Reflections on FESAC (Federal Economic Statistical Advisory Committee) "Ongoing Topics" by John Haltiwanger, University of Maryland[1]

Having served on FESAC since its inception in 2000, I have been privileged to watch the enormous progress by the US Federal Statistical agencies over many dimensions. I will highlight three here. First, there has been great progress in the development and use of administrative data. Second, surveys now regularly include rotating modules for key topics of current interest. Third, statistics are increasingly released on a granular basis (including more detailed geography and/or industry and also second moments). Overarching all of this is a commitment to greater collaboration among the agencies – this motivation is at the core of the activities of FESAC.

While great progress has been made, many economic measurement challenges (and opportunities) remain. Zvi Griliches' insight from his 1998 AEA Presidential Address that the economy is evolving towards business activities more and more difficult to measure is even more relevant today. The five topics listed below are not new but deserve ongoing attention.[2] This document focuses more on the measurement challenges than solutions (although some concrete suggestions are offered). All of the topics are areas of active research but further research and pilot projects are needed. I think FESAC could be very useful in providing guidance and feedback at both early and later stages of addressing these measurement challenges and opportunities. I hope that these remarks provide suggestions for possible future FESAC sessions.

Topic 1: New Uses of Existing Data

Timeliness of administrative data products: BLS and Census now have rich longitudinal business data products derived from administrative data to supplement their high value-added data products such as Quarterly Census of Employment and Wages (QCEW) and County Business Patterns (CBP). However, the data are not timely. The BLS Business Employment Dynamics (BED) and QCEW are significantly more timely and higher frequency than the Census Business Dynamic Statistics (BDS) and CBP. Even so, the BLS data are published with a six-to-nine-month lag (the latest BED data published is for 2022:3 – this compares with the latest BDS data published for 2020). It is a complicated process to track business dynamics. For example, one of the valuable contributions of the BDS is statistics on employer business startups (not just new establishments but true new firms with firms defined in an operational control manner – not just on the basis of taxpayer ID). Getting this right takes time. Having said this, there is a flow of administrative data in both the UI tax data and Federal payroll tax data system on almost a real time basis. That is, employers file quarterly payroll and UI tax reports. The UI data flow from state UI agencies to BLS to build the QCEW (along with multiple BLS business surveys) and the payroll tax data flow from IRS to Census to build the Census Business Register (along with

---

[1] *This document emerged from discussions with the current Chair of FESAC, David Wilcox. David and I have been chatting about the role of FESAC in the past, present and future. David suggested I write down some of my thoughts about current challenges and opportunities in economics measurement. The result is this document which reflects only my views. The discussion of challenges and opportunities for ongoing topics is not intended to be interpreted as explicit or implicit criticism of official statistics by the US Statistical agencies. The topics covered also partly reflect areas where I have been actively engaged in research that overlap with ongoing issues in economics measurement.*

[2] *This document discusses research and findings from many studies without full and appropriate citations. Citations available upon request.*

multiple Census business surveys). To build the high quality QCEW and CBP (and other) public domain data products takes time. However, given that there is a quarterly flow of data there is the potential for developing preliminary statistics on a more timely basis. While the statistics that would emerge from processing the data from the real time flow are not a substitute for the benchmark statistics, they could be an important and timely complement. Moreover, modeling of the relationship between the gold standard benchmark statistics and the real time flow (along with potentially other sources – see below) has great potential. The JEDx (Jobs and Employment Data Exchange -- see https://www.uschamberfoundation.org/JEDx) project for improving and enhancing the UI wage and employer records has many related ideas to push this agenda forward from the UI tax data.

Topic 2: New Uses of New Data

Digitized Data: The digitization of virtually everything including economics measurement has enabled enormous opportunities for improved economics measurement. It has led to alternative sources of key indicators (e.g., ADP employment report) but has also offered the statistical agencies opportunities to tap into this data as part of their data acquisitions. BLS, BEA and Census have active programs using such data in a variety of applications. However, there is an opportunity for fundamentally re-engineering key economic indicators from such digitized data.   The current infrastructure is still largely survey-centric with the digitized data currently used to supplement that data.   Moreover, as discussed further below, the survey data are from disparate business frames across the agencies. Instead of using the digitized data to supplement surveys, digitized data could potentially be used to replace the high frequency measurement of nominal output, prices and real output in at least some sectors. The most obvious opportunity is in Retail Trade where tracking item-level transactions in terms of prices, quantities and product characteristics is ubiquitous. Other sectors have substantial potential as well. The opportunity here is for higher quality, internally consistent, more timely, and more granular data.   Quality-adjustment of prices at scale is feasible with such data taking into account quality improvements from product turnover (currently BLS uses hedonics for only about 7% of products). Taking advantage of this opportunity would require a coordinating, cooperative multi-agency initiative. The AEA Committee on Economic Statistics (AEAStat) held a working session on this possibility at the AEA meetings in January 2022 (https://www.aeaweb.org/about-aea/committees/economic-statistics/data-needs).[3]

AI, Machine Learning and Blended Data: Rapid advancements in machine learning and AI offer new opportunities for improvements in economics measurement. Machine learning has the capacity to integrate data from many disparate sources at different frequencies and timeliness. Many of the issues discussed above and below could be addressed using these methods. Examples include: (i) Real time processing of administrative data and relating such flows to the historical higher quality more fully processed administrative data; (ii) Integrating the business lists from BLS and Census (see discussion below); and (iii) Integrating digitized and survey data (see discussion above).

---

[3] Full disclosure, I am part of large Sloan Foundation project exploring this opportunity (see htps://ebp-projects.isr.umich.edu/RESET/).

Topic 3: Importance of Firm Life Cycle (Entry, Youth and Exit)

Firm Entry and Exit Measures (and Associated Adjustments to High Frequency Statistics): We also need to high quality, timely measures of firm entry and exit. Research has shown that entry and exit are highly cyclical but are not well reflected in high frequency statistics. The high-quality administrative data generates measures of business entry and exit but with considerable lags. A recent exception is the newly developed Business Formation Statistics (BFS) which provides real time indicators of the formation of new businesses in the US.   As noted above, real time processing of the flow of administrative data could offer early model-based estimates of such indicators.   The lack of high frequency measures of firm entry and exit is related to another issue with the high frequency (e.g., monthly) indicators produced by the agencies. The Current Establishment Survey (CES) and Job Openings and Labor Turnover Survey (JOLTS) data from BLS have a birth-death adjustment model but it is based on limited real time information. This yields substantial benchmark adjustments with a lag. Census high frequency monthly surveys do not have a birth-death adjustment and in turn this underlies an important component of large benchmark adjustments.   The pandemic has highlighted the need for these statistics as there is an apparent surge in new businesses from the BFS but it is taking time to see how many of the new applications yield operating young businesses. Business exit surged in 2020 but there was no real time tracking of this activity. Evidence now emerging for 2022 suggests the surge in new businesses is also associated with a surge in business exit. Again, this information is all coming in with a substantial lag.

It's (Business) Age, Not Size: Research has shown that young businesses are of critical importance for job creation, innovation and productivity. Importantly and relatedly, it is primarily the high growth young businesses (only a small share of young businesses) that play a disproportionate role. The research demonstrating this has largely arisen from the developments of the longitudinal business databases at the US Statistical Agencies integrated with other data (e.g., patent data). The research has been undertaken by staff at the agencies as well as researchers with approved projects at the statistical agencies. This research has changed the way the economics profession thinks about the role of small vs. young businesses. Critically important, young businesses are small but not all small businesses are young. Small, mature businesses play an important role in the economy but are often niche, local businesses and are not the primary source of job creation, innovation and productivity growth. Why is this important? While the research underlying these findings has largely been developed at the US statistical agencies, this insight is not reflected in the surveys or data products by the agencies to a large extent. The business surveys are often designed to be representative by industry, business size and if possible, location. They are not representative by business age. Relatedly, many business data products by the statistical agencies release statistics by business size. It is only the data products from the longitudinal business data infrastructures that release statistics by business age.   Moreover, since it is especially high growth young businesses that play an outsized role, statistics that track high growth young business activity are important. Such statistics largely do not exist as regular data products.

Topic 4: Special Topics

Intangible Capital: The research community has increasingly recognized that to understand productivity and economic growth intangible capital is critical to model and measure. The most comprehensive definition of intangible capital is from the work of Corrado, Hulten and Sichel. Their characterization is that intangible capital investment includes any inputs into future rather than current period production. In practice, the measurement of intangible capital has focused on key but limited components of such inputs: R&D, training, brand capital. The US statistical agencies have helped contribute to the measurement of such activity but this is a wide-open area for economics measurement. For example, young businesses are inherently engaged in intangible capital investment. However, such investments are not well captured in current economic measurement.

Tasks: The traditional perspective on measuring productivity and business activity is via the specification of a production function relating outputs to inputs. The work of Daron Acemoglu, David Autor and Pascual Restrepo (and multiple collaborators) has highlighted the insights from a different specification relating business output to a series of tasks. These tasks can potentially be performed by either labor or capital. This perspective helps greatly in characterizing the developments of advances of technology such as robotics, automation and AI and the possible complements and substitutes with different types of worker skills. The statistical agencies are actively involved in core aspects of economics measurement in this area by tracking occupations in household and business data. Moreover, combined with the O*NET quantification of tasks by occupation from the Department of Labor, much progress has been made in the evolution of changing task requirements in the workforce. Relatedly, alternative data sources such as Lightcast (formerly Burning Glass) track changing task requirements of jobs. While conceptual and measurement progress has been made, much work needs to be done. The Acemoglu and Restrepo task approach to production implies a different approach to quantifying productivity. This would require a different approach to economics measurement for quantifying productivity. Relatedly but distinctly, rapidly changing technology such as AI implies a need for real time quantification of changing task requirements in the workplace. Also, efforts by the statistical agencies to track adoption of advanced technology should also be integrated with efforts to track the changing task content of production. The 2019 Annual Business Survey (ABS) module is an effort in that direction but much more needs to be done.

Gig Economy: While there has been much interest in the gig economy, the economics measurement has proved challenging. Substantial discrepancies between self-employment activity using household and administrative data have emerged. Over the post-2000 period, the household data show flat or even declining self-employment rates. In contrast, self-employment rates using administrative data are rising. Research conducted at the statistical agencies (and using IRS data) shows this reflects many factors. First, some independent contractors mistakenly classify themselves as wage and salary in the household data. Understanding this is important since this is a different type of employer-employee relationship than a wage and salary worker. Second, much self-employment activity is stopgap or supplemental and the household surveys don't capture multiple jobs or secondary activity well. For example, someone who regards themselves as primarily a student or even retired may be engaged in some secondary self-employment activity for pay but does not report that on a household survey. Prior to the

pandemic, much of the increase in the gig economy in the administrative data is in the ridesharing industry. Even here the household data do not capture this well compared to the administrative data. In the pandemic and its recovery, the BFS suggests that there has been a surge in individuals starting new businesses including nonemployer businesses (i.e., self-employment activity). This includes a surge in new businesses in Professional, Scientific and Technical Services which includes important components of High-Tech businesses.   It is too early to track this with the administrative data on nonemployers but again the household data do not seem to be capturing this pattern.

Topic 5: Importance of Collaboration

Progress on many of the above topics requires collaboration between BEA, BLS and Census. At the core of this need for collaboration is a longstanding vital issue that warrants more attention. Specifically, the fundamental discrepancies in the quantification of economic activity by US businesses between BLS and Census. These arise for complex reasons. At the core, BLS and Census use different business frames. FESAC and other advisory committees (AEAStat) have long advised for the enabling legislation that would permit sharing of FTI (Federal Tax Information) between BEA, BLS and Census with privacy protections from CIPSEA (Confidential Information Protection and Statistical Efficiency Act). But we know that such enabling legislation would only be a necessary first step. BLS and Census have naturally developed independent data infrastructures tracking US Businesses building on their independent frames.   For example, BLS works closely with its state partners along with its many business surveys including the Annual Refiling Survey, the Multiple Worksite Reports and the Current Establishment Survey. Census has the Economic Census, the Company Organization Survey (now called the Report of Organization) and its array of business surveys. The result is that BLS and Census business data do not agree on employment, payroll, industries, locations, number of establishments, or number of taxpayer IDs (EINs) at various levels of aggregation. These discrepancies on multiple dimensions limit the value added of the business data sharing that is feasible under CIPSEA without sharing of FTI. BLS shares its industry codes with Census under CIPSEA which is quite useful but the lack of common establishment identifiers, discrepancies in common taxpayer IDs, and differences in the classification of establishments as to whether they are part of single-unit or multi-unit firm imply that Census can use these codes to supplement but not to reconcile industry codes. Census shares its firm identifiers with BLS under CIPSEA which has the potential value of permitting BLS to generate statistics at the enterprise (firm) level based on operational control rather than taxpayer ID (the current practice at BLS). But again, these discrepancies make it challenging to use this information.

Why does this matter? It is of critical importance for the measurement of GDP, productivity and inflation. In producing the National Income and Product Accounts (NIPAs), BEA integrates the payroll, employment, hours, and price data from BLS with the revenue and expenditures data from Census. The payroll, employment, hours and producer price data depend on the BLS business frame and associated business data infrastructure. The revenue and expenditures data depend on the Census business frame and associated business data infrastructure.   The input-output tables critical for NIPAs and productivity are largely based on the Economic Census. The discrepancies in measuring business activity imply that the most basic core measures of real

output and productivity are adversely impacted.   Simply put, real output involves combining numerators (nominal output) and denominators (prices) from disparate business data infrastructures. Likewise, productivity measurement adds further complications to the mix with some inputs from BLS (labor) and others from Census (capital and materials).

It is worth noting that progress on key topics above can facilitate this reconciliation. For example, using integrated price and quantity data from digitized sources would overcome many of these limitations. Using machine learning to integrate and reconcile data from disparate sources can also play a critical role.

In closing, it has been an honor and learning experience to serve on FESAC all these years. I look forward to seeing progress in future years from the activities of FESAC.

June 9, 2023