search words bring in “False positives” (irrelevant works)
 - PPI – producer price index, to us but finds “protein-protein interaction” or “plasma protein isolates”
 - CPI – consumer price index, to us, but finds “candidate pandemic influenza”, “checkpoint protein inhibitor”
 - BLS – “basic life support”
 - Inflation – references to lung inflation, where we meant price inflation
- Better: economics, labor force, employment, current population survey, consumer price index, wages, salary, salaries, consumer spending, consumer expenditures
 - “economics” finds also “pharmoeconomics” and “employment” finds “unemployment.”
- This yielded 855 matches – many we have or would want, but many others too
 - A tiny fraction of the database
- Supervised learning can do better than a keyword search

How well can system determine relevance?

Supervised Machine Learning on our existing data

➤ Data

- 1,595 (~93%) papers on the wiki have abstracts
- Divide into **training set** of 1197 papers (75%), **test set** of 398 papers
- Response variable (1/0): high relevance: 69; lower relevance: 1526
- Predictors: Each word in the abstract, publisher (dummy variables)

➤ Models – Predicting classification with probability

- 3 classification models for text data chosen for experiment
- Regularized logistic regression (with lasso) – we'll focus on this
- Support vector machine
- Naïve Bayes model

Comparing 3 models on test set (n = 398)

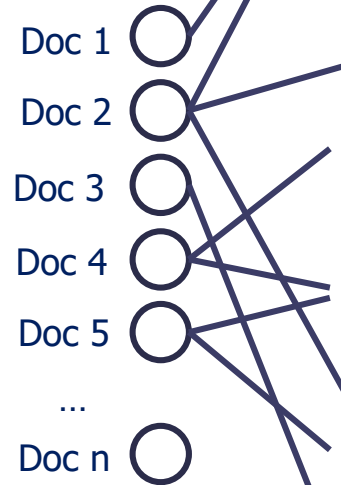
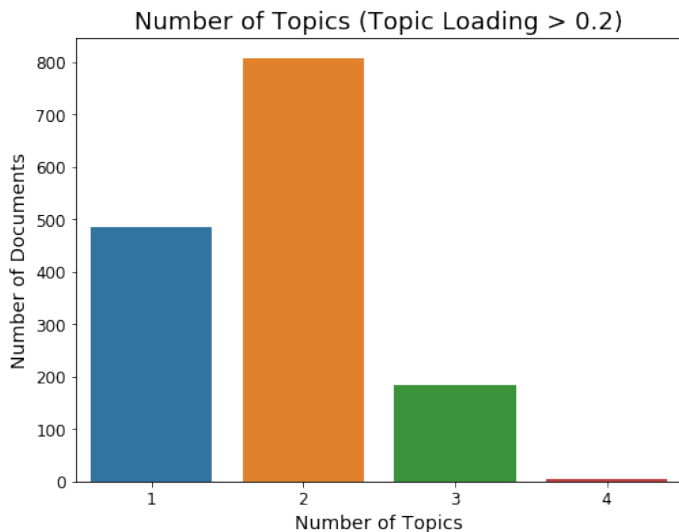
Model	Truth	Predicted High Relevance	Predicted Low Relevance
Logistic regression with lasso	High relevance	15	9
	Lower relevance	84	290
Naïve Bayes	High relevance	23	1
	Lower relevance	290	84
Support Vector Machine	High relevance	13	11
	Lower relevance	58	316

- Models use presence of words as predictors, e.g.: “estimates”, “economic”, “social”, “demand”, or bigrams: “19 pandemic”
- Tradeoff between specificity and sensitivity
- Dataset is class imbalanced: only ~4% of papers are marked as relevant
- Full text would be better than just abstract
- Could add data or adjust thresholds ; could add author information

Topic Modeling

The “topics” come from groupings of words across abstracts -- unsupervised

- Documents belong to multiple groups
- Strength of association with each group



Topic	Top 10 Words From Abstract
0	economic wealth public age social capital cost disease activity vaccine
1	bank program cost student experiment financial service decision information make
2	base time approach economic information new propose world identify technology
3	tax state order mobility consumer cigarette stay home home new county
4	health school public time woman impact lockdown stock policy economic
5	country income economic crisis household impact global policy government trade
6	death city case county number country rate change air pollution
7	firm market price shock demand policy rate sector productivity supply
8	work worker employment job labor home unemployment wage labor market occupation
9	social policy distancing social distancing infection economic lockdown individual disease intervention

Conclusions and future research

- Bibliography collection is a pilot project
 - We're not committed to continuing
- Can expand by Web scraping or downloading
 - CORD-19
 - ArXiv, SEAN, EUI, Federal Reserve, other sites
 - Places for improvement
- We would be interested to share bibliographies or techniques with CDCers

Contact Information

Peter B. Meyer

Research economist

BLS Office of Productivity and Technology

202-691-5678

meyer.peter@bls.gov

