Progress on Adopting Big Data in the US Consumer Price Index

Brendan Williams

FCSM 2024

23 October 2024

This presentation reports on the results of ongoing research and analysis undertaken by Bureau of Labor Statistics staff. It has undergone more limited review than official publications.

Overview

- Background on Alternative Data
- Update on Alternative Data and the CPI
- Alternative Data System
  - Allows flexible index estimation
  - Combines survey & alternative data
- Ongoing Developments
- Conclusion

Background on Alternative Data

Alternative Data and Official Price Indexes

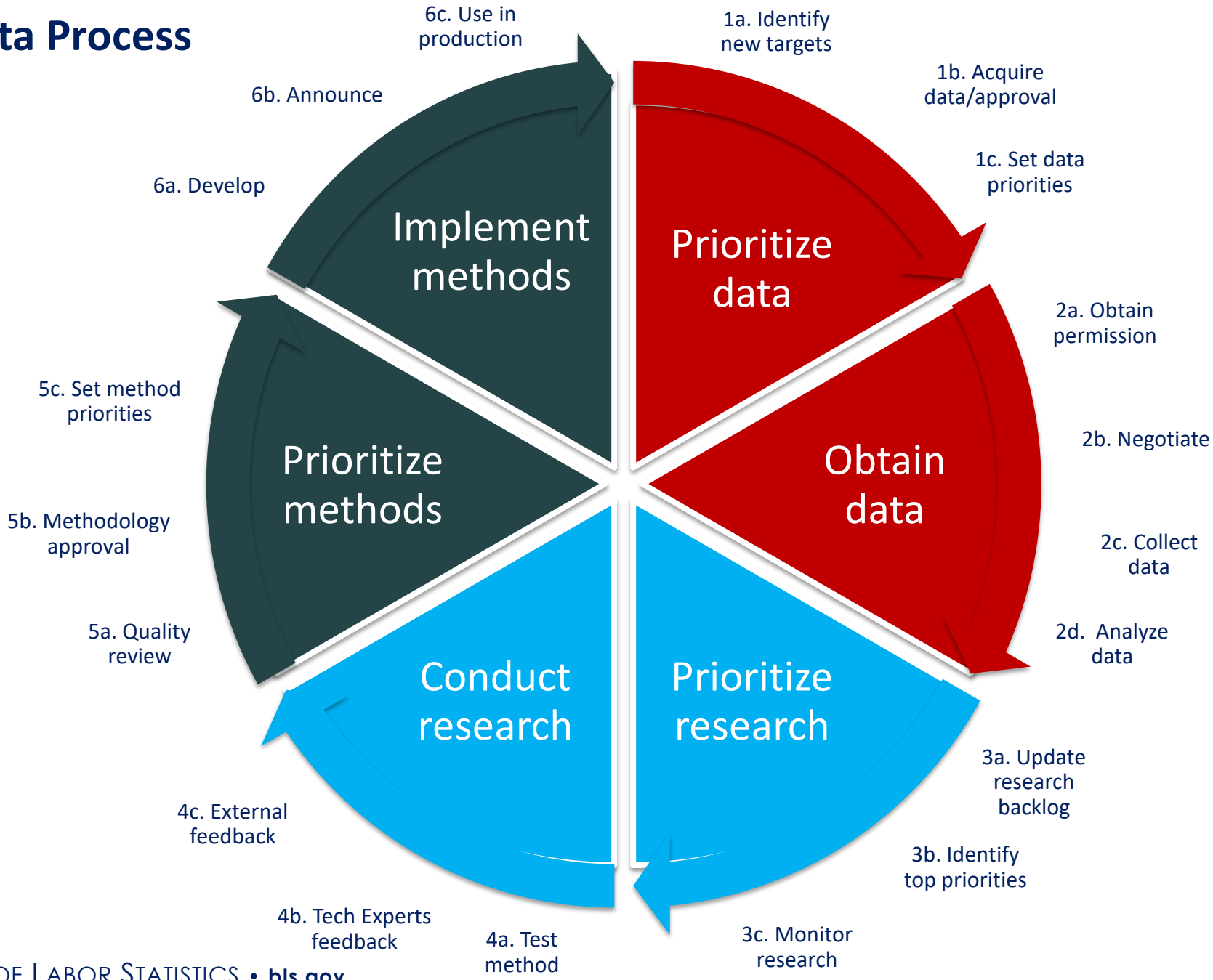
- Most CPIs have been based on survey data or representative products
- Many countries incorporated alternative data
  - Scanner or other transactions data
  - Webscraping
- BLS laid out plans for alternative data in Konny, Williams, Friedman (2022)
- Methodology development specifically for alternative data

Potential Benefits

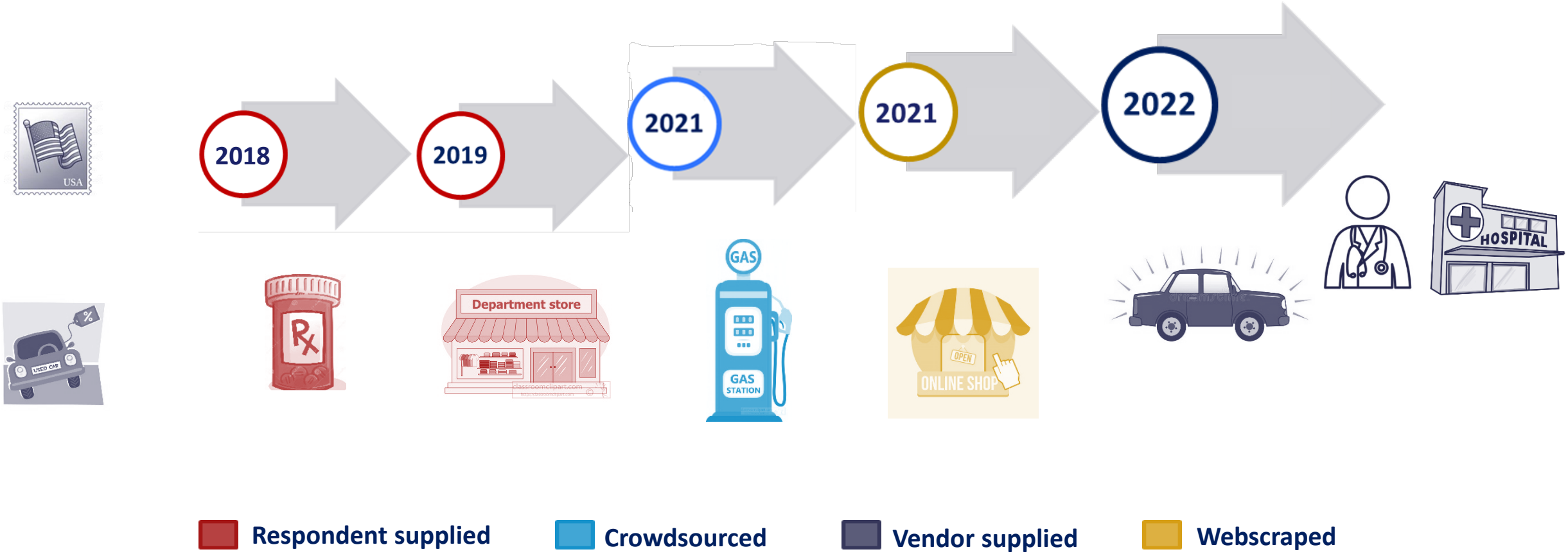
- Lower collection costs
- Reduced respondent burden
- Larger number of observations
- Sales and transactions data
  - Transaction prices
  - Real-time expenditure weighting
    - Better reflect consumer substitution

Update on Project Status

Alternative Data Process



AltData Implementation Timeline



Respondent supplied Crowdsourced Vendor supplied Webscraped

Ongoing Alternative Data Projects

- Medical claims (Scheduled for release of October 2024 data)
  - Bieler, et. al, *Monthly Labor Review* (2023)
- Vehicle leasing
- Wireless telecommunications
- Retailer API
  - Television, Laptops, Smartphones
- Diesel

Previous Approach

- Current systems have either replaced price quotes (CorpX, CorpY, Webscraping) or replaced an entire basic index (New vehicles & Gasoline)
- Different data structures (e.g. daily list prices compared to transaction records)
- Each data source required a dedicated software development project
- Little reuse of code between projects

Existing, Survey-Based System

- Centers on a “price quote”: a single price observation that represents price change for an item category
  - Fixed weight for four-year life of a sample
- “Price quotes” aggregate into “basic indexes” at the item-area level using geometric mean or Laspeyres-type index formulas
- Basic indexes aggregate into upper-level indexes
  - Laspeyres with annual weights for the CPI-U
  - Törnqvist with revised monthly weights for the Chained CPI-U

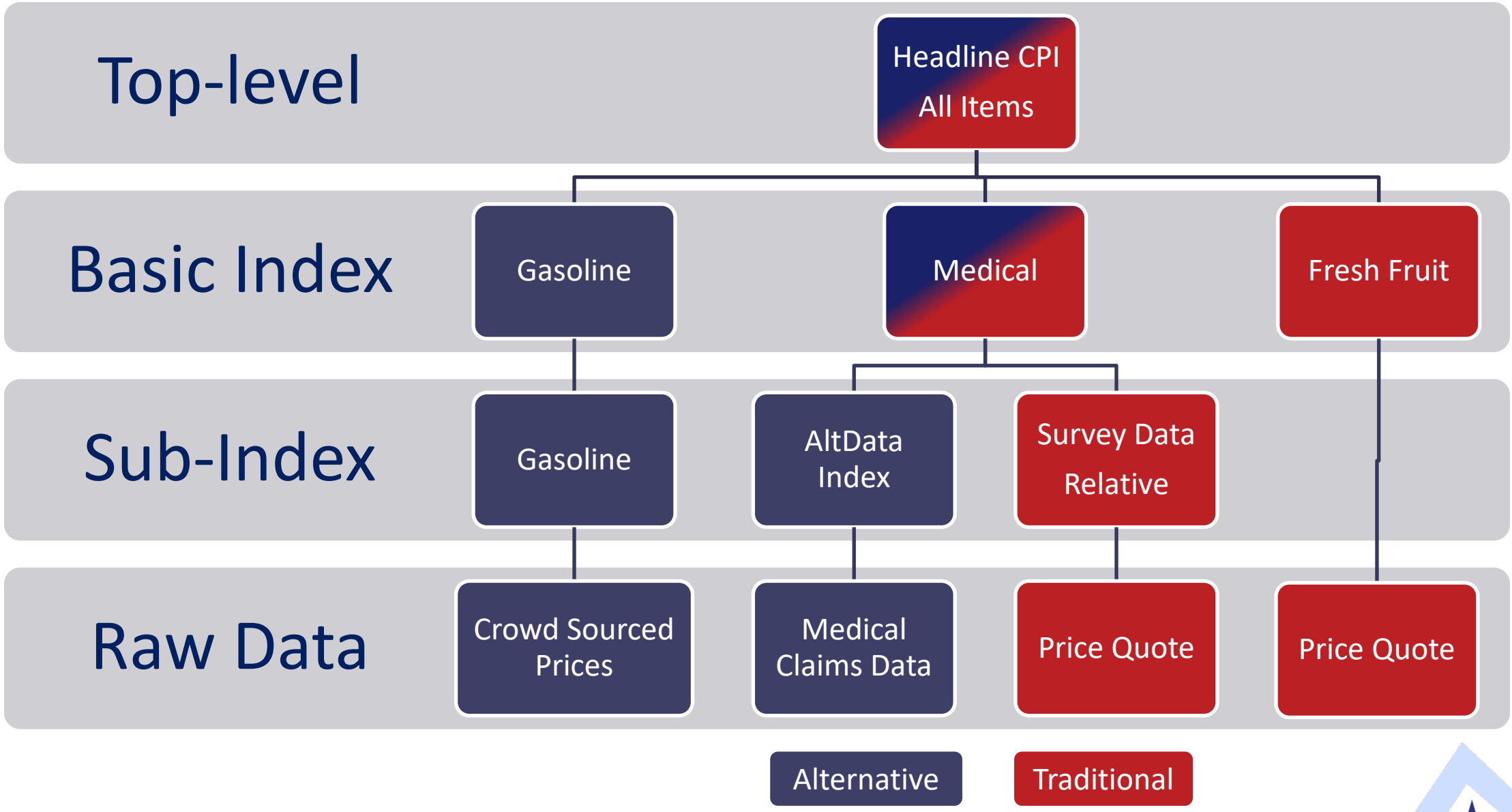
Alternative Data System

Alternative Data System Overview

- Allow generalized, flexible approach to calculate estimated price change
- Standardized but flexible input data format including
  - Product, store, and seller identifiers
  - Item categorization and geographic labels
  - Price, quantity, expenditure, and transaction count
- Sub-index: A price index that makes up a portion of a basic index
- Handle tax rates and replicate assignment for variance estimation

Creating a Sub-Index

- Subset the data based on eligibility and scope
- Define product and “by variables” or output level
- Specify the calculations used
  - Index formula
  - Outlier removal
  - Imputation
- Map the data to the rest of the CPI
  - Aggregate with survey data to form basic indexes (item and area)


Top-level
Basic Index
Sub-Index
Raw Data
Headline CPI All Items
Gasoline
Medical
Fresh Fruit
Gasoline
AltData Index
Survey Data Relative
Crowd Sourced Prices
Medical Claims Data
Price Quote
Price Quote
Alternative
Traditional

Aggregation of Survey and Alternative Data

- Geometric

$$Rel_{Basic,Geo} = \left(Rel_{Area,Item,C\&S}{}^{w_{C\&S}} \times \prod Rel_{Area,Item,Alt}{}^{w_{Alt,s}} \right)^{1/\left(w_{C\&S}+\sum w_{Alt,s}\right)}$$

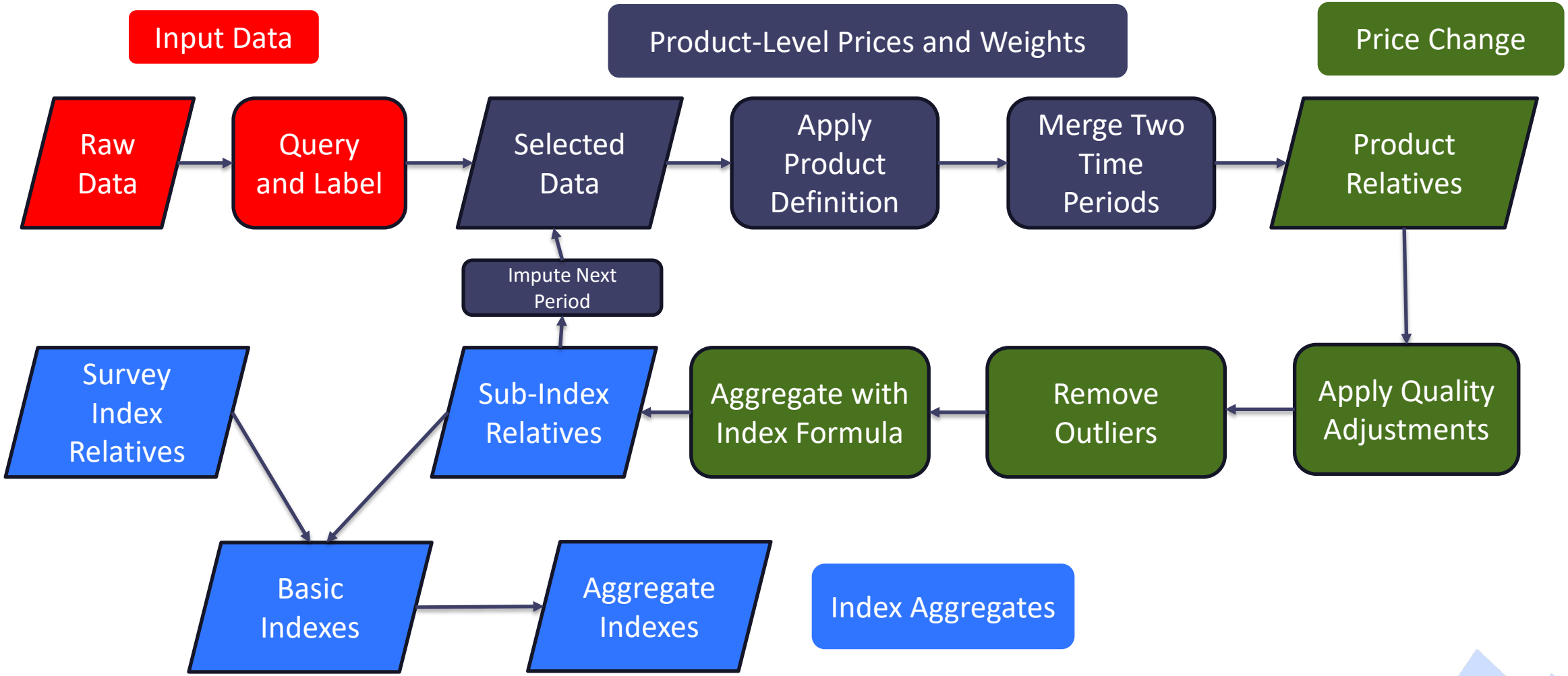
- Arithmetic

$$Rel_{Basic,Arith} = \frac{Rel_{Area,Item,C\&S} \times w_{C\&S} + \sum Rel_{Area,Item,Alt} \times w_{Alt,s}}{w_{C\&S} + \sum w_{Alt,s}}$$

Sub-Weights

- Share of expenditure on a sub-index for a given time period
- Medical Claims will use data sourced from the Medical Expenditure Panel Survey (MEPS)
- Weights can vary by population
  - Allow separate weighting for the U, W, and E populations

AltData System Flow

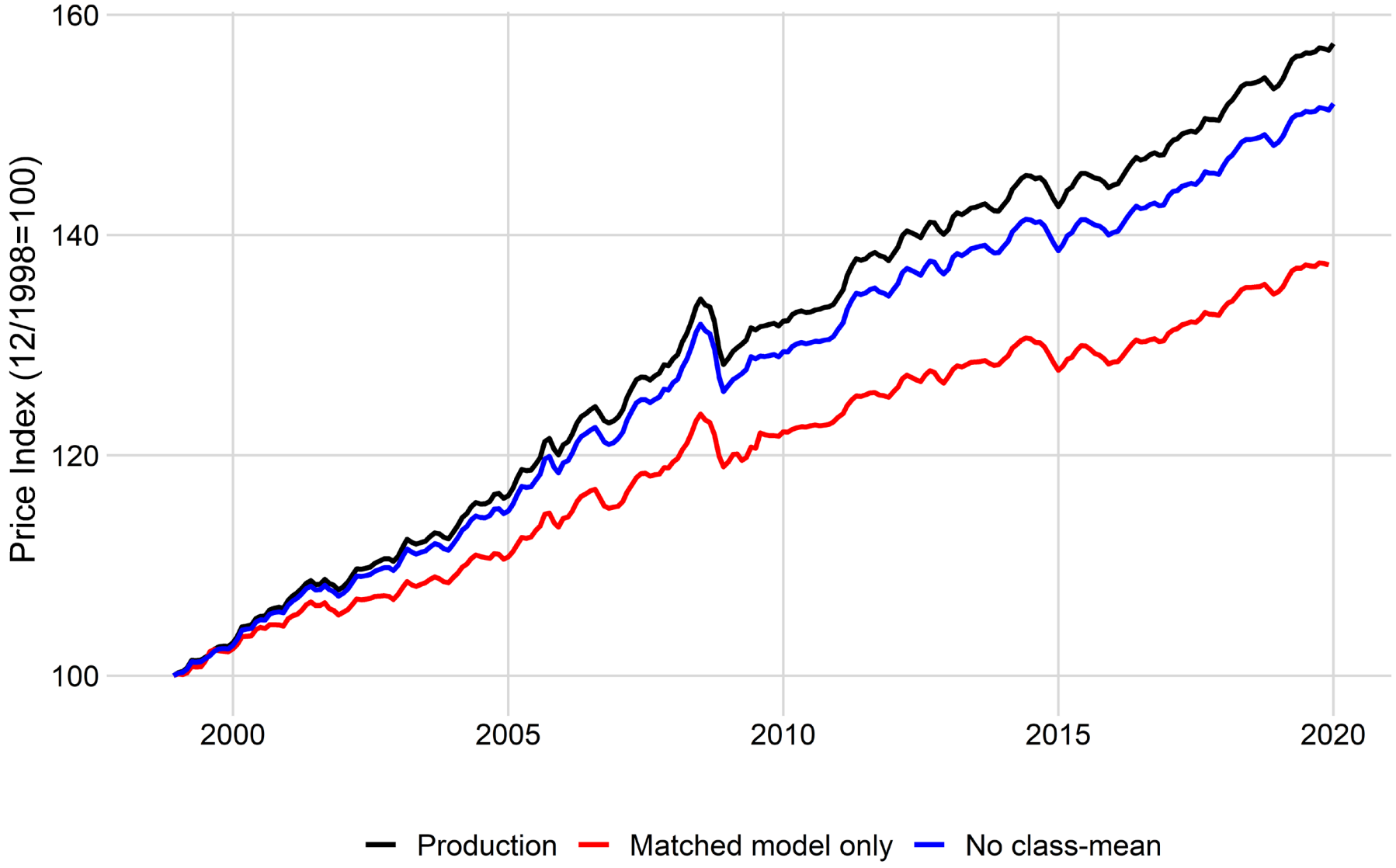


Ongoing Developments

Methodologies

- Product turnover central to many methodological issues
  - Price change is often associated with product change
- In the fixed-sample CPI, product turnover is one-to-one based on collection methods, and weights are static
- Alternative data shows reality of a constant flow of products and fluctuation of weights
  - Product matching and grouping can be ambiguous
- Hedonic imputation indexes

Item Replacement Effects in the CPI

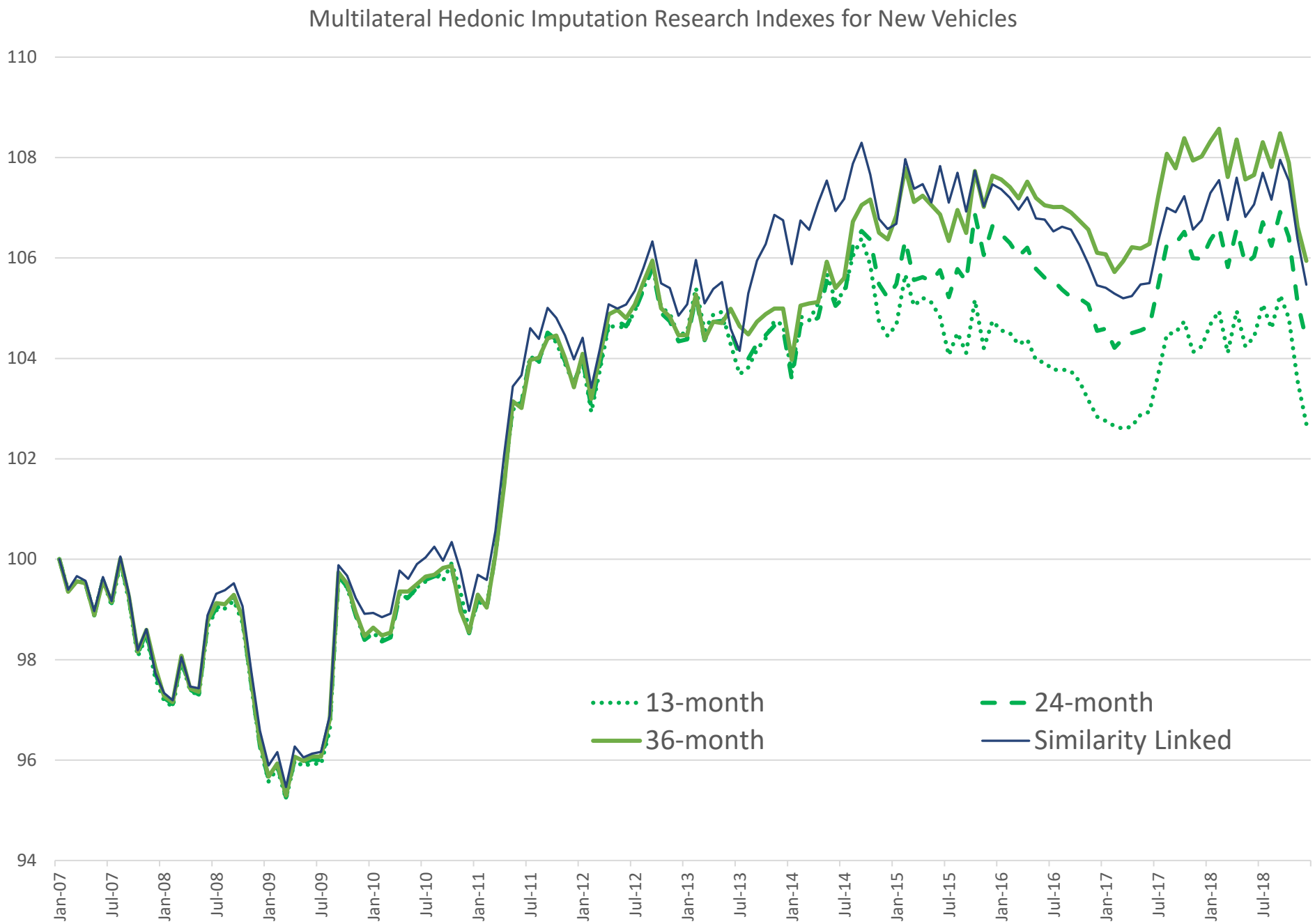


Methodologies (Continued)

- Product turnover central to many methodological issues
  - Price change is often associated with product change
- In the fixed-sample CPI, product turnover is one-to-one based on collection methods, and weights are static
- Alternative data shows reality of a constant flow of products and fluctuation of weights
  - Product matching and grouping can be ambiguous
- Hedonic imputation indexes

Hedonic Imputation

- Allows product change without product matching or grouping
- Brown and Smucker (2024)
  - Wireless telecom service
  - Television prices from national retailer
    - Incorporates adjustment for unobserved characteristics
- Combine hedonic imputation with multilateral indexes
  - Longer windows for multilateral indexes show less of a drop due to capturing more product turnover


Multilateral Hedonic Imputation Research Indexes for New Vehicles
110
108
106
104
102
100
98
96
94
Jan-07
Jul-07
Jan-08
Jul-08
Jan-09
Jul-09
Jan-10
Jul-10
Jan-11
Jul-11
Jan-12
Jul-12
Jan-13
Jul-13
Jan-14
Jul-14
Jan-15
Jul-15
Jan-16
Jul-16
Jan-17
Jul-17
Jan-18
Jul-18
13-month
24-month
36-month
Similarity Linked

Summary

- Incorporated several new sources of alternative data
- Adapting to alternative data with a new system gives us flexibility for new data sources and methods
- Researching methodological and other changes as we shift from a survey-based index number to a mixture of data sources

Contact Information

Brendan Williams
Senior Economist
Office of Prices and Living Conditions/CPI
www.bls.gov/cpi
williams.brendan@bls.gov