Capitalizing Data:

Theoretical Framework and Case Studies

By Rachel Soloveichik

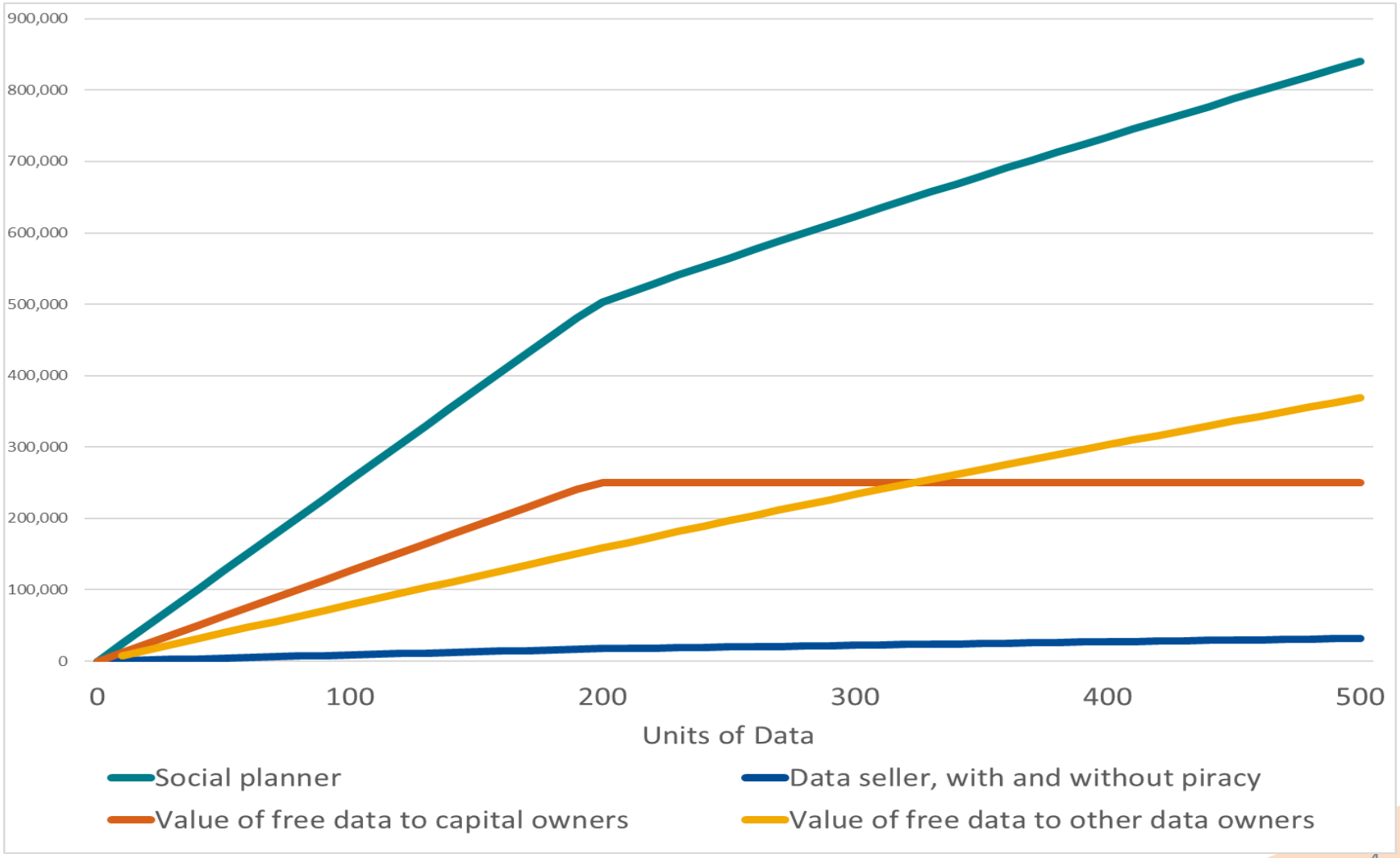
Outline of Presentation

- Data are defined as easy-to-copy information
  - Storage methods: Servers, CD's, paper, genetics, VHS tapes, and so on
- Theoretical framework
  - Data can be sold or given free by the owner of complementary capital
  - Identify parameters where free data yield more value than sold data
  - Argue that these identified parameters are common in the real world
- Case studies focus on four types of free data: tax, individual credit, driving, and marketing
  - These four types alone had $2.1 trillion of free data creation in 2017
  - Back-of-the-envelope calculations suggest that total privately funded free data creation may have been $6.7 trillion in 2017
- Recalculate GDP when data are capitalized

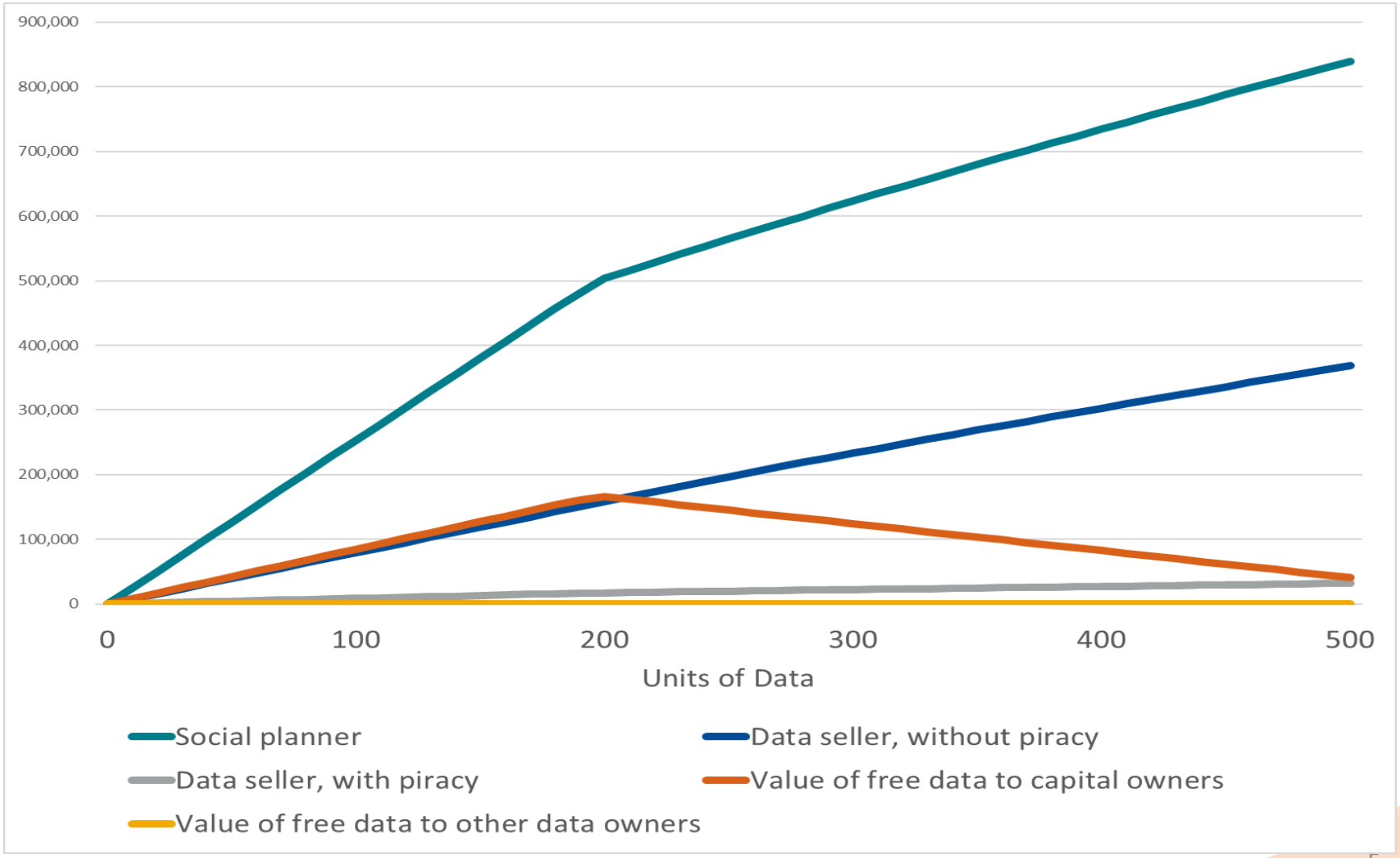
Theoretical Framework

- Firms 1 to n use v capital assets and w data types in a modified CES production function
  - Parameters s_1, s_2, …, s_n determines each firms' skill at using data
  - Parameter ρ determines complementarity between data and capital
  - Parameter σ determines complementarity between data types
  - Parameters $\beta^{1,1}$ $\beta^{2,1}$,. , $\beta^{1,v}$, …, $\beta^{w,v}$ determine how specific each capital asset is to each data type
- V separate capital owners rent separate capital assets, K^1 to K^v at fixed rental rates, r^1 to r^v
- Two ways to earn money from data:
  - Data owners can sell data to firms 1 to n at a price p per unit
  - Data owners can make their data free in return for a lump sum payment from either a capital owner or another data owner

Data Are Strong Complements to Data &
Data are Weak Complements to Capital

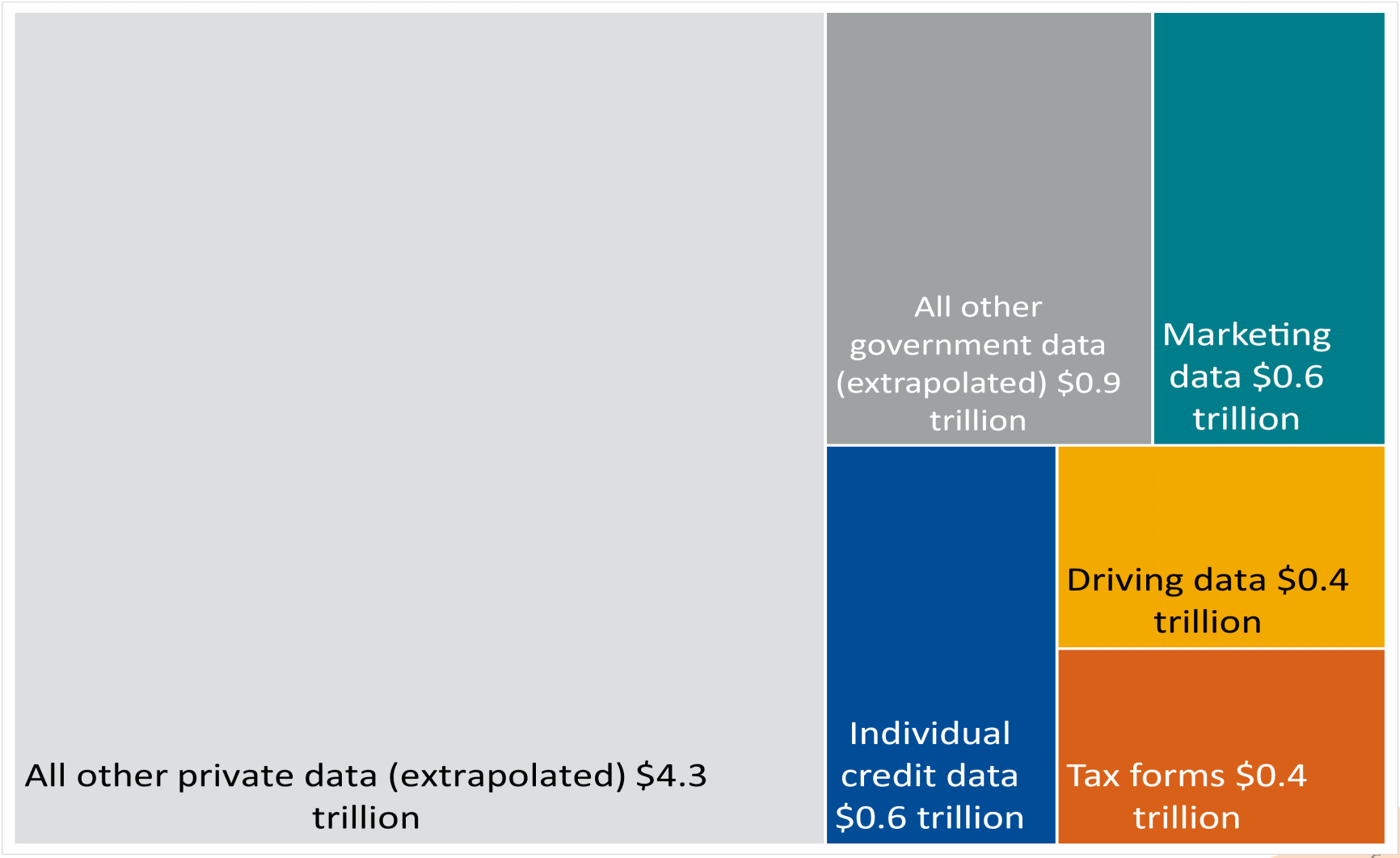


Data Are Substitutes to Data &
Data are Weak Complements & Specific to Capital

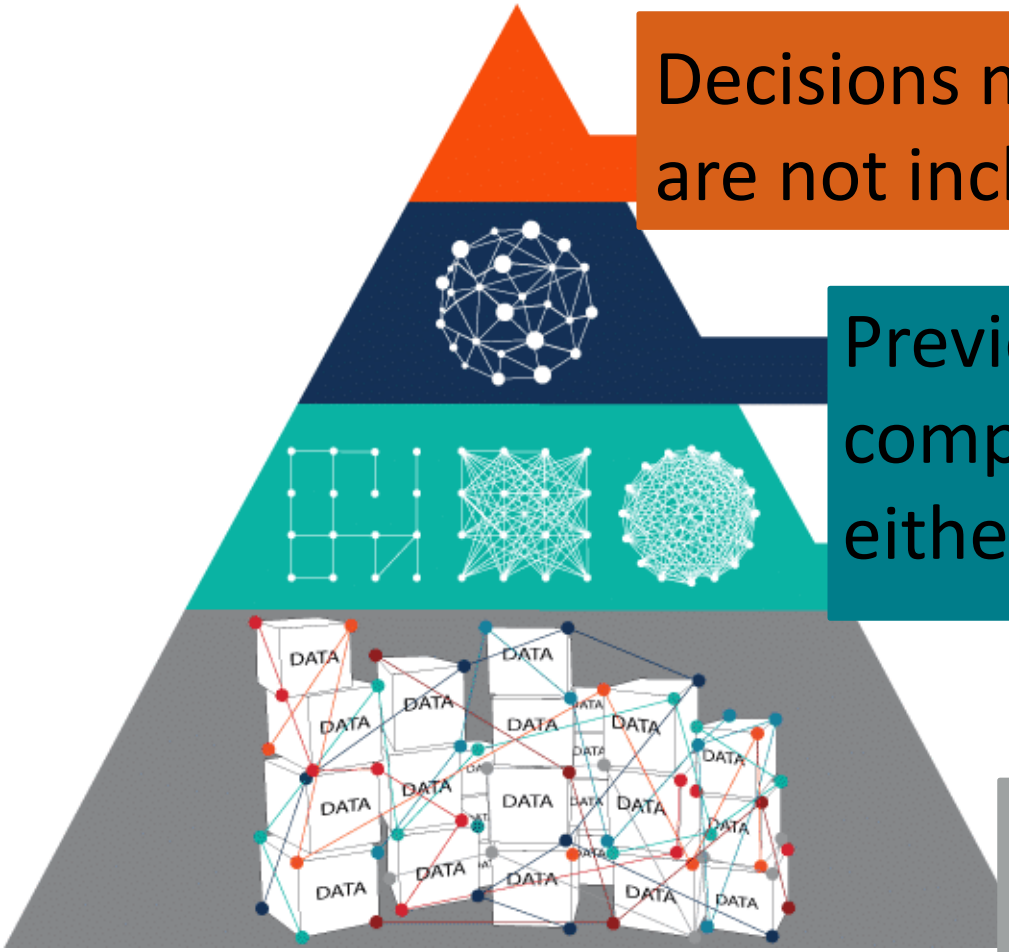


Total Free Data Creation in 2017

Extrapolated from Case Studies and Model



Data Pyramid: Free vs. Sold



Decisions made using data are not included in data

Previous research studied complex datasets that are either used in-house or sold

I study simple free data which underly complex datasets

Data Creators

Impact of Capitalizing Data on GDP by Industry

- Almost every activity generates data
  - Workers fill out tax forms when they start a job
  - Borrowers, banks, and debt collectors create credit data
  - Drivers, police officers, and insurers create driving data
- Data are sometimes primary output
  - E.g. laboratories produce medical data but not treatment
- Data are generally secondary output
  - Data given to governments are taxes in-kind
  - Data given to workers are non-cash benefits
  - Data given to customers are part of a bundled purchase
  - Primary output ↓ by the value of data given to customers
- Household data creation doesn't impact GDP

Platforms and Data Owners

Impact of Capitalizing Data on GDP by Industry

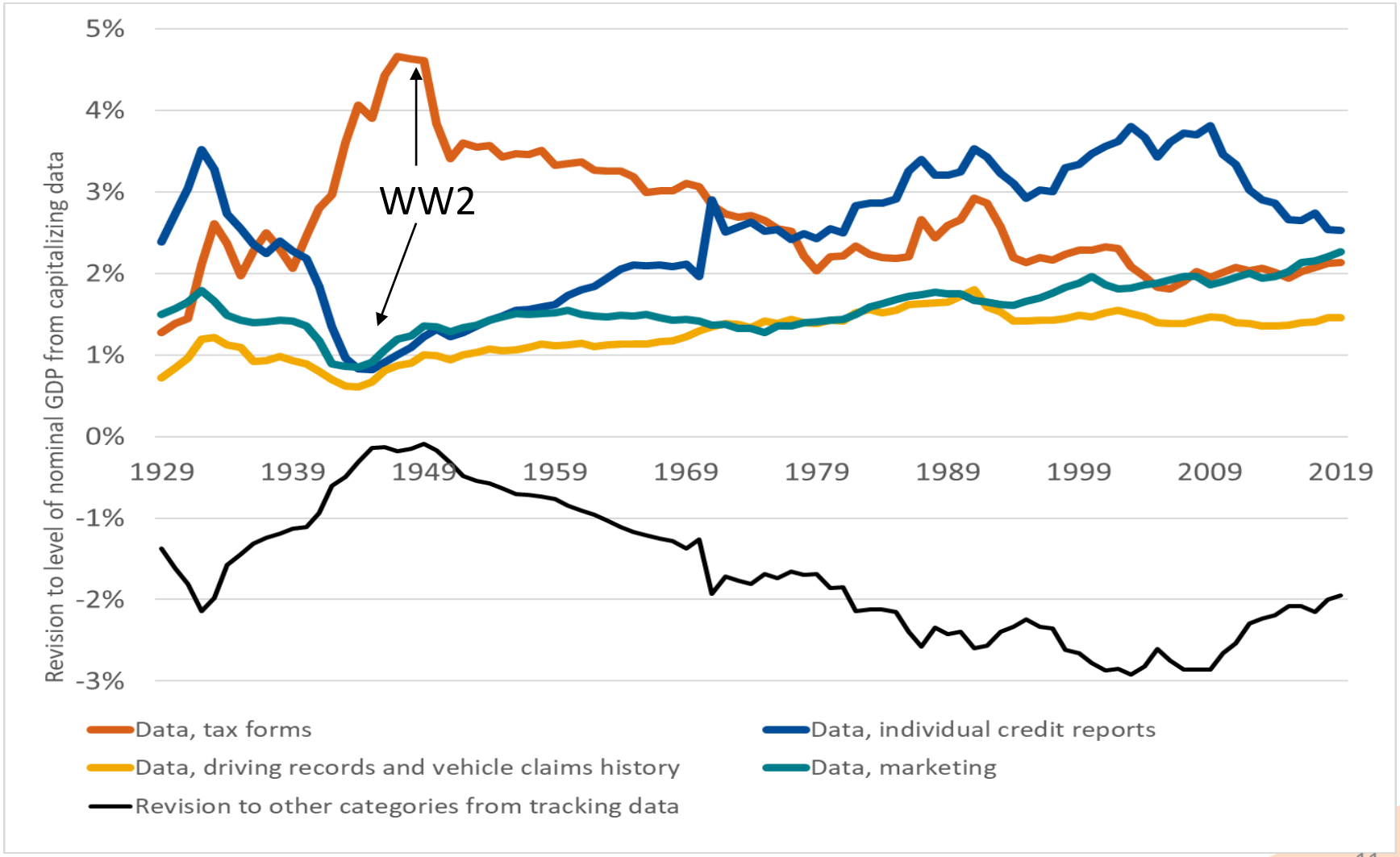
- Platforms organize data but don't control data
  - Neither inputs nor outputs change when data are capitalized
- Both sold and own-account data can be free
  - Sold data are owned by their purchaser
  - Own-account data are owned by their creator
  - This paper treats a government mandate to create data as a tax in-kind and therefore considers those data to be owned by the government
- Business data are tracked as intangible capital
  - Intermediate inputs ↓ by payments for data
- Consumer data are tracked as durables PCE
  - Nondurable PCE ↓ by payments for data

Data Users

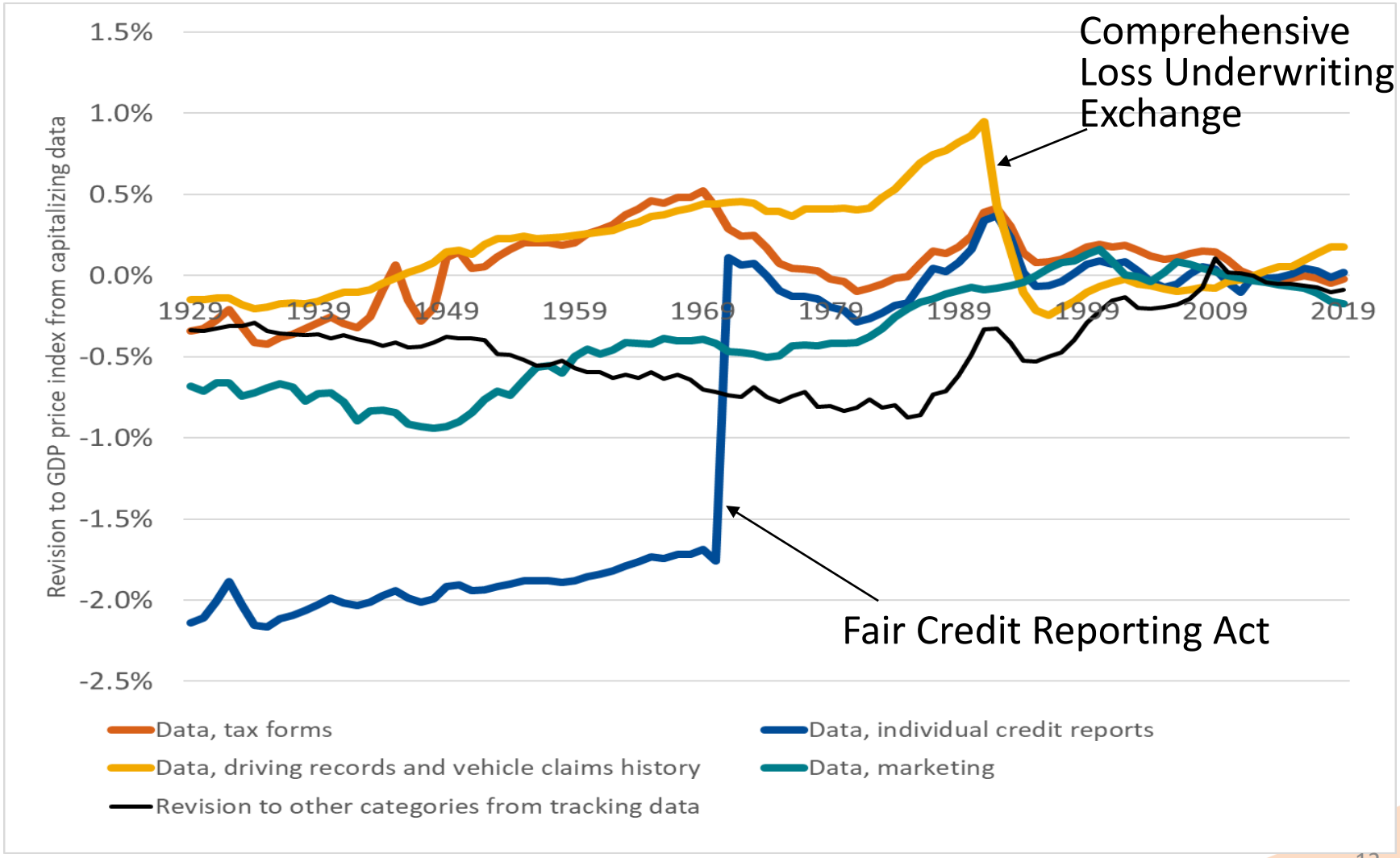
Impact of Capitalizing Data on GDP by Industry

- When calculating value-added by industry, data sharing is treated as a barter transaction
  - Customers who share their free data pay lower prices
  - Workers who share their free data about earn higher wages
  - Business owners who share their free data earn higher profits
- Businesses use data to target customers, hire workers, determine prices/wages, and so on
  - Intermediate input ↑ by the value of free data services used
  - Private output ↑ by the discounts given in return for customer data
- Governments use data to determine tax obligations, administer programs, and so on
  - Government output ↑ by the value of data services used

Nominal GDP Revision



GDP Price Indexes in Case Studies



Conclusion

- Theoretical framework where data can either be sold or given free
  - Identify plausible parameters where free data dominates sold data
  - Argue that many important and expensive to create data types have parameters that fall in the free region
- Privately funded data creation in 2017
  - Tax data: $0.4 trillion; individual credit data: $0.6 trillion; driving data: $0.4 trillion; marketing data: $0.6 trillion; other data: $4.6 trillion?
- Real GDP revisions in case studies
  - Tax data: growth rose a total of 3.2 percentage point between 1929 and 1948 due to Social Security and individual income taxes
  - Credit data: growth fell a total of 1.5 percentage point around 1970 due to the Fair Credit Reporting Act
  - Driving data: growth rose a total of 1.5 percentage point around 1992 due to the Comprehensive Loss Underwriting Exchange