THE AMERICAN ECONOMIC ASSOCIATION

Committee on Economic Statistics and Committee on Government Relations

Data Synchronization: The Time is Now

June 2023

Executive Summary

The problem: U.S. economic statistics could serve policymakers and the public better if not for the statistical agencies' uneven access to key source data. Under Section 6103 of the Tax Code, the Census Bureau can access federal tax information (FTI) on various types of U.S. businesses for statistical purposes. The Bureau of Economic Analysis (BEA) can access FTI, but for corporate businesses only. The Bureau of Labor Statistics (BLS) has no access to FTI, which also prevents BLS from accessing Census business data that incorporates FTI. This uneven access causes inconsistencies across key economic indicators, undercutting accurate measurement of changes in the U.S. economy and adversely impacting decisions made by businesses, households, and policymakers.

The result: Disparities in key measures of the strategically important semiconductor industry illustrate the consequences of current constraints on the agencies' abilities to access and share business data. According to Census data, there were 780 establishments and 95,400 employees in this industry in 2021, while BLS data indicated there were 1,900 establishments and 185,000 employees. Time trends in the Census and BLS measures also differ significantly. Thus, for this important industry we have vastly different measures of the level and growth of productivity and employment, hampering design of effective, well targeted policies to ensure continued growth and innovation in U.S. semiconductor supply. And this is just one example.

An essential step toward better statistics: Congress can amend section 6103 of the Tax Code to allow the disclosure of limited business tax data to BEA and BLS for certain statistical purposes. This change will remove obstacles to sharing business data across agencies, paving the way to substantial improvements in key economic indicators that businesses, communities, and governments rely on for decision making – including measures of job growth, inflation, gross domestic product (GDP), and productivity. It will also permit better measurement of employment, investment, technology adoption, and productivity growth in sectors of strategic importance for the U.S. economy, including semiconductors and advanced manufacturing. This small change would pave the way for many large improvements in the quality, accuracy, and consistency of U.S. economic statistics, with widespread benefits for the U.S. public and evidence-based policy.

AEA Contact: Martha Starr, PhD, Washington, DC, Representative of the AEA and Director of Government Relations. martha.starr@aeapubs.org. 703-283-0054.

Background

Business tax records provide a valuable input into the **Census Bureau**'s Business Register – a comprehensive list of U.S. businesses that Census uses to draw representative samples for its business surveys. These surveys provide statistics on key economic variables such as business sales and expenditures. The business surveys, including the quinquennial Economic Census and the Annual Report of Organization Survey, enable enhancement of the Business Register for the measurement of industry codes, as well as establishment-level activity for businesses operating in multiple locations. As businesses are required to file tax forms with the IRS on a regular basis, access to tax filings allows the Census Bureau to maintain a comprehensive, accurate, up-to-date list of U.S. businesses in operation, including businesses' names, addresses, principal activities, and other characteristics.[1] Section 6103(j) of the Tax Code[2] gives the Census Bureau access to tax information on businesses of all types "for the purpose of, but only to the extent necessary in, the structuring of censuses and national economic accounts and conducting related statistical activities authorized by law." Provisions of federal laws, Census Bureau policies, and data-use agreements strictly limit access to and use of tax data within the Census Bureau, and rigorous disclosure avoidance techniques are used to prevent any identifiable information from being publicly released. Given the close integration of survey and administrative data, most of the Census Bureau business data are considered commingled with tax data and cannot be shared with BLS or BEA. A key public-domain product based on the Census Business Register is the annual County Business Patterns data, which is published with about a one-year lag.

Like the Census Bureau, the **BLS** is charged with producing a set of key economic statistics that depend on data collected from U.S. businesses. In partnership with state agencies, BLS maintains its own business register, called the Quarterly Census of Employment and Wages (QCEW). BLS uses the QCEW as the sample frame for its periodic surveys of business activity which collect information on employment, occupations, hours, wages, and producer prices. For example, the monthly market-moving jobs counts from BLS's Current Employment Statistics program and Job Openings and Labor Turnover Survey rely on the QCEW business register. The QCEW collects administrative data on all U.S. employers that must participate in state unemployment-insurance (UI) systems (covering 95-98% of total U.S. employment). In the QCEW, comprehensive, frequent, and current UI records are augmented with information collected from businesses via the BLS's Annual Refiling Surveys and Multiple Worksite Reports. Such surveys play a critical role for the QCEW in measuring industry codes and establishment-level activity for businesses operating in multiple locations. QCEW records include business name, location, number of employees, industry, and quarterly payroll.[3] But BLS has no access to FTI under the Tax Code, which strictly limits its access to the Census business register because it includes FTI. As with the Census Business Register, provisions of federal laws, BLS policies, and data-use agreements strictly limit access to and use of QCEW data within the BLS, and rigorous disclosure-avoidance techniques are used to prevent any identifiable information from being publicly released. A key public-domain product based on the QCEW data is the County Wages and Employment data, released by BLS on a quarterly basis, with about a 9-month lag.

[1] B. De Salvo, et al., "Documenting the Business Register and Related Economic Business Data," Census Bureau Center for Economic Studies WP-16-17, March 2016.
[2] The term Tax Code refers to both Title 26 of the U.S. Code and the Internal Revenue Code of 1986.
[3] For more information about the QCEW, see https://www.bls.gov/cew/overview.htm.

Thus, the two agencies maintain separate business registers that do not record the characteristics of individual businesses in the same way.[4] The Census Bureau can (and does) use information collected by BLS to improve the accuracy of its business register as BLS industry codes are shared with Census, but BLS cannot access or use Census data that incorporates FTI. There is some sharing of business information from Census to BLS on firm identifiers that do not rely on FTI. A challenge for Census and BLS is that BLS's register has businesses not present in the Census data and vice versa. A related challenge is that different processes are used to measure establishment-level activity of multi-unit firms operating in multiple locations. Unresolved differences between Census and BLS business registers are thus a source of persistent inconsistencies in U.S. economic statistics, to the detriment of accurate measurement of economic activity. Critically, the agencies use these different business registers as sample frames for a wide range of business surveys that underlie key economic indicators. The inconsistencies in sample frames and information collected in the resulting surveys further contribute to inconsistencies in measurement of key economic indicators.

BEA differs from Census Bureau and BLS in that its primary responsibility is not to collect source data – but rather to produce the National Income and Product Accounts (NIPAs) using source data largely collected by other entities. The NIPAs tell us how fast the U.S. economy is growing, how much is being spent on business investment, the growth and composition of U.S. exports and imports, how much households are saving, levels of business profits, and more. Under the Tax Code, BEA receives tax information from the IRS for corporate businesses, which feed into the NIPA estimates. But it does not have access to FTI for non-corporate businesses (such as sole proprietorships and partnerships), nor to Census business data that incorporates FTI.[5] This implies BEA uses published aggregates from BLS and Census to build the NIPAs, rather than building up from business-level data. A consequence of these limitations is that BEA and BLS compute measures of productivity (e.g., output per hour) for U.S. industry by combining (published) industry output statistics from Census Bureau business surveys with (published) industry employment data from BLS business surveys, despite the apples-to-oranges issues caused by using unsynchronized business data from the two sources.

A related problem is that the Producer Price Index (PPI) produced by BLS is impacted by inconsistencies in the sample frames. BLS collects price information via businesses surveys based on its sample frame, but uses weights from Census's business surveys and Economic Census (which are based on the Census sample frame) to construct indices. This is a potential source of measurement error in this key measure of inflation that also impacts the NIPAs in a fundamental way. BEA estimates real output at the industry level by taking data on nominal output from Census sources and adjusting the data for inflation using BLS's PPI. Thus, the inconsistent frames not only adversely impact measures of productivity (output per hour), but the measure of output itself.

[4] For detailed discussion of the differences between the two business registers, see R. Becker, et al., "A Comparison of the Business Registers Used by the Bureau of Labor Statistics and the Bureau of the Census," August 2005, and K. Fairman, et al., "An Analysis of Key Differences in Micro Data: Results from the Business List Comparison Project," October 2008.

[5] The IRS can provide BEA with certain tabulations from tax return information.

What's needed

The Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002 authorized data sharing between the Census Bureau, BLS, and BEA, with the objectives of: increasing the efficiency of the statistical system; improving the quality, accuracy, and consistency of economic statistics; and reducing respondent burden. It was recognized then that **companion legislation to amend Tax Code Section 6103** would be needed to realize these benefits. But this companion legislation has not yet been enacted, which leaves the statistical system operating below potential and key measures of changes in the U.S. economy laden with "noise" and uncertainty because business data are not synchronized across statistical agencies. This impairs data-driven decisions by businesses, governments, and communities. It also impairs the accumulation of evidence on the effectiveness of government-provided incentives or investments in achieving policy goals, as required by 2018's Foundations for Evidence-Based Policymaking Act.

The required enabling legislation should be viewed as a necessary first step to pave the way for improved economic statistics. Due to current differences in registers, sample frames, data collected, and data infrastructures at the agencies, substantial time and resources will be required to better synchronize the agencies' business data. The enabling legislation is a necessary first step to make significant progress in realizing gains from synchronization.

Flawed economic statistics, flawed economic decisions

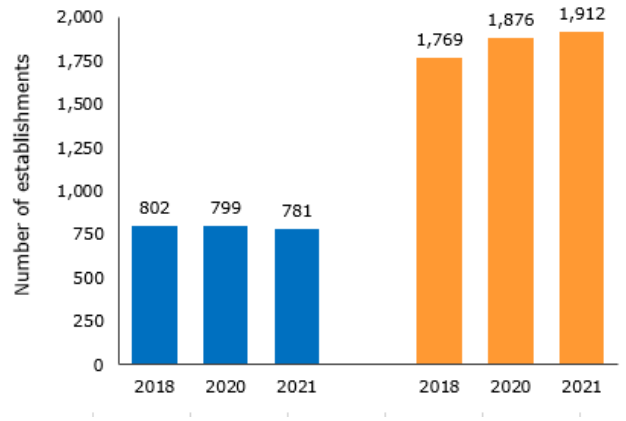
Economic statistics affect Americans' lives in many significant ways. Private businesses use economic statistics to track changes in the economy and plan investments, hiring levels, location decisions, and other activities. Governments and communities depend on economic statistics to devise policies and programs appropriate for the populations they serve and, increasingly, to accumulate evidence on effects of policies and programs in addressing needs. Consumers hear news about job growth, inflation, and economic growth and may adjust their spending, saving, and borrowing accordingly. Current data gaps and deficiencies impair the value of economic statistics to all these sets of users, because they provide noise-ridden pictures of economic trends.

As a timely example of economic statistics that need upgrading to meet evidence-based policy needs, the **CHIPS and Science Act of 2022** allocated $52 billion to strengthening the U.S. semiconductor industry, including $32 billion to encourage new investments in U.S.-based fabrication equipment and facilities. But the picture of the U.S. semiconductor industry available from economic statistics is fuzzy at best. As the chart below shows, there were 781 establishments producing semiconductors in the U.S. in 2021, according to the Census Bureau – or 1,912 according to the BLS. According to Census, the number of establishments fell by 21 between 2018 and 2021, while the BLS data show an increase of 143. Census data show a modest decline in semiconductor employment between 2018 and 2021, while BLS data show an increase – but the level of employment in the BLS data is substantially above the Census level.

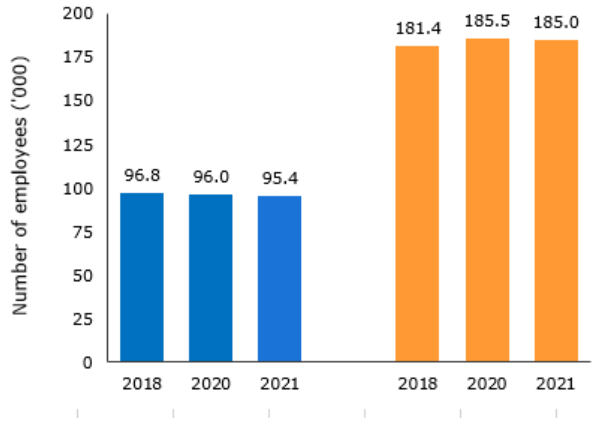
Census Bureau vs. BLS data on semiconductor manufacturing

NAICS 334411 Semiconductor and related device manufacturing

Number of establishments



Number of employees ('000)



County Business Patterns (CBP) -- tallies from Census Business Register

Quarterly Census of Employment and Wages (QCEW) -- tallies from BLS Business Register

A key problem that gives rise to these discrepant views of the semiconductor industry concerns North American Industrial Classification System (NAICS) codes, used to classify business establishments by industry. The Census Bureau assigns NAICS codes to businesses using multiple sources, including tax filings, its own data collections, and data collected by BLS. The BLS asks businesses to identify their NAICS code and validates entries as needed. While Census can check its NAICS codes against BLS codes for businesses identified by both agencies, the opposite is not true – because Census codes rely in part on tax data that BLS employees cannot access. Amending the Tax Code to undo this constraint would give the agencies two-way access to each other's data, providing new opportunities to synchronize their business lists and measure industry trends more consistently. Two examples demonstrate the importance of having accurate industry statistics for U.S. economic policy. The 2022 CHIPS and Science Act calls for defining a "critical manufacturing industry" based on the value-added and employment shares of the industry. Relatedly, an Executive Order of the President (14017) requires an assessment of critical supply chains supporting the Information and Communications Technology (ICT) industries. This assessment includes quantifying the impact of outsourcing on the ICT workforce. These requirements and assessments are difficult given the large discrepancies between BLS and Census data.

The reliability of a wide range of other economic statistics would increase if obstacles to data synchronization across the statistical agencies are removed – to the benefit of the U.S. public, businesses, government agencies, and communities. As some notable examples:

- *Better information on productivity growth*. Rising labor productivity is the fundamental driver of improvements in living standards for the American workforce. BLS and BEA measure productivity trends across industries by dividing Census Bureau estimates of output in an industry by BLS data on hours worked in an industry. But we know that, because NAICS codes are not consistent across data sets, establishments producing the output may not be classified the same way as the establishments employing the workers in an industry. By one estimate, establishments may be coded differently

across data sets as much as 30% of the time. In an era when significant technological changes are underway that should have strong potential for raising labor productivity – including digitization, cloud computing, artificial intelligence, and advanced manufacturing – consistently categorizing establishments by industry would substantially improve our ability to track the changes these trends bring about for American workers.

- *Better data on U.S. trade in services.* The U.S. is the world's largest exporter of services, with U.S. businesses supplying financial services, transportation services, intellectual property services, business services, and more to businesses, consumers, and governments in other countries. As no other statistical agency regularly collects information on international trade in services, the BEA conducts quarterly business surveys to fill in the gap. Yet firms replying to BEA surveys are known to underreport service exports relative to what they report in the Census Bureau's Economic Census (conducted every five years). If the BEA could access data for both corporate and non-corporate businesses from the Economic Census, it could draw better samples of businesses to respond to its surveys and would have a statistical basis for addressing problems of underreporting. With the relative importance of U.S. service exports rising, reliable data for this component of national output is again key for accurately measuring U.S. economic growth.

Use of tax data is subject to substantial protections and limitations

Protecting the confidentiality of both FTI and statistical data is essential. A 200+ page IRS publication spells out many rigorous disclosure-prevention provisions that government agencies must follow to prevent unauthorized access or use or disclosure of FTI. These include: keeping the data within a secure perimeter, limiting access to authorized need-to-know personnel, utilizing access control systems that can be audited, using encrypted data storage, undergoing rigorous disclosure review before publishing statistical information, and many more. CIPSEA Part B, similarly specifies how the statistical agencies are to collect, manage, and use individually identifiable information, so as to safeguard the protection of all data collected under a pledge of confidentiality. Both sets of provisions specify civil and criminal sanctions (including fines and imprisonment) for unauthorized disclosure or inspection of confidential information. The Census Bureau, BLS, and BEA already operate data systems built around the need to protect individually identifiable data (including business data), with only authorized personnel having access to them.

The Tax Code clearly specifies that the statistical agencies can use FTI only for statistical purposes, namely, "structuring censuses and national economic accounts and conducting related statistical activities authorized by law," and "only to the extent necessary" for this work. CIPSEA Part C, §3575(4) pointed to specific ways in which increased data sharing for statistical purposes would improve agencies' ability to track "the large and rapidly changing nature of United States business," by allowing them to "ensure that businesses are consistently classified in appropriate industries, resolve data anomalies, produce statistical samples that are consistently adjusted for the entry and exit of new businesses in a timely manner, and correct faulty reporting errors quickly and efficiently." CIPSEA Part B, §3576(3) also spelled out that increased data sharing would raise the comparability and accuracy of federal economic statistics by allowing the statistical agencies to "update sample frames, develop consistent classifications of establishments and companies into industries, improve coverage, and reconcile significant differences in data produced by the three agencies."

Thus, amending Section 6103(j) of the Tax Code would enable specific projects needed to improve the value of economic statistics to stakeholders, subject to well-established provisions for safeguarding privacy that are firmly in place.

Broad-based support for data synchronization

Proposals to amend the tax code to permit increased sharing of business data for statistical purposes have broad-based support in key stakeholder communities.

- The American Economic Association has written multiple briefs highlighting the benefits of data synchronization.[6]
- Think tank scholars across the spectrum have equally chimed in.[7]
- Numerous advisory committees and expert panels have called for timely progress on data synchronization, including the National Research Council,[8] the Federal Economic Statistics Advisory Committee,[9] and the Committee on National Statistics of the National Academy of Sciences.[10]

The U.S. Treasury's "General Explanations of the Administration's Fiscal Year Revenue Proposals" ("Green Books") for 2022, 2023, and 2024 have included a proposal to "allow officers and employees of each of BLS, BEA, and the Census Bureau to access the same FTI for businesses, and would permit BLS, BEA, and the Census Bureau to share such FTI amongst themselves," subject to "CIPSEA confidentiality safeguard procedures, requirements, and penalties" and other applicable standards. The 2024 Treasury proposal provided specifics about the data synchronization as follows:

- The proposal would give officers and employees of BEA access to FTI of those sole proprietorships with receipts greater than $250,000 and of all partnerships. BEA contractors would not have access to FTI.
- The proposal would give BLS officers and employees access to certain business (and tax-exempt entities) FTI, including: Taxpayer Identification Number (TIN); name(s) of the business; business address (mailing address and physical location); principal industry activity (including business description); number of employees and total business-level wages (including wages, tips, and other compensation, quarterly from Form 941, Employer's Quarterly Federal Tax Return, and annually from Form 943, Employer's Annual Federal Return for Agricultural Employees, and Form 944, Employer's Annual Federal Tax Return); and sales revenue for employer businesses only. BLS would not have access to individual employee FTI. In other words, the proposal would allow officers and employees of each of BLS, BEA, and the Census Bureau to access the same FTI for businesses, and would permit BLS, BEA, and the Census Bureau to share such FTI among themselves (subject to

[6] AEA, "Statement on Data Synchronization," April 2015; AEA Letter to Treasury Secretary Janet Yellen, July 2021; AEA Policy Brief, "How Data Sync Can Save Official Statistics," September 2021.
[7] M. Strain, "Data Synchronization: The Time Is Now," American Enterprise Institute, August 2016. D. Whitmore Schanzenbach and M. Strain, "America's small investment in government data has big payoffs," Brookings Institution, March 2017. N. Eberstadt, et al., "In Order That They May Rest Their Arguments on Facts: The vital role of government-collected data," a Joint Hamilton Project-AEI Report, March 2017.
[8] National Research Council, "Improving Business Statistics Through Interagency Data Sharing," 2006.
[9] FESAC, Statement on Data Synchronization, December 2014; reaffirmed June 2021.
[10] CNSTAT, "Toward a 21st Century National Data Infrastructure: Mobilizing Information for the Common Good," 2023.

restrictions described below). This sharing would include commingled Census business survey and tax data. No BLS contractor would have access to FTI.

The proposal would require any FTI to which BEA and BLS would have access, either directly from the IRS, from the Census Bureau, or from each other, to be used for statistical purposes only, consistent with the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA). The three statistical agencies would be subject to taxpayer privacy law, safeguards, and penalties. They would also be subject to CIPSEA confidentiality safeguard procedures, requirements, and penalties. Conforming amendments to applicable statutes would be made as necessary to apply the taxpayer privacy law, including safeguards and penalties to BLS as well as the Census Bureau and BEA. [11]

Bottom line

Amending Section 6013 of the Tax Code to enable the synchronization of business data collected and produced by Census, BEA, and BLS is a small change that can be counted on to pave the way for many improvements in the quality, accuracy, and consistency of U.S. economic statistics, with widespread benefits for the U.S. public and evidence-based policy.

[11] U.S. Treasury, "General Explanations of the Administration's Fiscal Year 2024 Revenue Proposals," March 2023.