



# The Potential for Alternative Data in Official Statistics

John Stevens

Federal Economic Statistics Advisory Committee Meeting  
June 2020

Any views expressed here are those of the presenter and not those of the Board of Governors of the Federal Reserve System or the Federal Reserve System.



# Outline

- Official statistics as ground truth
- Value/uses of alternative data
- Pandemic-related data challenges
- Two risks with alternative data

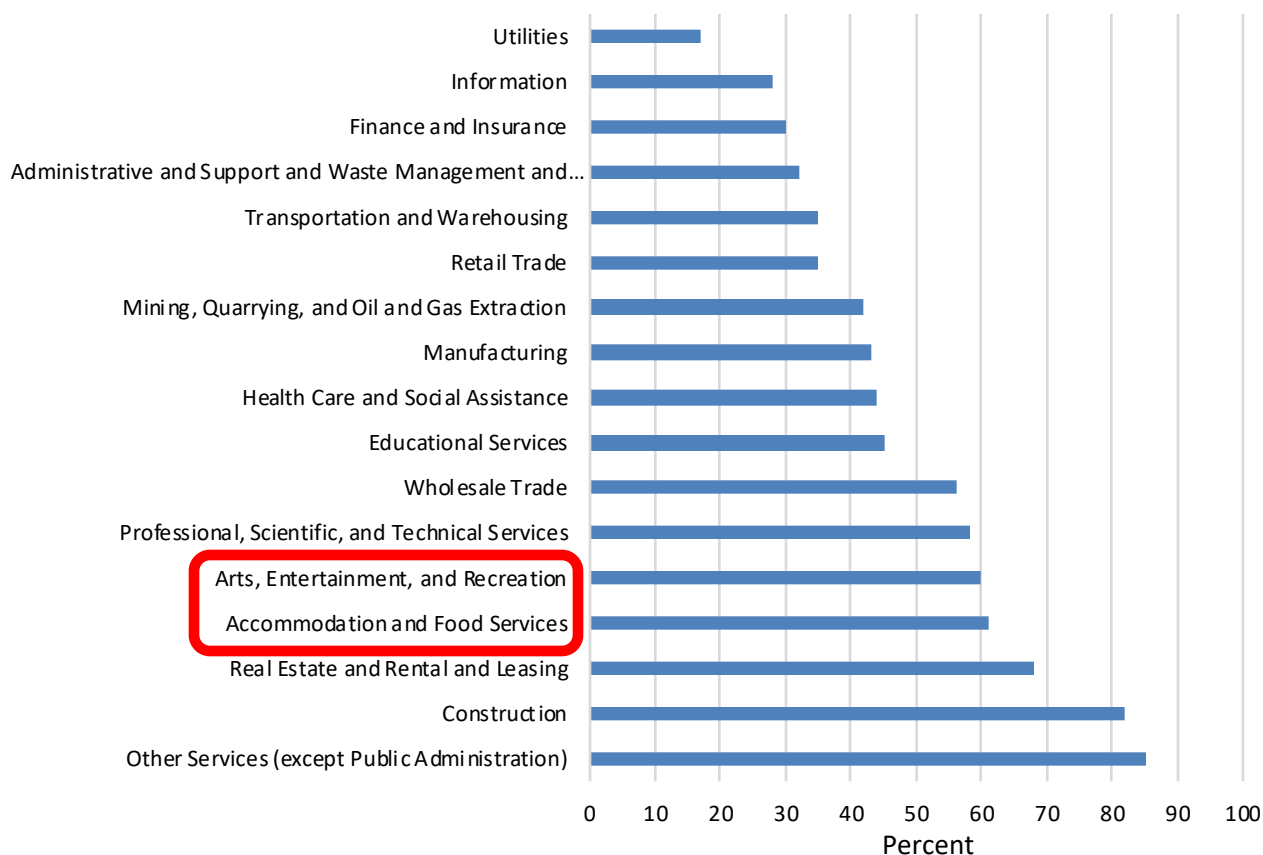


# Official Statistics as Ground Truth

- During a pandemic, basic facts are needed for policy development
  - What is the size distribution of firms?
  - How many businesses have closed?
  - How much cash do business have on hand?
  - What are the longer-term risks to firm creation?
- Creates important role for official statistics
  - Alternative data are often suggestive but not dispositive

# Official Statistics as Ground Truth

Share of Employment in Small Businesses

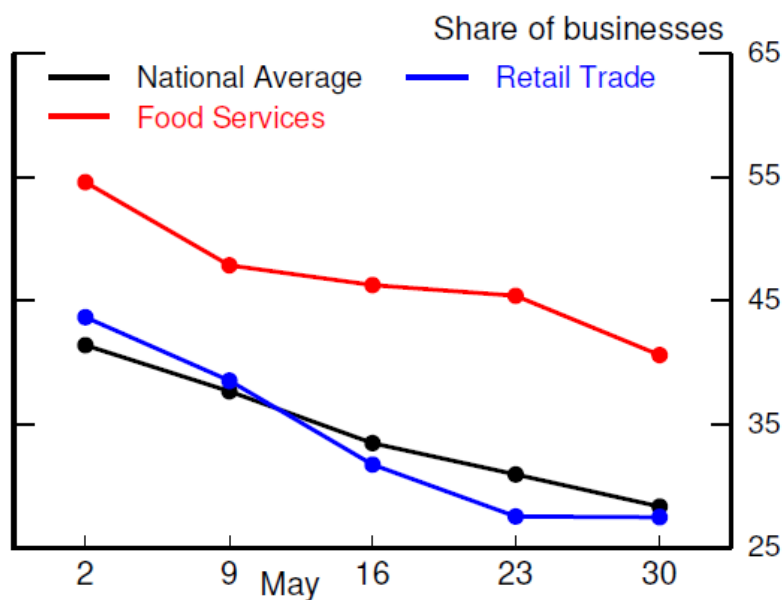


Source: Census Bureau's Statistics of U.S. Businesses.

Note: "Small business" here means as an enterprise with fewer than 500 employees.

# Official Statistics as Ground Truth

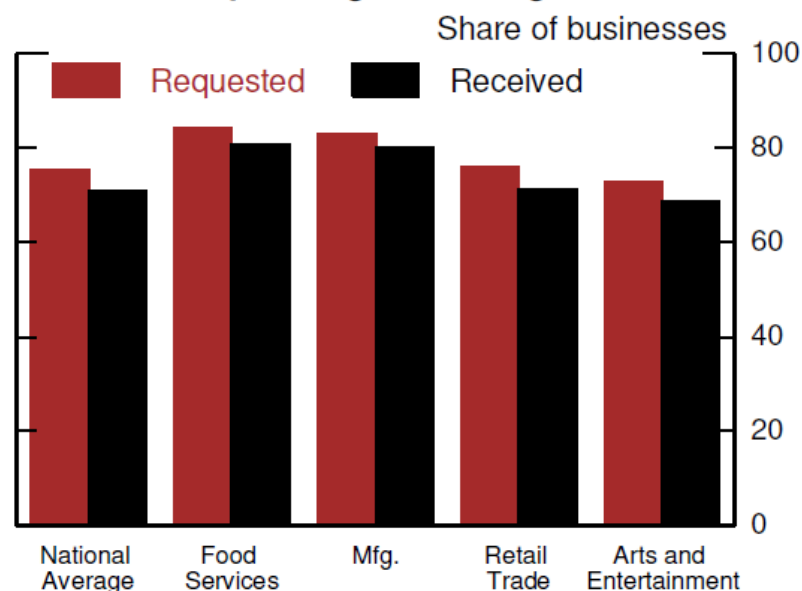
## Fraction Reporting Temporary Closings



Note: Survey question - "In the last week, did this business temporarily close any of its locations for at least one day?".

Source: Census Bureau Small Business Pulse Survey.

## Fraction Requesting/Receiving PPP Assistance

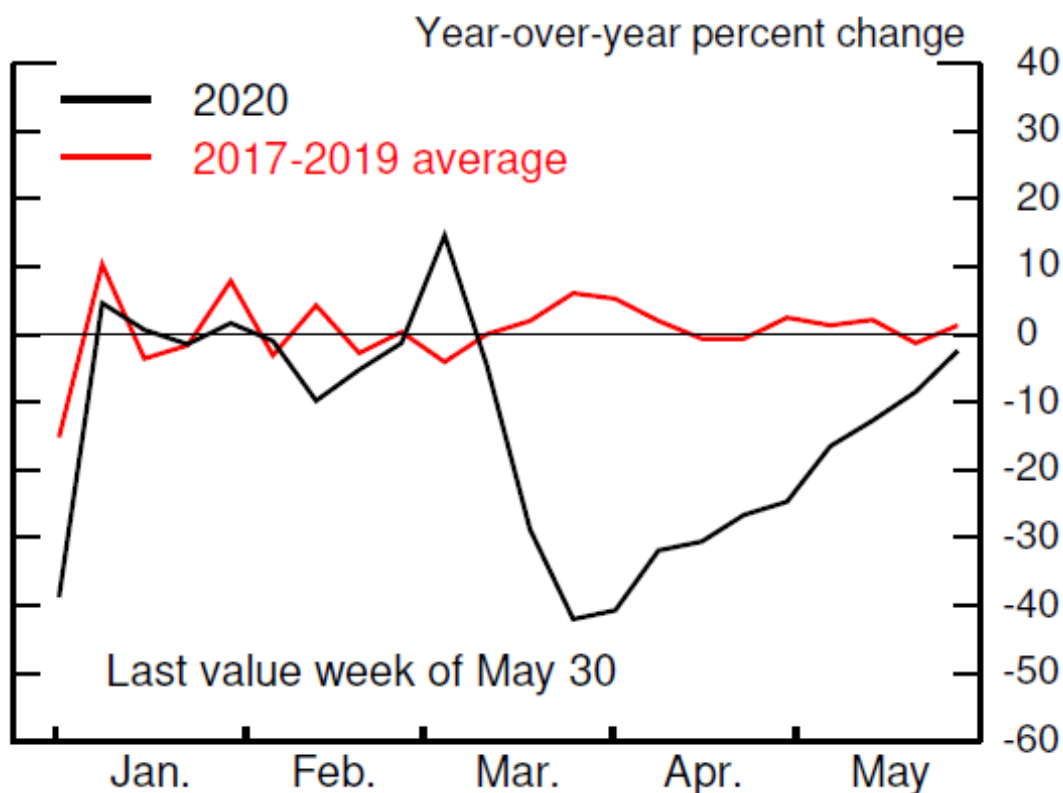


Note: Survey question - "Since March 13, 2020, has this business requested/received financial assistance from any of the following resources?".

Source: Census Bureau Small Business Pulse Survey.

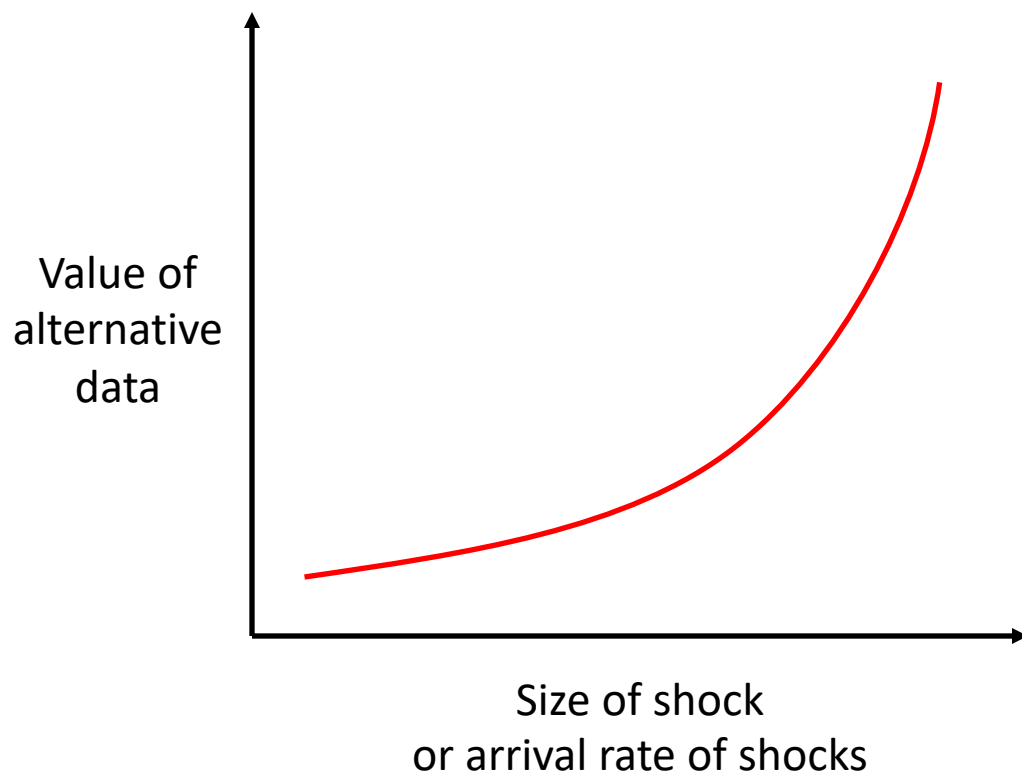
# Official Statistics as Ground Truth

## New Business Applications

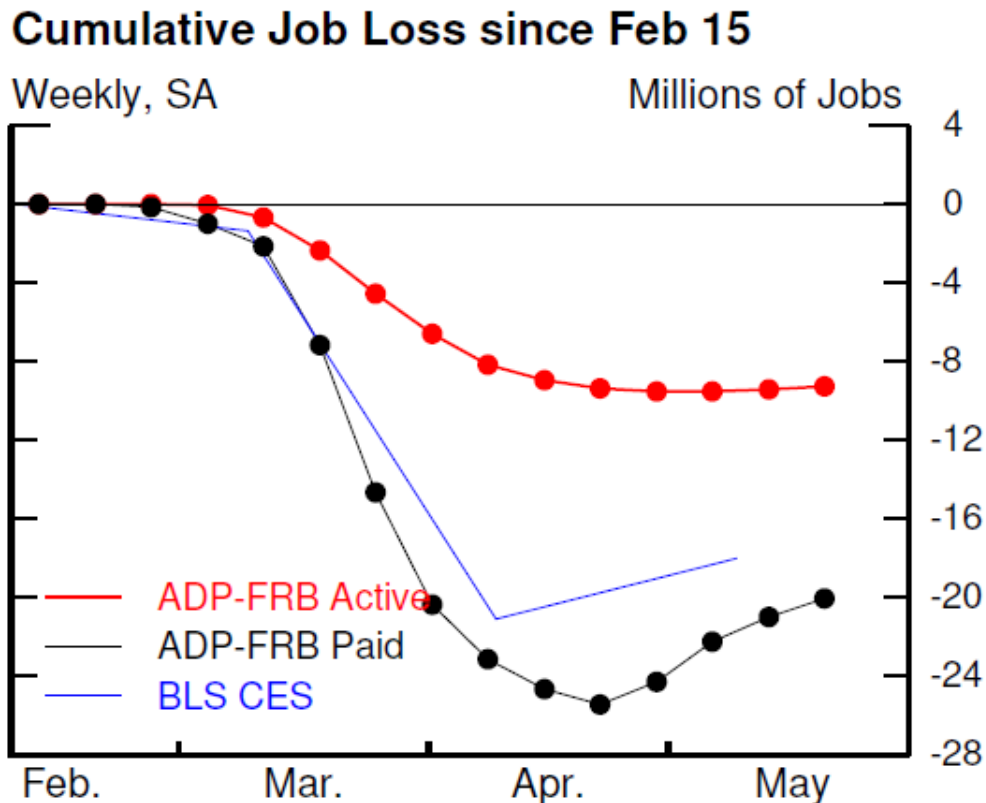


Note: Data derived from applications with planned wages for EIN.  
Source: U.S. Census Bureau.

# Value of Alternative Data



# Uses of Alternative Data



Note: Paid employment denotes workers who were issued a paycheck in a given pay period, while active employment also includes unpaid workers who remain active in the payroll system (e.g., are on unpaid furlough). Data extend through May 23.

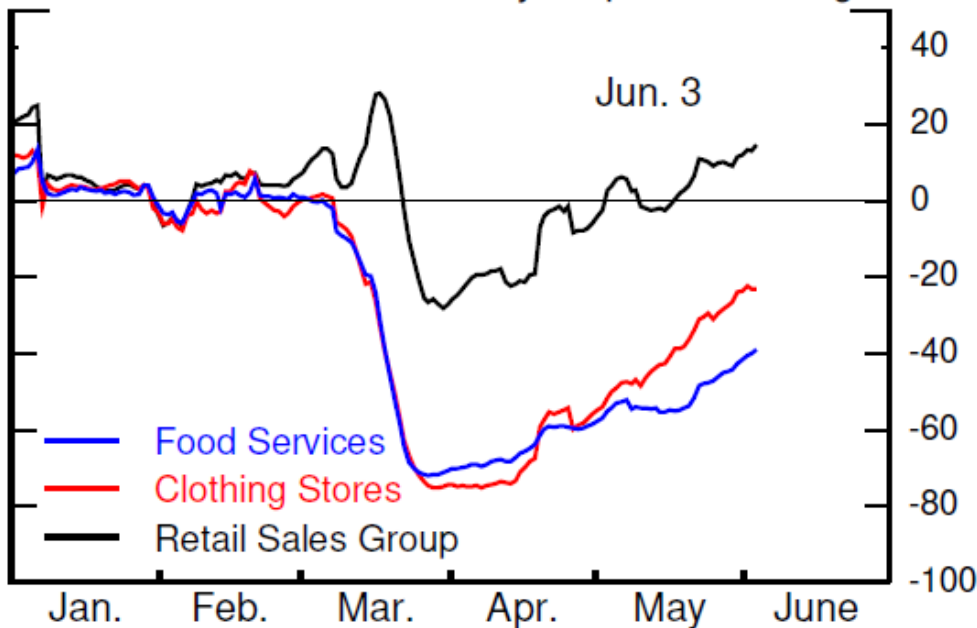
- Create *timely alternative estimates* that complement official estimates



# Uses of Alternative Data

## Daily Credit/Debit Spending

Year-over-year percent change

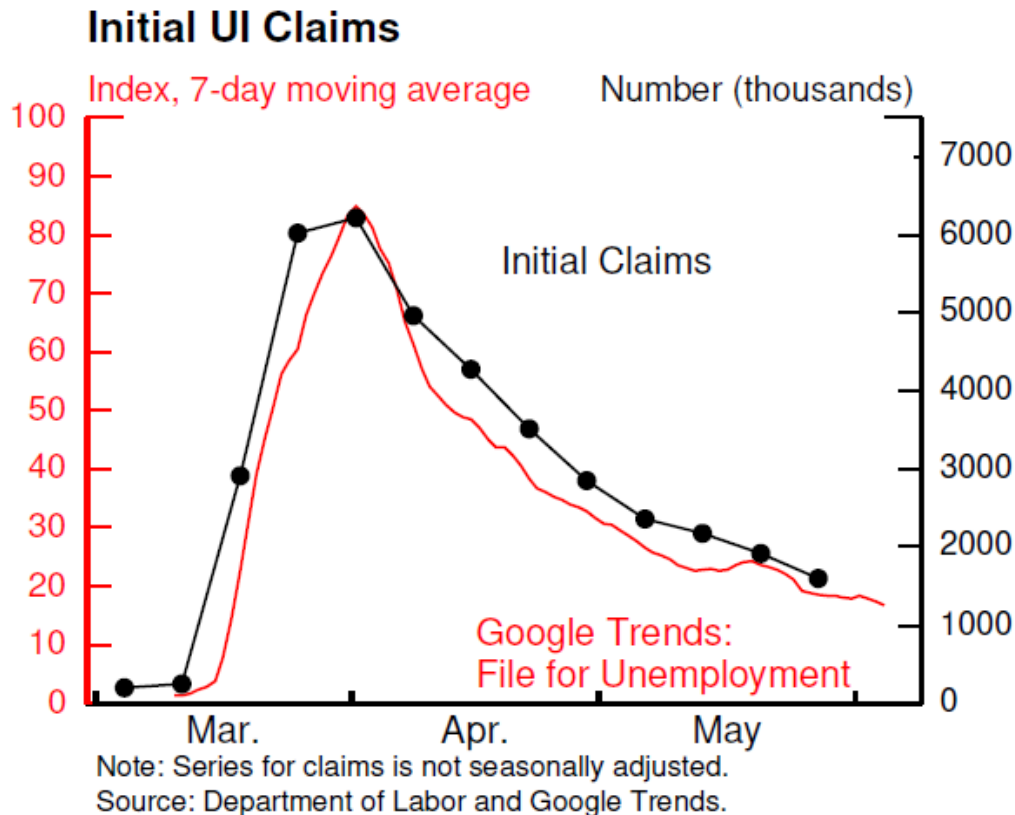


Note: 7-day moving average. Retail sales group excludes non-store merchants.

Source: Fiserv, Inc.

- *Verify changes in official statistics*
- *Answer questions that require more granular data*

# Uses of Alternative Data

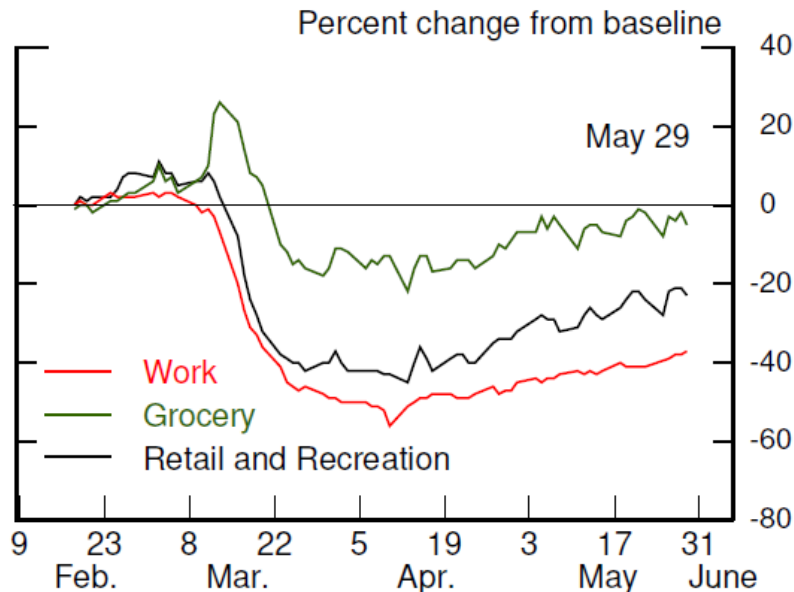


- Fill in *knowledge gaps*
- Examples  
timeliness,  
frequency,  
coverage,  
geography,  
industry,  
products, ...

# Uses of Alternative Data

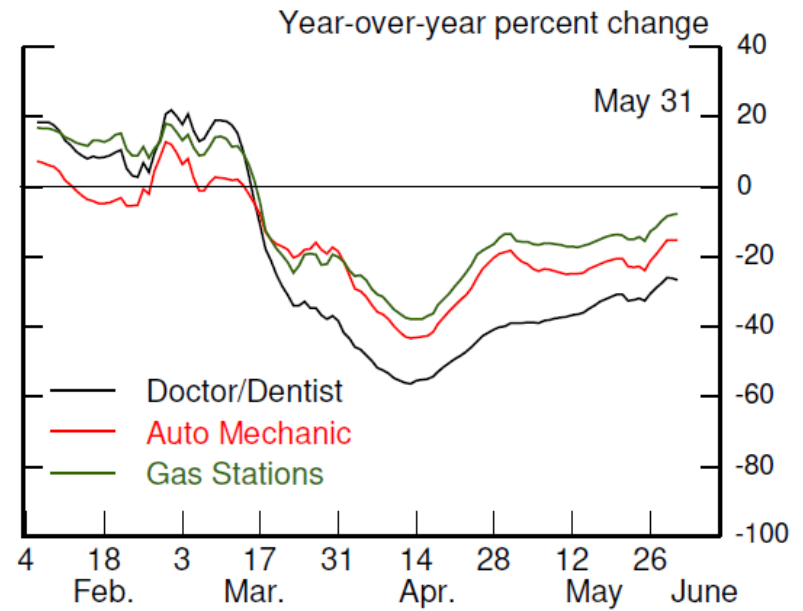
- Indicators w/o counterpart in official statistics may nonetheless help with a *narrative* about economic developments
- Example: Extensive vs. intensive margins

Google Mobility Index



Note: Baseline defined as median of Jan. 3 to Feb. 6 period.  
Source: Google Mobility Reports.

Daily Visits to Select Locations



Source: SafeGraph, using cell phone geo-location data.



# Pandemic-related Data Challenges

- Understanding establishment closures in real-time
  - Does non-response indicate a closure?
  - Are closures temporary or permanent?
  - Should a value of zero be imputed?
  - Always a challenge but more pronounced in a pandemic
- Aggregating data when weights are shifting rapidly
- Applies both to official statistics *and* to alternative data



# Alternative Data Risk #1: Bias


- “Big Data Paradox: The bigger the data, the surer we fool ourselves” (Xiao-Li Meng, 2018)
- How to mitigate?
  - Reweight using low-frequency, benchmark information from official statistics
  - Careful selection of sources for alternative data
  - Use multiple sources



# **Alternative Data Risk #2:**

## **Lack of Diversity in Data Sources**

- Alternative data source that becomes a critical input may expose an agency to:
  - Risk of hold-up in contract negotiations
  - Risk of disruption from contract termination
  - Risk of reduced innovation (from not looking at additional data sources)
  - Risks from non-response or data transmission problems
- Could take months or years to recover from an unexpected contract termination



# An Example:

## Nonresponse/Transmission Risk

- “Traditional” data
  - In 2012, four firms accounted for 45% of revenue in “petroleum and coal products manufacturing”
  - Nonresponse → industry-level non-disclosure
- Alternative data
  - Four payroll processors issue ~50% of employees’ paychecks
  - One payroll processor covers a share of employment that is comparable to the BLS’s responses to the establishment survey
  - Using payroll processing data as a survey substitute could create a non-disclosure risk *for the entire economy*



## Even so...

- Alternative data are worth pursuing...
  - The questions they help answer are important
  - A few data sources almost cover the population
- ...but take steps to mitigate risks...
- ...and don't lose sight of official statistics' comparative advantage...





# Comprehensive (often low-frequency) Data

- It is not flashy or timely, but...
- ...it is critical to informed policymaking
- ...it is needed for industry definitions and to benchmark high-frequency data (including alternative data)
- ...it helps bound our understanding
  - What establishments are in the path of a hurricane?
  - If a policy targets small firms, how much funding may be needed?
- (Higher-frequency official statistics are still useful though!)