

Alternative Sources of Price Data and Current Uses

James Thomas

Chief, Branch of Index Methods, Analysis, and Evaluation

International Price Program

U.S. Bureau Of Labor Statistics

Thomas.jim@bls.gov

Main reasons for alternative data source usage in the Statistical Community

- Faster, more timely statistics
- Reduction of respondent burden
- Modernization of the statistical production process
- New products and services (e.g. map development, visualizations)
- Cost reduction
- Meeting new demands such as Sustainable Development Goal Indicators

Source: United Nations Global Working Group on Big Data for Official Statistics

U.S. Statistical Stakeholders

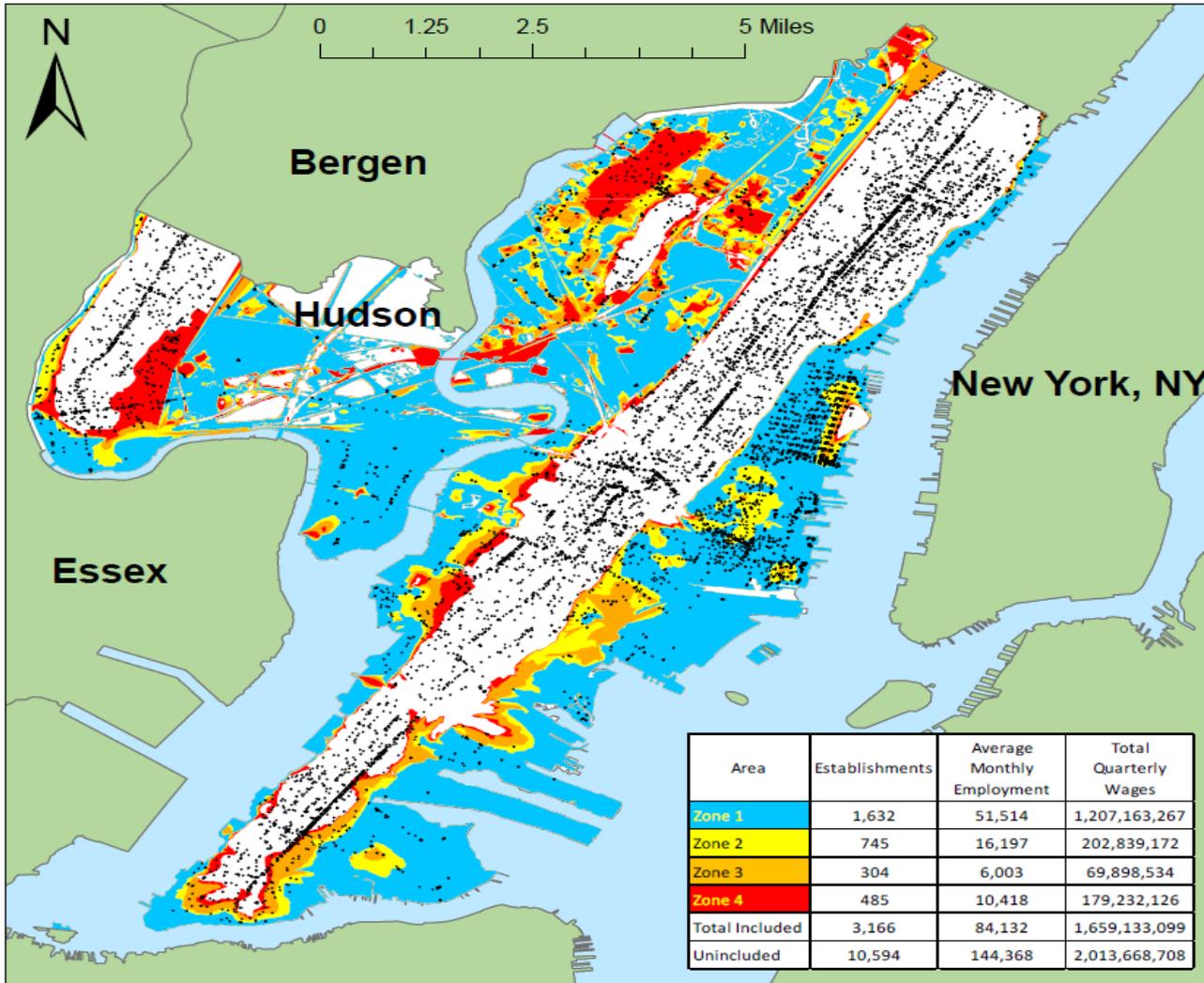
- In order to incorporate alternative data sources, BLS has fostered partnering arrangements with government agencies, private providers of data, research and academic communities.
- Examples include big data working sessions with Federal Economic Statistics Advisory Council (FESAC), Agency Advisory councils, and US Committee on National Statistics (CNSTAT).

Government Administrative Data Uses to Produce Price Indexes

- Sampling and Initiation: Source for sample frames used for drawing stratified probability samples to initiate companies and items for price statistics (e.g. IPP uses Customs Trade data for sampling).
- Re-pricing: Source of price estimates for index calculation in lieu of directly collected prices. (eg IPP EIA data on crude petroleum, PPI DOT info on baggage fees, PPI CMS for hospital and physician treatments).
- Weights: Source of weights for calculating price indexes. (IPP uses Census data to annually update export and import price index weights)
- Publication: Linking data series between agencies. (BLS price indexes used by BEA to estimate real GDP).

Government Administrative Data

Employment in Hurricane Storm Surge Flood Zones, Hudson County, NJ



Note: Flood zones represent a conservative estimation of areas that would experience flooding in the event of a hurricane. Zone 1 is the area that would be flooded by a Category I hurricane. A Category II hurricane would cause flooding in Zones 1 and 2. The hurricane categories reference the Saffir-Simpson Hurricane Scale.

Black dots represent establishments.

Data Source: U.S. Bureau of Labor Statistics, Quarterly Census of Employment and Wages; 2012 Third Quarter
<http://www.bls.gov/cew>

Flood Zone Source: National Hurricane Program.



Using Private Administrative Prices Data

- Sampling: JD Power used car frame for the CPI.
- Initiation: Scanner data used to compare distribution of selected products with results from CPI disaggregation process.
- Re-pricing: Electronic data collection for PPI and XMPI.
- Estimation: Stock Exchange Security Trades used by PPI for index estimation, medical expenditure panel survey household health claims data used to construct CPI experimental disease based price indexes.

Uses of web scraped data to calculate price measures

- Experimental Price Indexes (e.g. CPI used and new book indexes).
- Create data base of product characteristics for quality adjustment hedonic regression models for CPI (e.g. televisions, camcorders, cameras, and washing machines).
- Examining promising areas for Consumer Price Indexes and/or use in the International Comparison Program (food, cable TV, airline, and courier services, etc.).

Methodology Questions when Blending Price Data

- Understanding if alternative data is representative of the target population under study.
- Calculating weights for the construction of indexes.
- Determining when to use alternative price data based on goodness of fit and future availability.

Future Challenges for using Alternative Price Data Sources

- Enhancing data access, clarification on permission and informed consent.
- Providing transparent methodology for Price Index estimation.
- Establishing a Quality Framework for blending alternative data sources into price indexes.

Future Advantages of Using Alternative Price Data Sources

- Symbiotic relationship between official statistics and alternative data source providers (calibrate and validate).
- Speed of processing (Hadoop, simulation models, etc.).
- New uses filling in price gaps. [imputing or estimating missing prices for price time series and International Comparison Program (ICP) interspatial comparisons].
- Estimating weights at lower levels (CE uses EIA info on purchases of regular, mid, and premium gas to assign lower level proportions, ICP could possibly use data to disaggregate expenditures to unpublished lower levels)
- New and timely outputs.

THANK YOU
