Business Employment Data:
YTS versus CES and CPS

Gary Kunkle, PhD, Senior Research Fellow
Business Dynamics Research Consortium
May 2018

Gregg Cole, Data Science Research Lead
YourEconomy
March 2025

Many researchers require comprehensive and accurate data that measures U.S. and regional employment or that tracks business-level employment and location in order to execute their projects. This paper compares a leading private sector source of business establishment data, YourEconomy Time-Series (YTS), with two employment datasets released by the U.S. Bureau of Labor Statistics (BLS) - the Current Employment Statistics (CES) and the Current Population Survey (CPS). Before researchers decide which sources of employment and business-level information they will use in their studies, they should review some of the evidence regarding the strengths and limitations of YTS verses alternative government sources. Our findings are in the Conclusion section.

What is YTS?

YTS (YourEconomy Time Series) uses the Data Axle (see below) Business Data Historical files as well as its quarterly data full release business data to develop its semi-annual establishment-level time-series database used by academic researchers and economic development analysts following companies at their unique locations across the U.S. It focuses on establishments that are "in-business" meaning they are intent on conducting commercial activities, or have a DBA (doing business as) physical location. YTS tracks for-profit (both privately-owned and publicly-traded), non-profits, and government establishments[1].

---

[1] For additional information regarding YTS or other datasets maintained by YourEconomy, or their related services, please visit their website at: https://youreconomy.org

A Google Scholar search in March, 2025 yielded 61 peer-reviewed journal articles and white papers completed using the YTS database (see Appendix).

YourEconomy (formerly knows as BDRC - Business Dynamics Research Consortium) is a fourteen-year-old non-profit research organization with a mission to provide unique, valuable, and otherwise unattainable information to scholars that seek to understand business activities, employment creation and job flows. YourEconomy is developed inside the Universities of Wisconsin Office of Business and Entrepreneurship located in Madison, Wisconsin.

Data Axle

Data Axle[2], established in 1972, is one of the world's leading providers of business and residential data to business, government, and education marketplaces. Their datasets are designed to support marketing and business research initiatives.

To develop its datasets, Data Axle operates a 225-seat call center that makes contact with over 60,000 businesses each and every day in order to record and qualify company information. During a typical month, 15% of the entire Data Axle business dataset is re-verified. On average, 150,000 new businesses are added while 100,000 businesses are removed each month, capturing the dynamic business churn happening in the economy.

Data Axle's team also identifies new companies through U.S. Yellow Pages, county-level public sources on new business registrations, industry directories, and press releases.

Ongoing company changes are monitored via news feeds and web research, annual reports and 10-Ks, and real estate deeds and assessments.

CES and CPS

The Bureau of Labor Statistics (BLS) oversees the creation of both the Current Employment Statistics (CES) and the Current Population Survey (CPS). The CES is commonly referred to as the 'payroll survey' and the CPS as the 'household survey'.

---

[2] More information about Data Axle: https://www.data-axle.com

CES collects survey data from a sample of employers then adjusts the data to reflect insights from ES-202 data. CPS only surveys households, inquiring about the employment status of residents.

ES-202 data forms the base of a wide assortment of U.S. government employment and wage statistics. BLS uses Unemployment Insurance ES-202 data to compile the CES survey, its Quarterly Census of Employment and Wages (QCEW), and the County Business Patterns. In addition, the Census of Manufacturers (CM), the Longitudinal Research Database (LRD), the Longitudinal Employer Household Dynamics (LEHD) data are based on Census Bureau and ES-202 data (Kolko & Neumark, 2007).

The CES reports industry-level estimates of total employment of nonfarm payrolls by surveying nearly 150,000 businesses and government agencies. CES covers all 50 states, the District of Columbia, the U.S. Virgin Islands, and Puerto Rico.

The CPS is based upon a monthly survey of about 60,000 households. It reports estimates of the U.S. employment and unemployment rates as well as the number of people employed and those not in the labor force.

"YTS seems to be the better choice for researchers looking for an employment dataset that is as encompassing as the CPS data, as reflective of downturns as the CES data, but has more annual sensitivity to changes than either CPS or CES."

The Covered Employment and Wages Program is the formal name of the ES-202 collection process, which operates under a cooperative agreement between the BLS and state employment security agencies. States collect and process the ES-202 data based on the condition of confidentiality with BLS. As a consequence, government statistics are released in a highly aggregated form based on a limited number of categories for geographic region, industry (at four-digit SIC level) and employer size.

Unlike with YTS, CES's underlying establishment- or firm-level data is not routinely available to researchers. While it is technically possible for researchers to access the raw data, the complex and time-consuming approval process substantially limits access. Furthermore, scholars are restricted in the level of disaggregation

they can report in their work. Because of these severe restrictions, researchers are generally limited to a small number of variables (location, industry and firm size) contained within aggregated data to test economic theories.

Critiques of Private Sector Data Sources

It is the responsibility of all researchers to question the limitations of their data. This should extend to both government-sourced data such as CES and CPS as well as to emerging alternatives such as YTS.

As previously mentioned, YTS is comprised of data originating from Data Axle's historical files. Because this underlying data originates from a private-sector source rather than a government agency, it is appropriate to first address the critiques presented when David Birch originally used private-sector sources in his landmark study on job creation back in the 1980s.

David Birch, a pioneer researcher in the field of small business growth studies, used a dataset somewhat similar to Data Axle's historical files. He compiled his dataset using Dun and Bradstreet's (D&B's) Market Identifier (DMI) files on a bi-annual basis between 1969 and 1986 (Birch, 1987). Birch's findings were widely quoted by the U.S. Small Business Administration (SBA) and others as evidence that small businesses create the vast majority of new jobs in the U.S. economy.

Several scholars harshly criticized the use of DMI data by Birch and the SBA and, by extension, the use of private sector sourced data in general. They presented three main arguments (see Davis, Haltiwanger and Schub, 1998). First, DMI records are unsuitable for regional employment research because they are prepared for commercial purposes rather than as a tool for statistical analysis. Second, DMI files do not adequately account for establishment births, nor do they accurately track younger firms. Finally, there are discrepancies in the total U.S. employment figures cited by DMI files when compared with statistics published by the BLS.

We now review each of these possible critiques of private-sector data:

1) Commercial datasets may be unfit for analysis

Although Data Axle sells business information, they are not in the economic data generation business. They do not smooth their data through seasonal revisions nor by comparing it with other macroeconomic trends as is done with CES. Data Axle simply conducts checks on the validity of the data and scrubs it of anomalies, such as missing data and miscoded information. Hence, Data Axle, and by extension YTS, can be considered to be raw business data, and neither Data Axle nor YTS make any further claims. Experienced researchers know that data from sources other than the government have frequently been used in peer-review statistical studies and been found to be insightful.

2) Births and younger/smaller businesses may be underreported

The private sector has introduced many substantial improvements in the methodology used to gather, screen, and clean establishment-level data since 4Birch's dataset was developed in the 1980s (Neumark, Zhang, and Walls, 2005). For example, in 1991 the regional Bell telephone companies were allowed for the first time to sell the information they collected. This enabled private-sector business information firms like Data Axle to greatly increase the number of establishments they reported and verified, simply by using the Yellow Pages to identify new business units.

YTS and Data Axle will usually start tracking a business that only consists of the founder within one year from its start date. In contrast, CES's ES-202 data would not begin tracking the business until it hires one full-time worker and began submitting unemployment compensation filings.

There still remains some lag in the detection and verification of some establishments in the first year of operations in the Data Axle (YTS) data. Yet, all available sources have difficulty identifying companies at the precise moment of birth, government sources included. The magnitude of this difficulty has been greatly reduced in recent years among private sector sources due to advances in information technology and "big data" sources (e.g. internet-based searches, on-line directories and real-time company announcement, yellow-pages availability, etc.), which has increased the speed of detection and the accuracy of the information regarding new and small firms.

3) Differences in total employment

As shown in the table below, YTS reported 15,402,508 more businesses in 2017 than CES (a difference of 9.7%).

T1: Total Employment Reported by CES, CPS, and YTS (2008-2017)
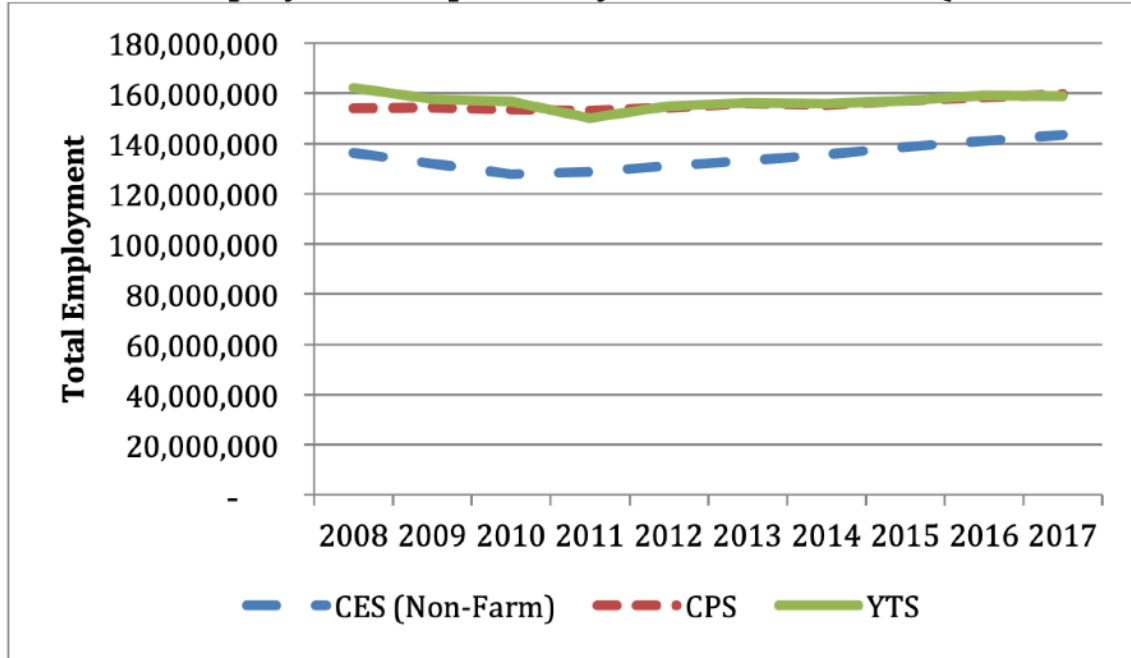
YTS counts employment in all businesses in the economy including in the farming and agricultural sector, whereas CES only reports non-farm 5employment. In sharp contrast, CPS and YTS both count jobs in all sectors, and their difference in 2017 was a mere 0.58%. Essentially, CPS and YTS report nearly identical levels of total national employment.

CES and CPS Divergence

Perhaps one of the more troubling issues with the BLS data is the wide disparity in total employment reported by CES and CPS. As stated by Ben Bernanke, former Chairman of the Federal Reserve, "We do not fully understand the differences in employment reported by the payroll and household surveys, and the truth probably lies in between the two series" (Kane, 2004).

Even once farming sector jobs have been removed from the household survey (CPS) the discrepancy is large and cyclical (Kane, 2004; Neumark, Zhang, & Walls, 2005).

The gap in employment between the CES and CPS surveys widens during economic downturns and narrows during recoveries.

The primary problem appears to be that CES does not count self-employed workers while CPS (and YTS) does. Yet researchers have tracked a large surge in self-employment over recent years in the form of limited liability companies typically used by contract-based workers such as independent consultants and other contract employees.

As Tim Kane at the Heritage Foundation states,

> *"Analysts know intuitively that today's economy is structurally different from the economy of five or 10 years ago, but the consequences of the new economy are difficult to predict. Perhaps payroll jobs are weak simply because the modern company relies less on payrolls for engaging the labor force."*

CES Data smoothing

The BLS creates the CES based on an annual sample of establishments, then uses statistical modeling to cross-check and modify other data that was collected and submitted by states as part of their ES-202 unemployment compensation reporting requirements.

In fact, CES' state-submitted employment data has been modified many times before it is published. Unemployment compensation figures are rounded and smoothed to eliminate "peaks and valleys" by averaging the data over multiple years. It is adjusted and revised to eliminate seasonal variations and business cycles. And it is synchronized with other federally- generated statistics such as Census Bureau data. Statistical modeling and estimation is also needed because only about 57 percent of manufacturers, for example, respond to the CES survey (Copeland, 2011). This creates substantial sampling bias in the original survey data.

Every state agency modifies their unemployment compensation data in unique ways before giving it to BLS. Indeed, there is such a large statistical variation in the reporting of employment statistics by states that the BLS website cautions against aggregating data from the different states. They advise in bold type: **Due to these statistical limitations, BLS does not compile a "sum of states" employment**

series and cautions users that such a series is subject to a relatively large and volatile error structure (see https://www.bls.gov/web/empsit/cesfaq.htm).

The table below shows a close-up of the same data presented in T1 so that we can compare annual changes.

T1: Close-Up, Total Employment Reported by CES, CPS, and YTS (2008-2017)
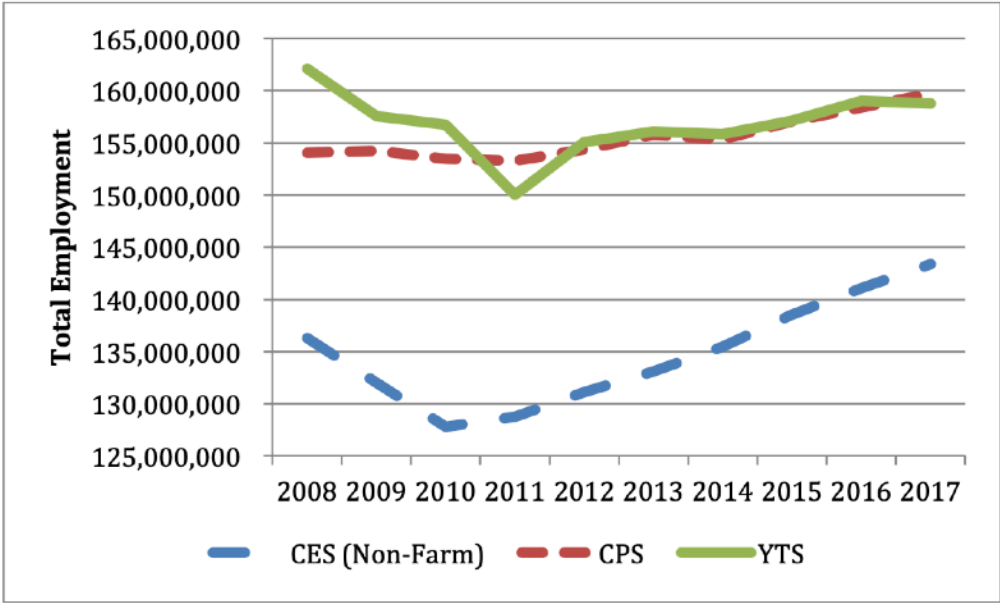
We can make several important observations based on the table. First, as stated before, CES reports far less total employment than either CPS or YTS; about 10% less for most years. Second, since 2012 YTS and CPS have reported nearly identical levels of overall employment in the economy. It is also interesting to note that during the slowdown of 2008-2011, CES and YTS showed a similarly strong decline in total employment, whereas CPS's downturn was modest. Fourth, CES shows the downturn ending in 2010, in contrast to both CPS and YTS indicating an employment bottom in 2011.

Despite their difference in total employment magnitude, the overall direction of the lines before and after 2011 are more similar between YTS and CES than they are between YTS and CPS or, more notably, between CES and CPS.

Lastly, we can see that the YTS data appears to be more sensitive to changes in the

economy than either CES or CPS. We can visually observe only slope change in the CES employment line and perhaps four small variations in slope in the CPS data, whereas the slope of the line for the YTS data appears to change nearly every year. This may be due to the fact that CES and CPS undergo revisions, adjustments, and data smoothing, whereas YTS does not. Regardless, it appears that YTS is more reflective of actual annual changes in the economy than either CES or CPS.

YTS thus seems to be the better choice for researchers looking for an employment dataset that is as encompassing as the CPS data, as reflective of downturns as the CES data, but has more annual sensitivity to changes than either CPS or CES.

CES Inconsistencies and Inaccuracies

According to Lars Vilhuber at Cornell University, there is a "widespread perception that administrative data are objective and comprehensive" (Vilhuber, 2005). Yet an audit by BLS in the late 1980s found an average error rate of 7.8 percent with a high variation across states. In 1997 (the first reported year of the YTS data) the BLS admitted that the true error rate is likely to be even higher.

Errors have been found in the handling of the data by state workers compiling CES's ES-202 data; the likely result of random coding errors. Vihuber writes that these errors seem to be persistent and uncorrected because, "none of the involved parties has strong incentives to actively search for and obtain more accurate records on an on-going basis." This, of course, is not the case with YTS's Data Axle data that is collected for business purposes, as the company has a vested commercial interest in ensuring the accuracy of the data it sells.

There may be more purposeful reasons why some firms submit false ES-202 reports. Systematic tax evasion by corporations is well documented. Corporations are obliged by law to report their number of employees and wages in quarterly ES-202 unemployment compensation filings and are subject to statutory payroll taxes based on the finlings. It is reasonable to expect that at least some businesses file ES-202 with incorrect information for tax evasion purposes, thus casting some doubt upon the accuracy of the employment statistics derived from this data.

A study by the IRS in 2001 estimated that the rate of corporate income tax noncompliance was 17 percent (Slemrod, 2007). This problem appears worse for both small and private firms, which form the majority of private-sector businesses.

For example, the IRS estimates noncompliance for corporations with less than $10 million in assets at approximately 29 percent. An independent study found that corporate tax filing deficiencies at public companies were 12.5 percent, whereas the rate at private companies was 17.1 percent.

The problems with ES-202 extend beyond errors and intentionally filed false information. Problems also occur in classifying employment when the employer 8firm experiences an ownership change; assigning workers to specific locations operated by multi-unit employers; tracking differences in operational activity at individual locations of multi-unit firms; and accounting for undocumented workers.

Each company that files an ES-202 report is assigned a unique employer identification number. If a firm changes ownership, the employer identifier associated with jobs at that establishment changes, which often is reflected in the records as a mass layoff and closure at one business and a sudden hiring at a new or existing business – even if all the workers are retained through the ownership transfer. As researchers that have studied this problem point out, "spurious apparent employer changes are known to induce biases in both employment and job flow statistics" (Abowd, et al., 2006).

ES-202 data performs poorly in assigning geographic location to employment for multi-unit businesses. Approximately 30-40 percent of workers at the state level work at companies with more than one establishment (Abowd, et al., 2006). These workers are reported under the same employer identification number regardless of location, obscuring the employment location data.

A similar problem occurs for companies that divide corporate activities across locations. For example, many multi-unit companies have separate facilities for headquarters activities, manufacturing and warehousing. All workers, regardless of their location or occupation, would be assigned the same geographic and industry code as the parent company. This makes it almost impossible for researchers to spatially detect corporate activities at multi-unit firms. In contrast, YTS tracks each establishment separately with their own unique business identifier.

To compound problems even further, it is also possible that some employers are in noncompliance by failing to declare employment of undocumented workers.

Researchers at the BLS have recognized there is a problem with accounting for

undocumented workers. As they explain,

> *"While it is an undisputed fact that these workers are part of the U.S. economy, how many there are and their relative prevalence in the labor force remains a subject of debate. Regardless of the extent to which undocumented workers are employed by U.S. businesses, though, they probably are more likely to be reflected in the CES employment figures than in the CPS." (Nardon, et al., 2003)*

The evidence suggests that ES-202 data is laced with its own errors and inconsistencies. Some of these are caused by accidental or intentional misinformation submitted on the unemployment insurance filings, while others are caused by administrative handling of the data. These observations cast doubt that either CPS or CES represents a 'gold standard' to which all other business data sources should aspire.

Conclusions

Researchers looking for datasets that track employment and establishment-level information should consider how their needs match with the quality and flexibility of all reasonably accessible sources. This paper has evaluated the BLS's CES and CPS along with the private-sector YTS sources.

We find that for annual tracking of employment data at the national level YTS and CPS are of the same reported magnitude, while CES only tracks non-farm employment. For analyzing establishment- or firm-level data the YTS and CES datasets are the only options possible because CPS does not supply that data.

When evaluating the merits of the all three sets to perform either an annual national-level employment-tracking project or to analyze specific businesses, YTS offers advantages over the BLS sources that can be summarized:

1) **Information richness**

YTS's Data Axle historic files are more information-rich than CES or CPS data. YTS allows a researcher to track changes in employment for all sectors of the economy, unlike CPS which only covers non-farm sectors. Even if accessed at the raw data file level, ES-202 information is limited to a few

variables such as the number of full-time employees, wages, industry, and business location. In contrast, Data Axle (the source of YTS information) allows subscriber access to dozens of demographic, operational, and performance variables for each establishment as needed.

2) Sensitivity

YTS has time-series employment, sales, location, and corporate structure data at the establishment level dating back to 1997. Over this period, researchers can observe more clearly annual changes in employment and other variables, whereas CPS and particularly CES shows notable data smoothing with far less change from year to year.

3) Accuracy

BLS has well-documented problems with the accuracy of their data throughout the collection, coding, and classification stages. Scholars have suggested that public-sector employees at BLS have less incentive to provide accurate information than employees at private sector firms such as Data Axle.

4) Self-employed

Start-ups with only a Founder and/or partners, and businesses that represent self-employed persons, are not picked up in the CES data. CES does not begin tracking the start-up until the Founder hires the first employee, and entirely ignores self-employed persons. With YTS, unlike with either CES or CPS, researchers can track start-ups and freelance workers and identify them as such with their proper industry codes, location, and sales as early as the first year of their operation.

5) Ease of access

Researchers that require establishment or firm level data find it much easier to gain access to YTS than CES's ES-202 filing, and there is no access at that level to CPS data. ES-202 filing data is highly restricted due to confidentiality requirements involved in its collection. Researchers wishing to access raw CES data are required to obtain permission through a long and

complicated application process. If they are allowed to use the raw data, they face further restrictions as to how their findings can be released. For example, the results must be presented at an aggregated level. Thus, using ES-202 – even when available – can dramatically inhibit timely research and limit the potential value of research findings.

6) Data flexibility

YTS's time-series business data at the establishment level allows the identification and analysis of specific businesses or subsets using variables that can change over time. CES and CPS employment data is aggregated, severely restricting its research usefulness for business-level projects. YTS allows analysts to create nearly endless subgroups, which is simply not possible with BLS data. For example, researchers may wish to compare the productivity or relocation patterns of healthcare providers in Texas against those in New York. In addition, researchers could, if needed, conduct a follow-up survey of these groups of companies with YTS data. This is possible with YTS but not with either CES or CPS.

References

Abowd, J. M., Stephens, B. E., Vilhuber, L., Anderson, F., McKinney, K. L., Roemer, M., et al. (2006). The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators. In Longitudinal Employer-Household Dynamics (p. 139). Center for Economic Studies, U.S. Census Bureau.

Birch, D. L. (1987). Job Creation in America. New York: The Free Press.
Copeland, K. R. (n.d.). Nonresponse Adjustment in the Current Employment Statistics
Survey. Retrieved 2011, from US Bureau of Labor Statistics:

Davis, S. J., Haltiwanger, J. C., & Schub, S. (1998). Job Creation and Destruction. Cambridge: The MIT Press.

Kane, Tim. (2004). Diverging Employment Data: A Critical View of the Payroll Survey. The Heritage Foundation.

Kolko, J., & Neumark, D. (2007). Business Location Decisions and Employment Dynamics in California. Public Policy Institute of California.

Neumark, D., Zhang, J., & Wall, B. (2005). Employment Dynamics and Business Relocation: New Evidence from the National Establishment Time Series. NBER Working Paper Series, 11647.

Slemrod, J. (2007, Winter). Cheating Ourselves: The Economics of Tax Evasion. Journal of Economic Perspectives, 25-48.

Nardone, T., Bowler, M., Kropf, J., Kirkland, K., & Wetrogan, S. (2003). Examining the Discrepancy in Employment Growth between CPS and CES. Federal Economic Statistics Advisory Committee.

Vlihuber, L. (2005). Adjusting Imperfect Data: Overview and Case Studies. Cornel University.

Appendix

Recent Scholars Using YTS Data

- Massachusetts Institute of Technology
- Universities of Wisconsin
- Bureau of Economic Research
- Harvard University
- The University of British Columbia
- Michigan State University
- University of Florida
- Cornell University
- UNSW Australia
- Central Michigan University
- City University of Hong Kong
- Monash University
- St. Louis University
- University of Nevada, Reno
- University of New South Wales

For YourEconomy's list of recent academic publications using YTS data, see:
Recent Academic Publications Citing YourEconomy Time Series (YTS) Data