Big Data and the Measurement of Prices and Real Economic Activity: A Better, Faster, Cheaper Approach

Stephen J. Redding
Princeton and NBER

David E. Weinstein
Columbia and NBER

June, 2016

Bar Codes and Measurement: New Approaches

- *Better*: more general approach than that currently used
  - Ability to integrate all products and services for which markets operate and prices and quantities can be measured: e.g. all goods, transportation, retail and wholesale trade, lending and insurance markets, etc.
- *Faster*: Capacity to develop economic statistics in real time
  - What is real output or inflation today?
- *Cheaper*: Exploit existing databases without need of field agents
  - Computation of CPI corresponding to 20 percent of consumer expenditures can be done with millions of observations on a small server in minutes

Bar-Code Data: Challenges

- Product turnover is phenomenal
  - In a typical year, 40% of household's expenditures are on goods that were created in the last 4 years; 20% are on goods that will not survive 4 years.
  - How do we measure prices when the set of goods is changing?
- Conventional price indexes (e.g. Laspeyres versus Jevons) can yield very different inflation measures
  - Need to think about what we mean when we talk about inflation
- No single product firms or industries
  - How do we move from information on bar codes to firms to industries to aggregate output?

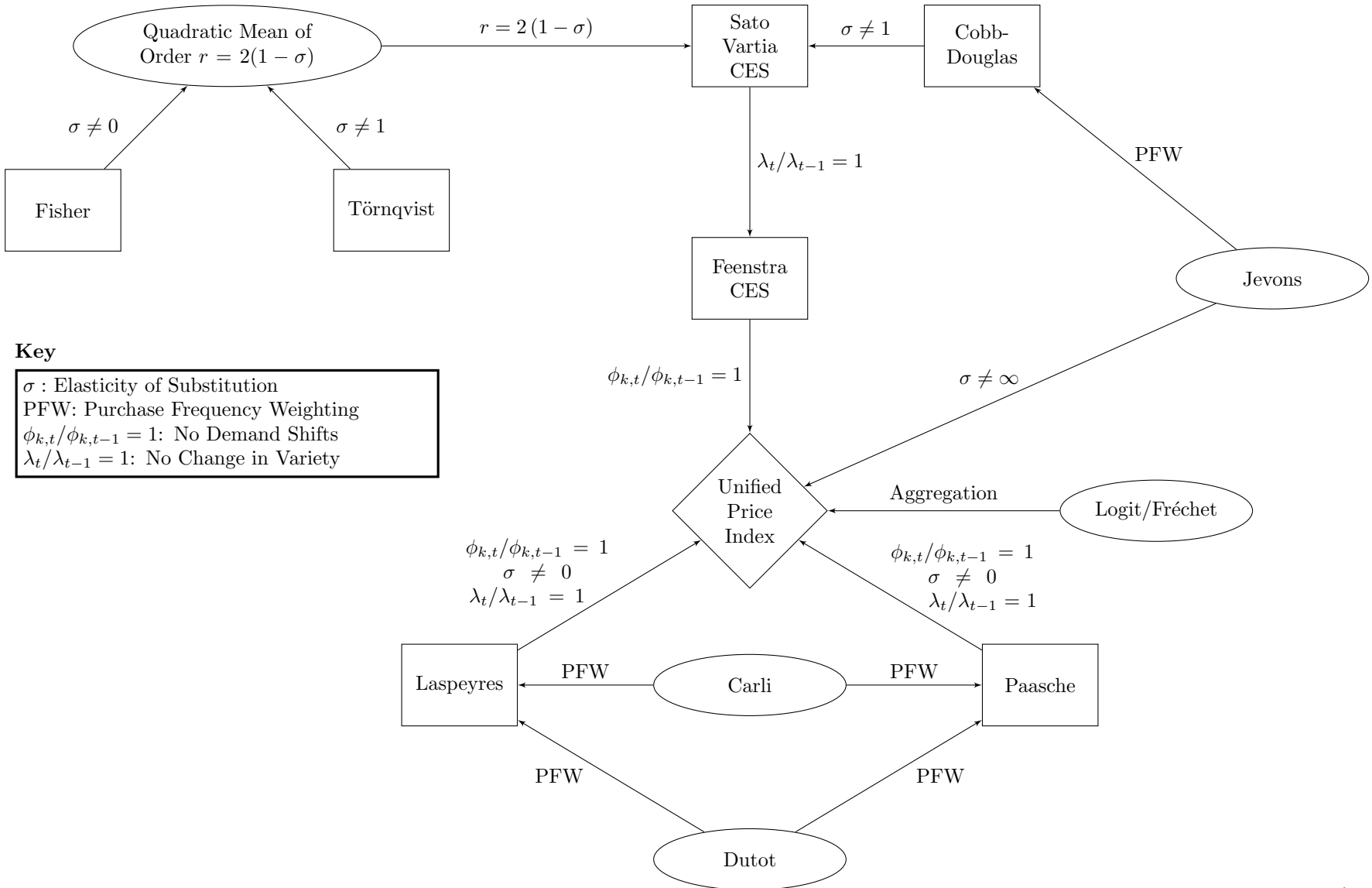
The State of the Literature

- The Axiomatic Approach
  - Dutot (1738), Carli (1764), Jevons (1865), Laspeyres (1871), Paasche (1874), Fisher (1922), Törnqvist (1936) developed indexes that have "common sense" properties *but are not based on consumer theory*
  - First three constitute basis for 97 percent of measures of inflation and real output used in official statistics
- The Economic Approach
  - After Konüs (1924), economists have believed that price indexes should be based on consumer theory
    - A price index is the ratio of two unit expenditure functions
  - Economists can derive standard price indexes when the number of goods and demand for each good are constant
- Existing indexes are inconsistent with duality, time reversibility, and/or aggregation when the number of goods and demand for each good vary over time
- Our method solves this problem

Unified Price Index (UPI)

- We develop a "unified approach" that consistently estimates welfare and demand *even when demand for each good is time varying*
  - Requires only data on prices and expenditure shares
  - Allows for entry and exit of goods over time
  - Identifies a unique elasticity of substitution (σ)
  - Satisfies constant aggregate utility function
  - Yields consistent aggregation from micro to macro
  - Nests all major micro, macro, and statistical approaches to price measurement
  - Generalizes to heterogeneous groups of consumers
- Existing exact price indexes are biased in the presence of mean zero demand shocks
  - *Substitution bias :* consumers substitute away from goods whose price has risen
  - *Consumer valuation bias :* consumers substitute towards goods that they desire more

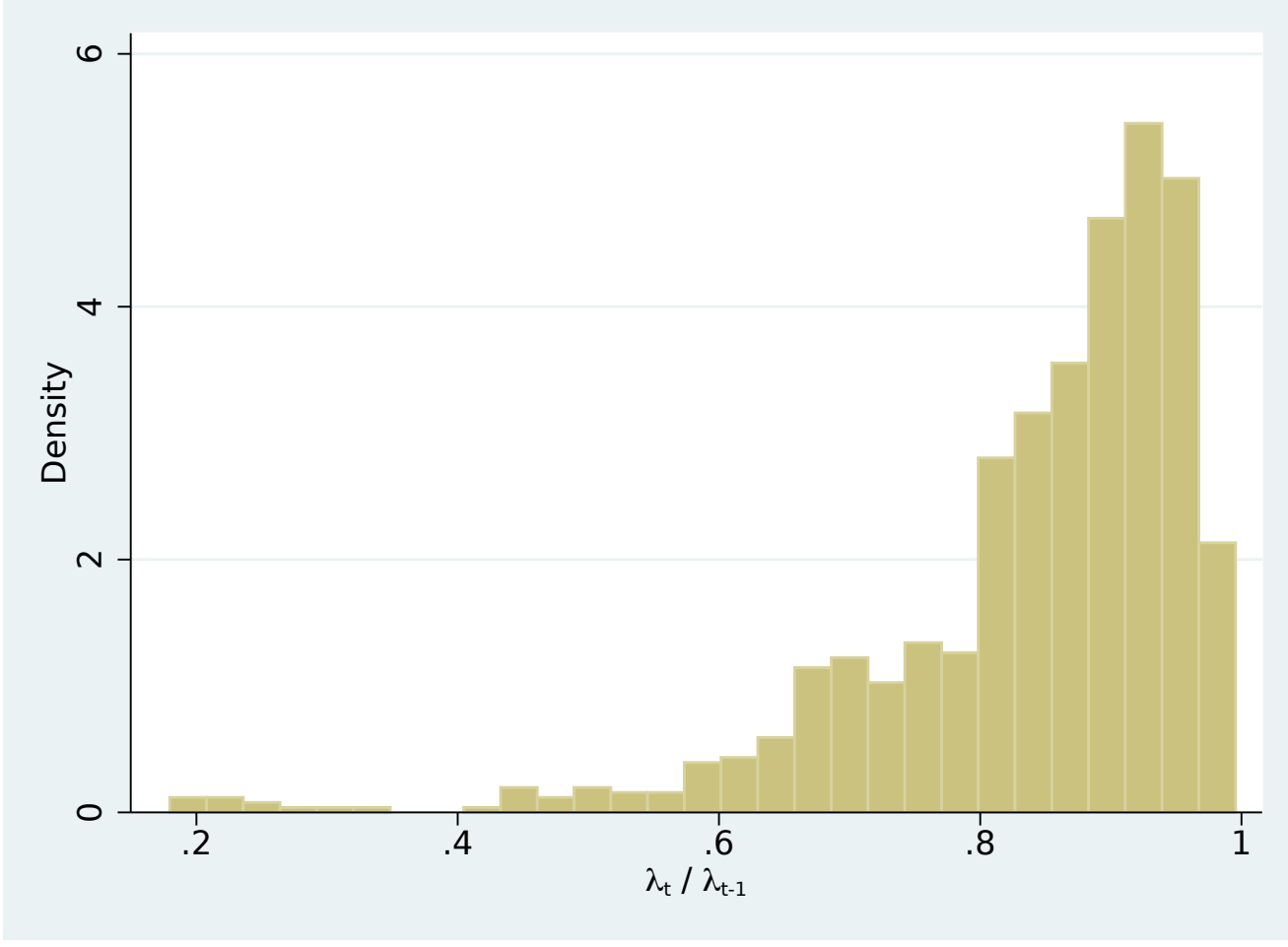
The UPI and Extant Price Indexes



Pilot Using Nielsen HomeScan Data

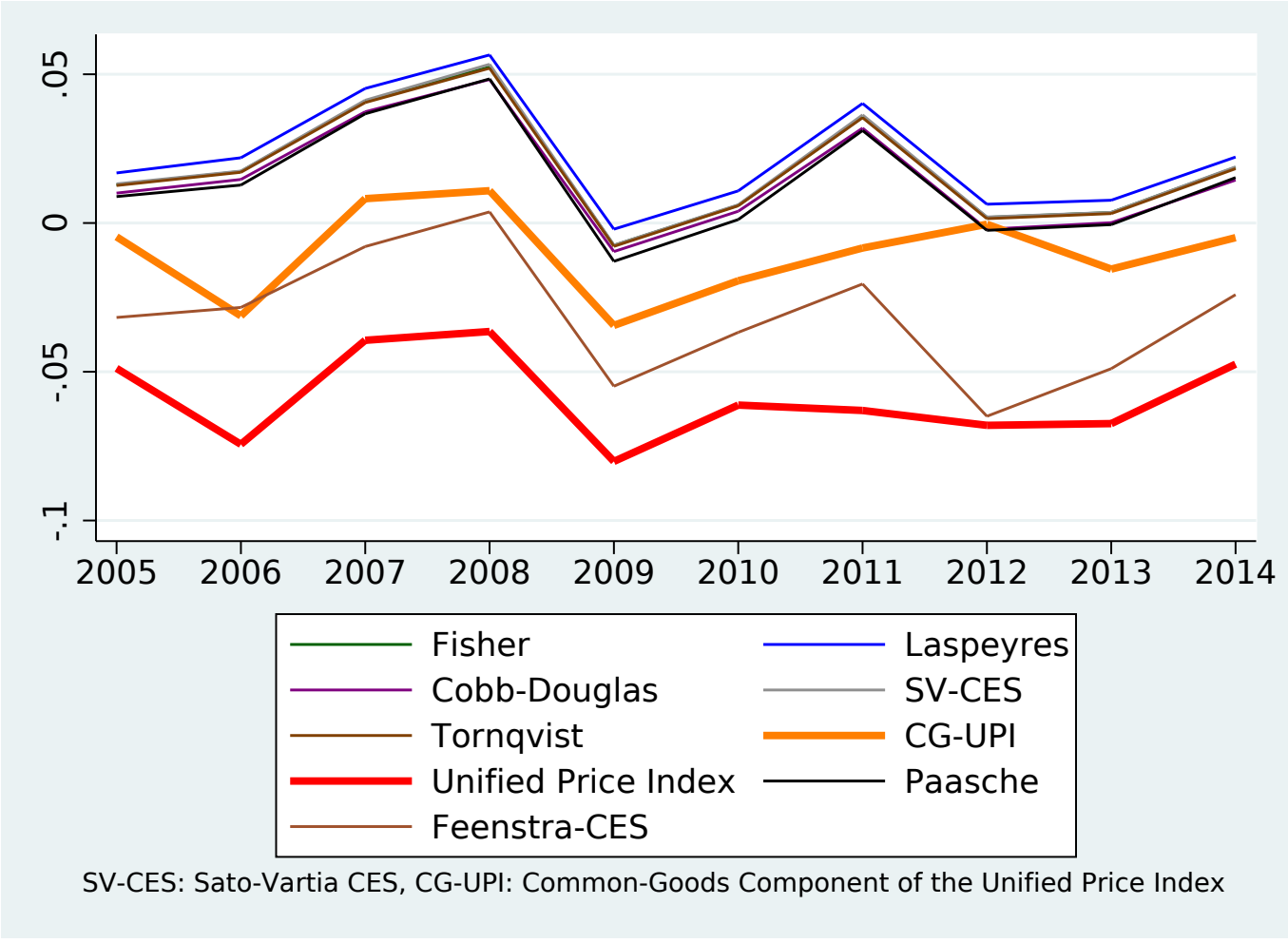
- Approximately 55,000 households scan in every purchase of a good with a barcode
- Observe price paid (including coupons) and total quantity purchased in common physical units (e.g. volume, weight, area, etc.) by UPC
- Around 670,000 different Universal Product Codes (barcodes) sold in each quarter, aggregated into 87 product groups
  - Largest four are carbonated beverages, pet food, paper products, bread
- We aggregate to the national level for Q4 (2004-14) using nationally representative household weights from Nielsen to measure average price per UPC and total quantity sold.

Importance of Product Turnover



One indicates new goods are as good as exiting goods.
Zero indicates that no one wants to buy pre-existing goods.

Aggregate Price Indexes



SV-CES: Sato-Vartia CES, CG-UPI: Common-Goods Component of the Unified Price Index

Between 2004-14, cost-of-living increases were much lower and productivity growth was much higher than is being measured by conventional methods

Sectors Amenable to This Approach

- Items in red represent projects in process
  - Durable Goods: Automotive data available from car manufacturers, Furniture and other data available from store databases
  - Non-durables: Food/Packaged Goods (Nielsen), Clothing (Internet Retailers),
  - Transportation and Hotel (Expedia, Travelocity),
  - Housing (Zillow, Trulia, Real Estate Records)
  - Insurance transactions (online)
  - Retail Productivity: data from Census and scanner transactions
  - Import/Export data from Census transactions

Potential for BLS, BEA, and Census

- National and regional measures of cost-of-living indexes
- Productivity, real output, and innovation by sector
- Cost-of-living changes by income class
  - Improved measures of poverty and income inequality
- High-frequency price and output indexes: daily measures of inflation/output changes
- Customizable cost-of-living measurement: allow people to pick the assumptions they like (e.g., product substitutability, existence of new goods, and existence of demand shocks) when measuring inflation and real output