

BLSTAC - Future Priorities discussion, June 17, 2016

Disease Based Price Indexes (OPLC)

Federal statistical agencies are being encouraged to publish medical statistics by disease rather than by goods and services. This means that medical expenditures would be reported by disease (i.e. cancer, heart disease, etc.) rather than by service (i.e. physicians, hospitals, etc.). Disease based price indexes are a necessary tool to deflate nominal disease expenditures into real disease expenditures. Many issues arise when deciding how to construct these disease based price indexes. First, we must decide which data set to use. Do we use expensive private insurance claims data that comes from convenience samples or do we use publicly available databases that are free to the public but have much smaller sample sizes. Second, we must decide how to treat comorbidities. For example, if an individual visits a physician to treat both her diabetes and asthma, what fraction of the visit gets allocated to diabetes and what fraction gets allocated to asthma? A third issue involves the choice of a disease classification system. The ICD-9 (International Classification of Diseases) system has too many categories and most items would not be populated. However, a classification system such as the Chapter of the ICD-9 are so broad that each category contains a widely heterogeneous group of patients. Finally, there is the quality adjustment issue. Any medical price index needs to be quality adjusted, but at this point in time, we do not have the necessary outcome database to make these quality adjustments.

Questions:

1. Do you think that it is better to report medical statistics by disease than by service?
2. Do you believe that it worth the additional expense of purchasing the larger private insurance claims databases as a source for computing these disease based price indexes, or can we rely on the Medical Expenditure Panel Survey?

Exploratory Online Book Price Indices (OPLC)

With a few exceptions, the vast reservoir of data available on the web remains largely unexploited by the Bureau of Labor Statistics (BLS). Since August 2013, price quotes have been collected for between 260,000 and 330,000 books per month from a major online retailer of books. In addition to price quotes, sales ranks have been collected which have been shown to be inversely related to quantities sold (e.g., Hackett, 1967; Brynjolfsson et al., 2003, 2012; Chevalier and Goolsbee, 2003). Using inverse powers of sales ranks as a proxy for quantities sold, a variety of price indices were constructed (Carli, geometric mean, Laspeyres, Paasche, Törnqvist) and compared them to a geometric mean index computed using BLS collected, online book prices. The weighted price indices (Laspeyres, Paasche, Törnqvist) all show a significant jump in October 2013 that is unobserved in the BLS data. Two hypotheses for this deviation are: 1) the BLS's small sample of online price quotes and/or 2) the BLS's sample is not sufficiently representative of actual purchases.

References:

- Brynjolfsson, E., Y. J. Hu and M. D. Smith, 2003, "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety," *Management Science*, 49(11):1580–1596.
- , 2012, "The Longer Tail: The Changing Shape of Amazon's Sales Distribution Curve," Working paper.
- Chevalier, J. and A. Goolsbee, 2003, "Measuring prices and price competition online: Amazon and Barnes and Noble," *Quantitative Marketing and Economics*, 1(2):203–222.
- Hackett, A. P., 1967, *70 Years of Best Sellers, 1895–1965*, R. R. Bowker Company, New York.

Questions:

1. Should the BLS increase its proportion of collected online quotes, especially for products where online retail is dominant? 25% seems too small for books. It is expected that the root problem is TPOPS.
2. Should the BLS increase the frequency with which it resamples books in order to maintain representativeness? The average age of the BLS sample is 2006; the "expenditure" weighted age of the webscraped sample is 2010.
3. Would it be useful for the BLS to create a new "national online region" and an associated "regional" price index? The issue is that online prices are generally national and so in principle they should be treated as national quotes.

Current Employment Statistics (CES) Establishment Age/Size Class Estimates (OEUS)

For the Current Employment Statistics (CES) survey, experimental establishment size class estimates are expected to be published in 2015. An earlier iteration of CES firm size class data was released that utilized a base-period sizing for monthly estimates and an end-period resizing for generating a benchmarked series. This benchmarking methodology provided information about change to employment levels from one end-period size classification to the next. However, most interest in size class data is in answering where growth is occurring. For this, a consistent base-period sizing methodology appears to be a better choice. As a result, CES has been researching methods to do just that. The methodology chosen has led to choices about the data that we can publish (changes but no levels). Additional choices include the size classes to use, and the industries to publish.

The methodology chosen for developing estimates of employment change by business size may also be applied to develop data based on business age. This work is in an earlier stage of development.

Questions:

1. Does the Committee have any concerns with or suggestions about the methodology chosen to produce firm-size class estimates for CES?
2. Should BLS pursue specific additional research before considering the firm size class benchmarking methodology final?
3. Does the Committee have any concerns with or suggestions about applying the methodology to produce data based on firm-age class for CES?

Matching establishment data bases at BLS (OEUS – November 2016)

The goal of this session is to describe the various recent establishment ‘matching’ projects we have been undertaking at BLS and to see the advice of the TAC in terms of future directions for matching at BLS. Last year, Associate Commissioner Michael Horrigan began a BLS matching team that brought together researchers and program staff from across the person to explore the potential for matching across various BLS data sets, but in particular the Quarterly Census of Employment and Wages (QCEW). Our initial focus was to simply share information on the matching projects that were underway and to brainstorm on other potential matching projects that could be undertaken. These meetings have served to focus our attention on both the enormous potential benefits and challenges of matching: The potential of matching to create new rich data sets to analyze complex economic issues and create new products as well as the significant challenges that matching can pose such as the amount of time and resources matching can take and the legal barriers to matching across federal data systems. In this session Associate Commissioner Michael Horrigan will provide an brief overview of the work of this matching group and lay out a draft vision for the potential matching in the future at BLS. This will be followed by three presentations: Elizabeth Handwerker will present work on matching the QCEW to the Bureau of Economic Analysis’s data on Foreign Direct Investment; Matt Dey/Dave Piccone/Laurie Salmon will present our plans for matching QCEW to the Occupational Employment Statistics Survey (OES); and Kristen Monaco / Mina Kim will present work we are doing matching the Survey of Occupational Injuries and Illnesses to QCEW and OES.

Questions:

1. What other matching projects should have high priority for BLS?
2. The FDI project will result in the creation of enterprise frames (for a limited number of years). Should BLS consider on-going creation of enterprise frames for matching projects and possibly for developing new surveys?

Identifying Factoryless Goods Producers in the U.S. Statistical System (OEUS – November 2016)

Since 1945, international trade has surged and the global economy has undergone rapid change leading to the rise of global value networks and the relocation of production activities across national borders (Bernard and Fort, 2015). Accurate classification of the primary business activity of establishments is the responsibility of the U.S. Economic Classification Policy Committee, which uses the North American Industry Classification System (NAICS) as the standard for federal statistical agencies when classifying business establishments. Currently, establishments that outsource all transformational activities, known as *factoryless goods producers* (FGPs), are not included in the NAICS, and thus are inconsistently classified across U.S. statistical programs with a possible impact on U.S. economic measures.

Doherty (2012) points out that outsourcing of manufacturing activities and the subsequent fragmentation of the traditional production model (where all production activities are handled within the same establishment or firm) are not consistent with the current structure of the NAICS. Consequently, Doherty also notes that to allow for consistent classification of FGPs a definitional model is needed that reflects modern-day production decisions made by firms in response to ongoing changes in global economic conditions.

In response to this challenge, the Bureau of Labor Statistics (BLS) initiated research to explore key FGP concepts and characteristics. Specifically, this paper summarizes the results of two field tests conducted by BLS, where about 550 establishments in each test were mailed a paper form and then a subset were selected for telephone debriefing interviews. Results suggested that the classification scheme originally developed does may not align well with how establishments operate, and the language used to describe FGP characteristics is ambiguous and industry-specific. This presentation will summarize results from the field tests, present lessons learned, and describe future challenges and research related to FGPs.

Questions for TAC:

1. Do you consider factoryless goods production to be best measured at an enterprise or an establishment level?
2. Do you have suggestions for future research to pursue?

Internet Scanning for Fatal Occupational Injuries (OCWC – November 2016)

The Bureau's Office of Safety and Health Statistics is investigating how to systematically implement a way to identify information found on the internet in an attempt to address two limitations of the current Census of Fatal Occupational Injuries and Illnesses (CFOI) process. First, many outside researchers would like access to the original source documents used to produce CFOI numbers, but we cannot share these because many are collected under a pledge of confidentiality. Second, information posted on the internet is often a useful source

of information about fatal occupational injuries, but manually sifting through this information can be a time consuming endeavor.

This project aims to address both of these issues by continually scanning the internet for new information related to fatal occupational injuries and then saving the resulting articles and any relevant information that can be automatically extracted from them to a searchable database accessible to both outside researchers and CFOI staff. We currently have a working prototype that uses Google Alerts to monitor the internet for new articles related to fatal occupational injuries and a Python program to save those articles to a searchable database, but several outstanding questions remain:

1. Are there better ways to monitor the internet for new work related fatal injury articles than Google Alerts?
2. To what extent can we save and potentially redistribute work related fatal injury articles that might be protected by copyright or other laws?
3. How can we best extract the information we need from these articles?

Occupational Requirements Survey (OCWC – June 2017)

In 2015, the Occupational Requirements Survey implemented production of a nation-wide sample of establishments to support production of occupation specific data for most of the O*NET 8-digit SOC codes. Under the current design, establishments in all industries and sectors of the economy are sampled each year and data from three (or more) years of data collection will be used to produce the ORS estimates. This sample design was carefully tested to ensure that it would meet the goals for this survey but has some potential drawbacks.

The current design does not provide the opportunity to control the number of quotes collected for any specific occupations or even for specific occupational groups resulting in the potential of collecting too much data for occupations in which there is a lot of homogeneity for the ORS data elements and not enough data for occupations where there is wide variation in the ORS data elements. Although the design will provide data for the vast majority of occupations, it is possible that it will not support release of estimates for all the SOC codes that appear frequently in the disability claims processes. In addition, initial research into the frequency with which ORS data elements change over time indicates that ORS values will vary more often for some jobs or types of jobs than for other jobs or types of jobs. This suggests that it could be necessary to update ORS data for some types of jobs more often than for other jobs. But our current sample design only supports updating data for all jobs on a single schedule (i.e. all at once).

So research is being conducted into the potential for using other sample designs that would permit targeting specific occupations with the selection of each sample. Among the options being researched are the potential of selecting samples of occupations that have been collected by the Occupational Employment Statistics (OES) program and/or using prior OES data to adjust sample allocations by industry. Using OES data poses many challenges including acceptance of

OES non-response rates and potential occupation coding errors which vary by industry and occupation and may have a negative impact on future OES response rates. This presentation for TAC will include an overview of the challenges with the current design and a summary of the current research efforts to target occupations during sampling.

Questions:

1. What are additional options for targeting occupations within an industry-based sample?
2. Should we move completely away from an industry-based sample and use an occupation-based sample? If so, do you have suggestions aside from recontacting OES respondents?