Communicating the accuracy and relevance of statistical data during a pandemic

Michael Levi and Emily Liddel
U.S. Bureau of Labor Statistics

IMF Statistical Forum on Measuring the Economics of a Pandemic
17-19 November 2020
Trust and transparency

- BLS depends on the trust of data users (as do all stat agencies)
  - If customers do not believe our data are impartial and accurate, they will find other sources
- Trust is established and maintained through transparency
  - Announcing changes in methods
  - Announcing changes in operations
  - Alerting users to potential measurement anomalies
- COVID-19 exercised the BLS commitment to transparency

- BLS and Census cease all in-person data collection
- Telephone data collection centers close
- Staff are moved to maximum telework status
Impact: Collection modes

- In many, but not all, BLS surveys, the mix of data collection modes changed
- Ex: Consumer Price Index

<table>
<thead>
<tr>
<th>Mode</th>
<th>April 2019</th>
<th>April 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>In person</td>
<td>74%</td>
<td>0%</td>
</tr>
<tr>
<td>Telephone</td>
<td>10%</td>
<td>18%</td>
</tr>
<tr>
<td>Online</td>
<td>16%</td>
<td>82%</td>
</tr>
</tbody>
</table>
Response rates changed in many BLS surveys, but not uniformly across industries, demographic groups, or geographic areas

Ex: Producer Price Index

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mining, utilities, and construction</td>
<td>66</td>
<td>65</td>
<td>71</td>
</tr>
<tr>
<td>Retail trade</td>
<td>75</td>
<td>77</td>
<td>76</td>
</tr>
<tr>
<td>Entertainment, accommodation and food services</td>
<td>66</td>
<td>65</td>
<td>47</td>
</tr>
</tbody>
</table>
Impact: Methods

- Seasonal adjustment
- Birth-death model
Impact: Data sources

- Ex: Productivity program
  - Continuing Unemployment Claims
Impact: New data

- Additional questions on household and establishment surveys
- New survey (Business Response Survey)
Are the data still reliable?

- “Everything’s shut down”
- “Nothing’s in stock”
- “You’re not collecting the information”

BLS says, “All published data meet reliability standards”

How can users assess for themselves?
Communicating uncertainty

- Difficult under the best of circumstances
  - Confidence intervals
  - Revisions
  - Changes in methods

- Communicating a change in uncertainty is even harder
Starting mid-April 2020

- Calls pour in to BLS wanting information on
  - Data collection
  - Methods
  - Concepts and definitions
We had (most of) the answers

- For methods, go to the *Handbook of Methods*
- For changes to methods, see the announcement box in our release
- For response rates, go to a response rate page (9 months out of date)
- For modes of data collection, see link buried at the bottom of a web page
Early efforts

- We didn’t know
  - What paradata users wanted
  - What paradata we had readily available
  - How long the pandemic disruption would last

- Went through several iterations of Q&As, paradata reporting
Eventually: A common framework

- One measure is not enough
- What is meaningful for one program may not be so for another
- Value of consistency across programs
- Needs to be easy to update
Paradata framework

- Collection mode(s) and collection rates
- Imputation rates
- Cell or series suppression
- Statistical measures of error
- Changes to methods
- Q&As

- All with comparison to pre-pandemic measures
Release page

- Links to impact statements for each release
- “Evergreen” Q&A
Impact statements

Impact of the coronavirus (COVID-19) pandemic on the Producer Price Index data for July 2020

The Producer Price Index (PPI) is a measure of the average change over time in the prices domestic producers receive for the sale of their products. PPI estimates for July 2020 were published on August 11, 2020. The pricing date was Tuesday, July 14, 2020, which means that most producers, but not all, provided prices they received on that date. The summary statistics in tables 1, 2, 3 and 4 indicate the impact of the COVID-19 pandemic on the PPI data collection, index estimation, and the BLS ability to publish indexes was minor for the release of preliminary July data. No changes in estimation procedures were necessary.

This page provides further information on the effects of the COVID-19 pandemic for the July 2020 Producer Price Indexes news release.

Collection mode
The PPI collects nearly all of its repricing data by web collection. There were no changes to collection mode in response to the pandemic.

Response rates
Table 1 provides the July 2020 response rate by industry sector, along with comparison periods of June 2020, July 2019, and a 12-month average for March 2019 to February 2020. The July 2020 response rate for items requested shows a similar level of response for the vast majority of industry sectors. While the response rate for entertainment, accommodation, and food services sector remains down when compared to historical comparison periods, both the July and June 2020 response rates reflect an uptick in response as businesses begin to reopen.

It is important to note that typically, the PPI survey’s monthly price collection response rate averages 76 percent by the end of its 4-month index revision period. All PPI Indexes are recalculated 4 months after publication of preliminary data. Thus, July indexes may be revised when BLS publishes the PPI for November on December 11, 2020. All summary statistics in the tables below reflect comparisons to PPI preliminary data.

<table>
<thead>
<tr>
<th>Industry sectors</th>
<th>12-month average, March 2019 to February 2020</th>
<th>July 2020</th>
<th>June 2020</th>
<th>July 2020</th>
<th>Percentage point change from June 2020 to July 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Industrial sectors</td>
<td>69</td>
<td>71</td>
<td>72</td>
<td>73</td>
<td>+1</td>
</tr>
<tr>
<td>Agriculture, forestry, fishing and</td>
<td>47</td>
<td>50</td>
<td>48</td>
<td>47</td>
<td>-1</td>
</tr>
<tr>
<td>handling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mining, utilities, and construction</td>
<td>66</td>
<td>63</td>
<td>66</td>
<td>70</td>
<td>+4</td>
</tr>
</tbody>
</table>
Prominent misclassification error

- Operational paradata subtleties swamped by misclassification error in household survey
- Accuracy and impartiality of BLS data called into question
- BLS announced and described issue in every Employment Situation news releases
- Provided detailed information in supplemental documents
  - Including estimates of what the data might have shown without misclassification
- Published an explainer blog
Some positive reactions from website feedback


👍 I sense you’re on the bleeding edge of the data that’s going to point us out of the COVID mess.


👍 very helpful, thank you


👍 Extremely helpful. Thanks for keeping this marvelous data update..Great Job!!


👍 So happy to see the BLS moving so quickly to incorporate COVID questions to the CPS, and for making the information available as quickly as possible.
Critical reactions from website feedback


The title of this page is "Effects of COVID-19 Pandemic and Response on the CPI". I was expecting a simple answer such as "The CPI change was 50% lower than had been predicted in January". Instead there are details on the methodology. Frustrating


This says it all. Data inadequate
User reactions (quantitative)

- As of September 30, 2020:
  - 166,067 user views
  - 294 ratings
  - 138 comments
- Many of the ratings were unrelated to the pages themselves
- Once unrelated ratings are eliminated
  - 84.2 percent positive feedback
Final thoughts

- Users need reliable information about the accuracy and relevance of the estimates upon which they rely
- No single paradata measure is sufficient
- Framework: Spectrum of meaningful measures
- Flexibility: As conditions change, other measures may become more relevant
Contact Information

Michael Levi
U.S. Bureau of Labor Statistics
1-202-691-5100
levi.michael@bls.gov