

Quality Considerations for Alternative Data in the BLS Producer and Consumer Price Indexes

Crystal Konny (CPI)
Bonnie Murphy (PPI)



“Alternative Data”

- Administrative/Publicly available data
- Purchased data sets
- Company provided data
- Web scraping/application program interface (API)

QUALITY

Quality metrics are similar
regardless of the type of
alternative data

Data Uses

- Create sample frames
- Benchmark sample
- Supplement collected data to support hedonic modeling
- Replace/supplement current data collection methods
- Validation of survey data

Dimensions of Data Quality

- Accuracy
- Timeliness
- Reliability
- Coherence/Comparability
- Accessibility and Relevance
- Cost

Accuracy Considerations

- Transparency
- Sample versus universe
- Quality control
- Concordance (*next workshop*)
 - Mapping items into existing structures
 - Blending with existing survey data

Data Accuracy – Coverage Error

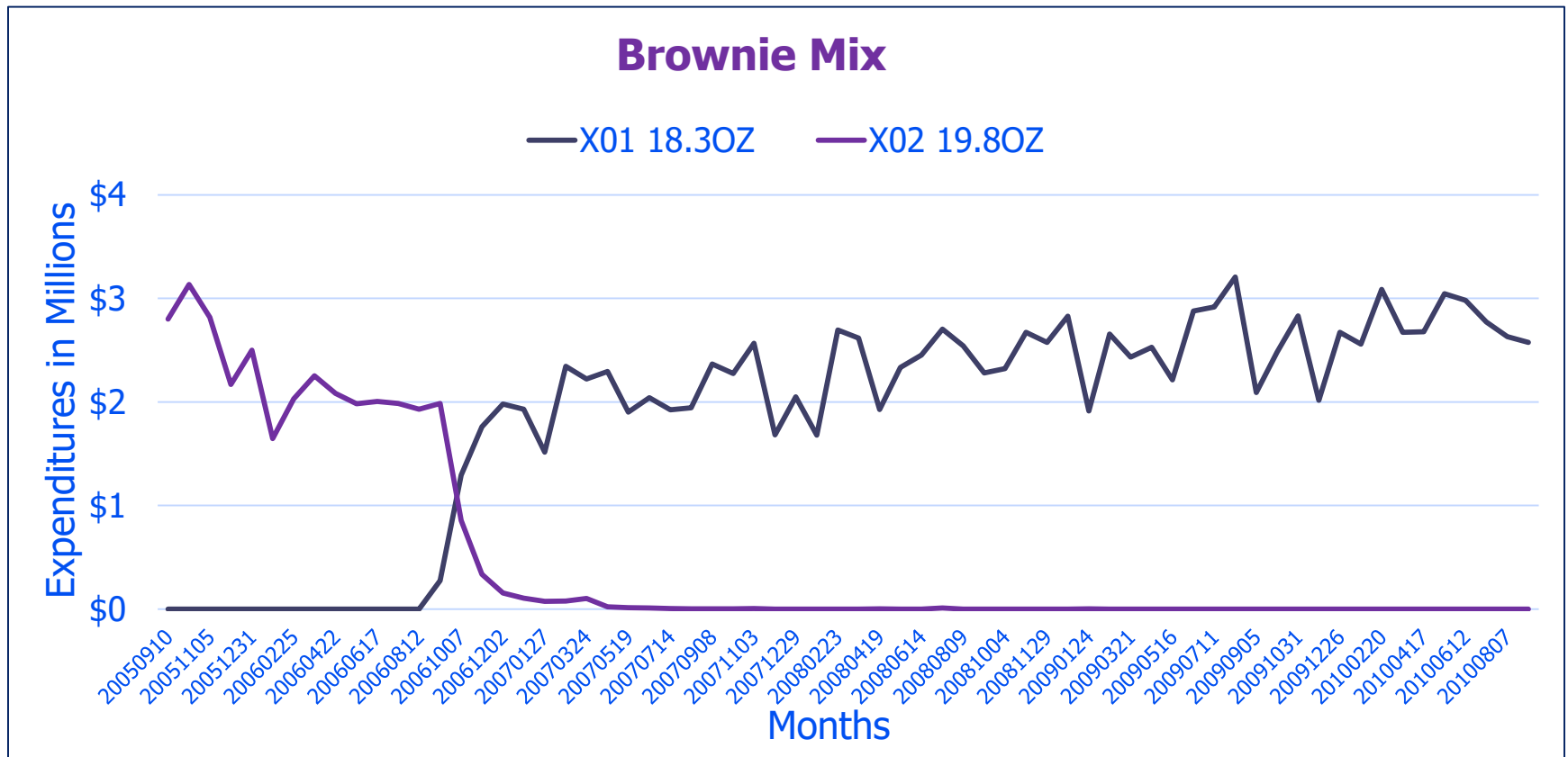
- Time period/history
- Geographic (different for CPI and PPI)
- Outlet/establishment
- Item

Data Quality at the Item Level

- Characteristic detail
- Price level of detail



UPC Example



Timeliness

- Timing of data procurement
- Timing of data delivery - control of schedule



- Limits on long term contracts

Reliability

- Will it exist going forward?
- Condition of data/ 'cleaning' or validating needed?
- Data security



Coherence/Comparability

- Methodological constructs
 - Fit into existing?
 - New one needed?
- Comparability of data over time



Accessibility and Relevance

■ Accessibility

- ▶ Resulting indexes will continue to be accessible to users
- ▶ Microdata
 - Survey microdata is accessible to “sworn agents” of BLS
 - Alternative source may not allow sworn agent access

■ Relevance – as long as our publication structures remain relevant, alternative data would not change their relevance

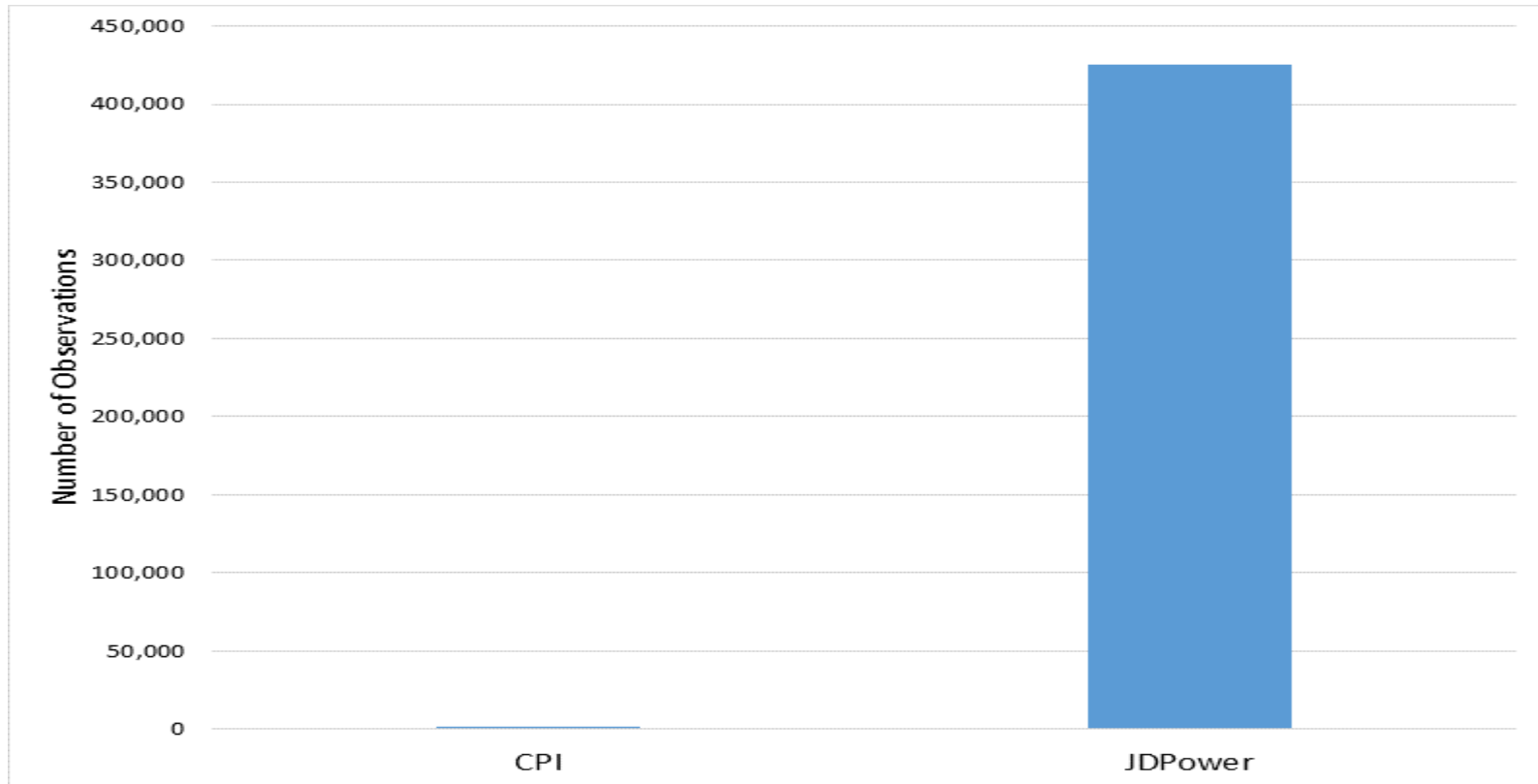
Cost

- Cost of data
- Cost of IT infrastructure
- Cost of processing changes to use data
- Incorporation into existing BLS structure
- Purchasing process- huge cost of switching vendors
- Staff – maintain skill set to manage/massage data
- Fall back plan

Benefits of Alternative Data

- Becoming increasingly more available
- Allows for evaluation & improvements
- May reduce collection costs
- Reduces respondent burden
- Increases sample size
- May increase data quality
- May increase timeliness of reporting

Sample Size

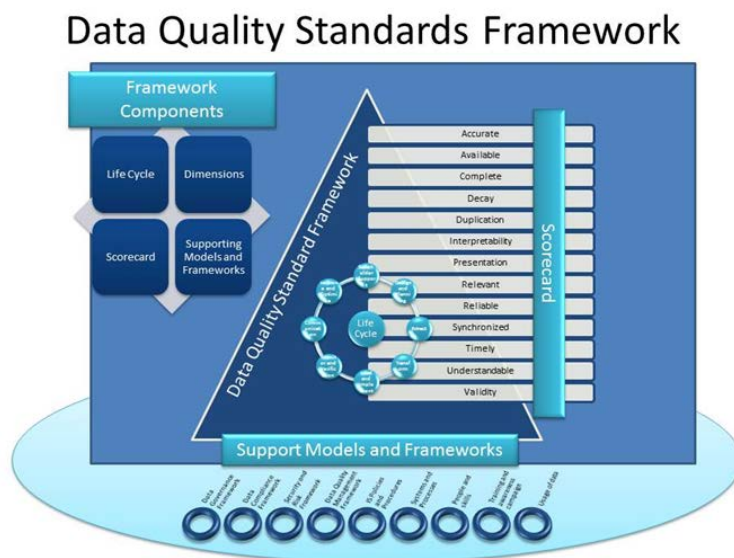


Current Index Quality Metrics

- Variance/standard errors for current price indexes
- Response rates
- Non-response bias studies (if necessary)
- Timely release of monthly data
- Size of Revisions (PPI)

Alternative Data Matrix

- Includes a cost–benefit analysis
- The relative importance of each quality metric may change depending on the intended use of the data



Alternative Data Matrix

Quality Metrics	Sample Frames	Benchmarking	Hedonics	Replace collection	Supplement Collection	Data Validation
Data Transparency- methods understood						
Granularity- Level of detail						
Quality of descriptive data						
Scope, type of price						
Coverage- items						
Coverage- geography						
Coverage- outlets						
Sampling procedures						
Data delivery reliable						
Viability of data source						
Data Usability						
Data Frequency						
Data Security						
Data delivery timeliness						
Data history						
Data Cleanliness						
Data Usability- mods to current system						

Alternative Data Matrix

Quality Metrics-- Cost Benefit Analysis

Dollar cost of data

BLS cost of using data

Reduced respondent burden?

Dollar savings if no data collection

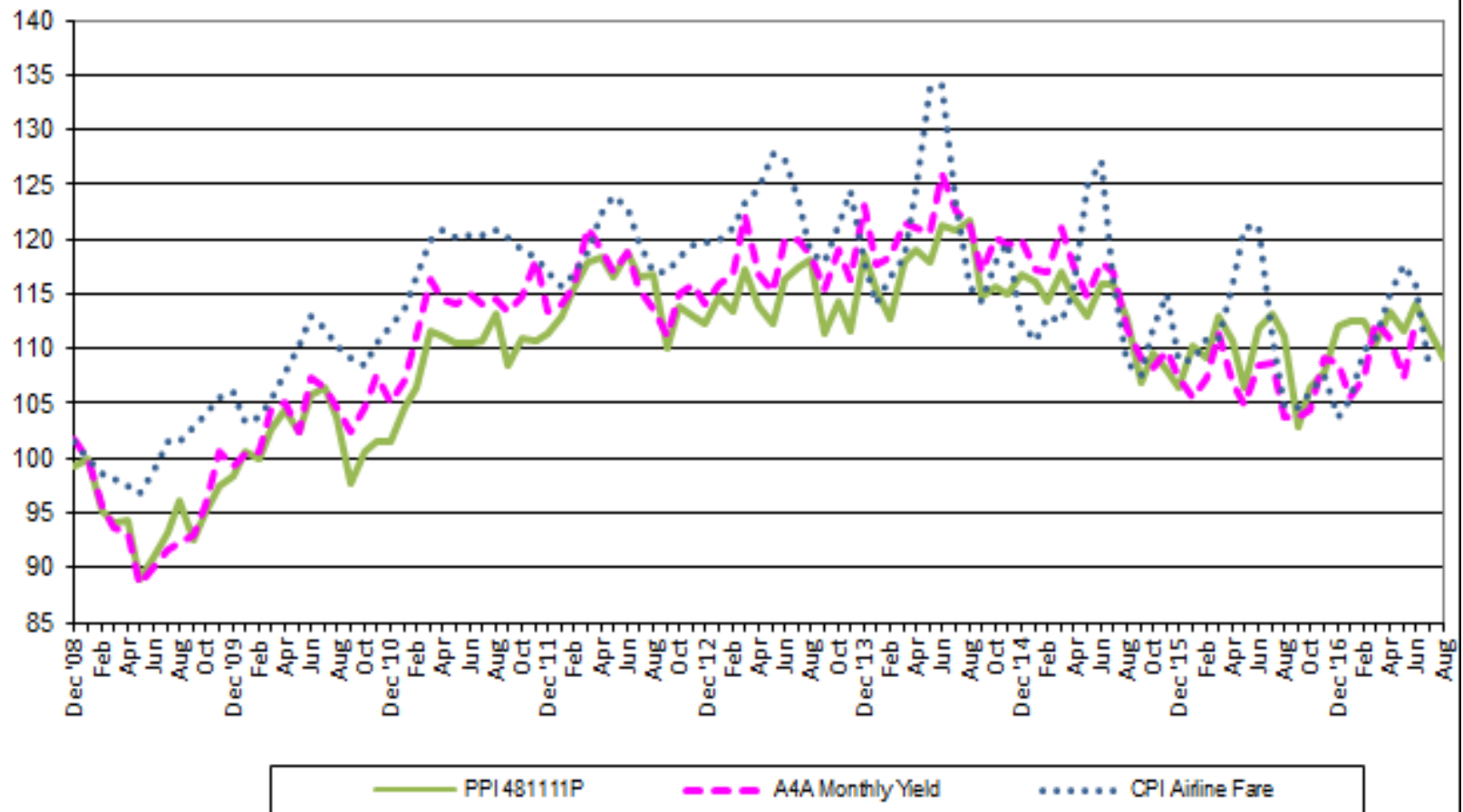
Increased Sample Size?

Improved timeliness?

Data quality improvement?

Specific issue/problem alt data will address

PPI, A4A and CPI Airline Indexes



Contact Information

Bonnie Murphy

Chief, Branch of Industry Pricing

Producer Price Index Program

Murphy.bonnie@bls.gov

Crystal Konny

Chief, Branch of Consumer Prices

Division of Consumer Prices and Price Indexes

Konny.crystal@bls.gov

