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 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

Brownie Mix

- X01 18.3OZ
- X02 19.8OZ

Expenditures in Millions

Months

Expenditures in Millions

- $0
- $1
- $2
- $3
- $4

- 20050930
- 20051105
- 20051231
- 20060107
- 20060422
- 20060617
- 20060812
- 20061007
- 20070127
- 20070342
- 20070519
- 20070714
- 20070908
- 20071103
- 20080229
- 20080461
- 20080609
- 20080814
- 20081004
- 20090129
- 20090321
- 20090516
- 20090711
- 20090905
- 20091126
- 20100120
- 20100317
- 20100612
- 20100807
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

<table>
<thead>
<tr>
<th>Quality Metrics</th>
<th>Sample Frames</th>
<th>Benchmarking</th>
<th>Hedonics</th>
<th>Replace collection</th>
<th>Supplement Collection</th>
<th>Data Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Transparency- methods understood</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Granularity- Level of detail</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of descriptive data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scope, type of price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage- items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage- geography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage- outlets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data delivery reliable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Viability of data source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Usability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data delivery timeliness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data history</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Cleanliness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Usability- mods to current system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Alternative Data Matrix

<table>
<thead>
<tr>
<th>Quality Metrics--</th>
<th>Cost Benefit Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dollar cost of data</td>
<td></td>
</tr>
<tr>
<td>BLS cost of using data</td>
<td></td>
</tr>
<tr>
<td>Reduced respondent burden?</td>
<td></td>
</tr>
<tr>
<td>Dollar savings if no data collection</td>
<td></td>
</tr>
<tr>
<td>Increased Sample Size?</td>
<td></td>
</tr>
<tr>
<td>Improved timeliness?</td>
<td></td>
</tr>
<tr>
<td>Data quality improvement?</td>
<td></td>
</tr>
<tr>
<td>Specific issue/problem alt data will address</td>
<td></td>
</tr>
</tbody>
</table>
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