Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

New U.S. Business Establishments: Surging or Stalling?

Dan Cao, Henry Hyatt, Toshihiko Mukoyama, Erick Sager

2026-040

Please cite this paper as:
Cao, Dan, Henry Hyatt, Toshihiko Mukoyama, and Erick Sager (2026). "New U.S. Business Establishments: Surging or Stalling?," Finance and Economics Discussion Series 2026-040. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2026.040.

New U.S. Business Establishments: Surging or Stalling?*

Dan Cao† Henry Hyatt‡ Toshihiko Mukoyama§ Erick Sager¶

June 2026

Abstract

Since the 1990s, the Bureau of Labor Statistics (BLS) has reported much more rapid growth in U.S. private sector employer establishments than has the Census Bureau – the gap reached roughly 1.6 million by 2023. Using linked BLS-Census microdata, we document two main drivers. First, a large and growing number of employers providing services to the elderly and persons with disabilities are in scope for the BLS frame but not the Census Bureau's. Second, many firms appear with substantially more establishments in the BLS frame. These discrepancies substantially affect the measured establishment size distribution and quantitative policy analysis.

JEL Codes: E24, J21, L11, O31

Keywords: establishments, multi-unit firms, concentration

*Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau or Board of Governors of the Federal Reserve System or its staff. We thank Keith Bailey, Michael Dalton, Ryan Decker, Barbara Downs, Andrew Foote, Christopher Goetz, Matthew Graham, John Haltiwanger, Erika McEntarfer, Hubert Janicki, Shawn Klimek, Kristin Sandusky, Martha Stinson, and participants at the 2024 Comparative Analysis of Enterprise Data conference, the 2024 NBER Summer Institute, the 2024 Southern Economic Association meeting, and the 2025 BLS-Census workshop for helpful comments. Yagmur Menzilcioglu, Jinxin Wu, and Jifan Wang provided superb research assistance. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Project 7515812: CBDRB-FY24-CES024-005).

†Department of Economics, Georgetown University. Email: dan.cao@georgetown.edu

‡Center for Economic Studies, U.S. Census Bureau. Email: henry.r.hyatt@gmail.com

§Department of Economics, Georgetown University. Email: tm1309@georgetown.edu

¶Board of Governors of the Federal Reserve System. Email: erick.sager@gmail.com

1 Introduction

In recent decades, the U.S. economy has experienced the rapid rise of "mega firms" which operate many business establishments. A growing literature seeks to understand the causes and consequences of this phenomenon. For example, Cao et al. (2020), Aghion et al. (2023), and Hsieh and Rossi-Hansberg (2023) highlight the importance of technological changes that make it easier for firms to add new branches, manage their establishment portfolios, and replicate low-cost technologies across their worksites. Moreover, recent studies show that the expansion of multi-establishment firms plays an important role in rising national market concentration and declining local market concentration (Rossi-Hansberg et al., 2021; Autor et al., 2023), declining labor share (Aghion et al., 2023), rising earnings inequality (Kleinman, 2023), and the transmission of local economic shocks (Giroud and Mueller, 2019).

There are two main administrative datasets that cover employer establishments in the U.S. The first is the U.S. Census Bureau's Longitudinal Business Database (LBD), which tracks employer establishments that appear in its Business Register (BR), which in turn is derived from tax records, surveys, and other data sources. A second dataset is the Quarterly Census of Employment and Wages (QCEW), which the Bureau of Labor Statistics (BLS) maintains in partnership with U.S. states, territories, and the District of Columbia.[1] Each of these datasets is associated with published aggregates. The LBD is tabulated as the Business Dynamics Statistics (BDS). The QCEW is published as tabulations with the same name and acronym, as well as through the Business Employment Dynamics (BDM). To use comparable definitions and scope, in this paper we consider the private sector only and exclude government employers.

These two data sources disagree substantially about the number of employer business establishments in the U.S. In 2023, the most recent year for which both data series (referencing March) have been published, the BDM (QCEW) reported 9.2 million private sector employer establishments, while the BDS (LBD) reported only 7.5 million. This difference of about 1.6 million establishments stems from a divergence between the series that began in the early 2000s. The difference in establishment counts between the Census and BLS business registers has been previously noted by Barnatchez et al. (2017) and Decker and Haltiwanger (2023).[2]

[1] Some papers have used the National Establishment Time Series (NETS), a private sector source of business microdata. Barnatchez et al. (2017) provide a comparison between NETS and QCEW.

[2] Earlier work by Becker et al. (2005) and Fairman et al. (2008) compared the BLS and Census establishment frames.

In this paper, we document these discrepancies, shed light on their sources, and consider their consequences. For the years 2004 to 2016, we can link the LBD with establishment-level QCEW microdata provided by all 50 U.S. states and the District of Columbia to the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program. This unique linked data source allows us to explore the details of the disagreements between the two datasets. The divergence between these two data sources arises from fundamental differences in how each statistical agency constructs its business frame. While the Census Bureau relies primarily on Internal Revenue Service (IRS) tax filings (Form 941) and periodic surveys to link establishments to firms, the BLS frame is built from state-level unemployment insurance (UI) records and the Multiple Worksite Report (MWR).

Our linked microdata analysis reveals that these distinct methodologies lead to two main drivers of the disagreement between the two datasets: disagreement in coverage and disagreement among the firms that appear in both datasets. Disagreement on coverage is mostly concentrated in one industry group, Education and Health Services (NAICS 61-62).[3] Nearly all of this disagreement is concentrated in one detailed industry, Services to the Elderly and Persons with Disabilities (624120). This industry includes home health aides and most of its employers historically (until 2013) were included by BLS in the Private Households (NAICS 814110) industry. Disagreement in other industries is mostly attributable to disagreement among the firms that appear in both datasets. For 2016, there are 167 thousand firms that appear as single-unit in the LBD but as multi-unit in the QCEW, observed via the LEHD Employer Characteristics File (ECF), where they operate a total of 621 thousand establishments. About a third of these firms (57 thousand) are in the Professional and Business Services (NAICS 54-56) industry supersector.

These findings highlight a remarkable difference in the measured complexity of U.S. firms between the LBD and the LEHD ECF (QCEW). For instance, our microdata analysis reveals that the LEHD ECF identifies more than twice as many firms as multi-unit as the LBD. Overall, the LEHD ECF characterizes the economy as surging in small and decentralized establishments, while the LBD characterizes the economy as still dominated by fewer but significantly larger ones. These findings also substantially alter the interpretation of the U.S. establishment size distribution. We also demonstrate that this discrepancy yields different quantitative conclusions when used in firm-level policy evaluation.

[3]Industry definitions are based on the North American Industry Classification System (NAICS). In this paper, we primarily use industry "supersectors" which are themselves aggregations of the roughly two-digit NAICS sectors.

This paper proceeds as follows. Section 2 describes how the BLS and Census Bureau establishment frames are collected. Section 3 describes differences in the published establishment counts from the BDS and BDM. Section 4 describes our administrative microdata and the empirical results we obtain from it. Section 5 considers the implications of our findings for interpreting size distributions among U.S. businesses and policy analysis. Section 6 concludes.

2 Two establishment frames

Two U.S. government agencies, BLS and the Census Bureau, maintain comprehensive lists of U.S. employer business establishments. This section provides an overview of each of the two establishment frames. The two frames are detailed in Appendix A.

2.1 Census Bureau

The LBD is an enhanced version of the Census Bureau Business Register (BR), a longitudinal database of business establishments and firms with coverage starting in 1976. The BR, in turn, is derived from administrative records data from the IRS supplemented with survey, census, and administrative records data. The main data sources include the IRS, the Social Security Administration (SSA), and BLS. A key enhancement to these IRS records is provided by what has historically been known as the Company Organization Survey (COS), which has collected information on how businesses are structured. The COS is administered more frequently to larger businesses, and businesses with 500 or more employees are sampled with certainty.

The LBD is constructed by linking annual cross-sectional files from the Census BR to provide a longitudinal history for each establishment. Because the LBD records every employer establishment in the BR, it also allows researchers to identify the parent firm associated with each establishment. The Census Bureau regularly produces several public-use datasets derived from the LBD, most notably the Business Dynamics Statistics (BDS).

2.2 BLS

The BLS establishment frame is the result of a partnership with U.S. states, territories, and the District of Columbia. When a private sector business hires employees in a covered industry (which includes

nearly all employees) in a given state, it must notify the state government of its new hires and start paying unemployment insurance (UI) taxes on them. These governments in turn provide these reports of employment and wage bills, as well as each establishment's location and industry, to the BLS each quarter. These QCEW microdata are the basis for the QCEW report and are derived from the UI accounting system in that state.

Establishment-level QCEW microdata are restricted to the projects approved by the BLS. A secondary and increasingly common path for researchers to analyze QCEW-based records in conjunction with Census data is through the Longitudinal Employer-Household Dynamics (LEHD) program. Our microdata analysis is based on the LEHD ECF. Because the LEHD program is housed within the Census Bureau, it performs its own internal crosswalk to link these state-reported UI records to the Census Bureau's Business Register. This unique infrastructure allows researchers within the FSRDC environment to compare the BLS-derived establishment counts in the ECF directly against the Census-derived counts in the LBD for the same underlying firms. The QCEW program publishes a wide array of high-frequency data, but for researchers studying business entry and exit, the primary public-use product is the Business Employment Dynamics (BDM).

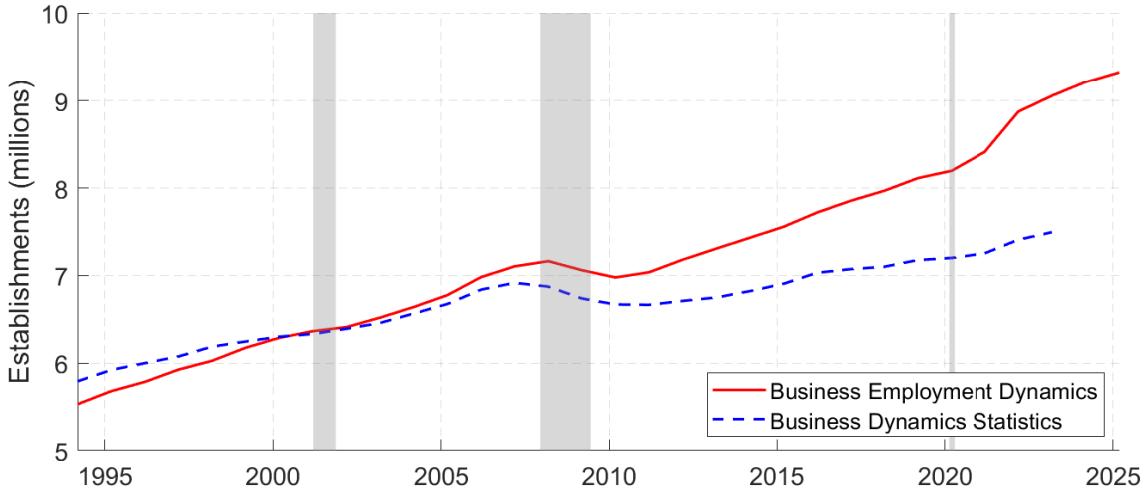
3 Differences in establishment counts

Before analyzing our microdata, first consider publicly-available tabulations of these two establishment frames: BDS for the LBD and BDM for the QCEW. For comparability, these were aggregated to roughly the NAICS supersector level (a residual category combines Agriculture, Forestry, Fishing, and Hunting (NAICS 11) and Other Services, Excluding Public Administration (NAICS 81)), and we consider private sector employment only. This comparison motivates our analysis using LBD and QCEW microdata in the next section.

3.1 Differences in high-level totals

The number of establishments is quite different in the BDS, compared to the BDM. See Figure 1, which compares the published aggregates from the BDS and BDM. In 2023, the BDM reported about 9.2 million employer business establishments, while the BDS reported 7.4 million. This difference is the result of a gap that has widened sharply over time. In the 1990s, the BDS reported more

Figure 1: Total U.S. establishments, March, 1994-2025



Notes: The Business Employment Dynamics (BDM) is derived from the Quarterly Census of Employment and Wages (QCEW) and published by the Bureau of Labor Statistics. The Business Dynamics Statistics (BDS) is derived from the Longitudinal Business Database (LBD) and published by the Census Bureau. Shaded regions indicate recessions.

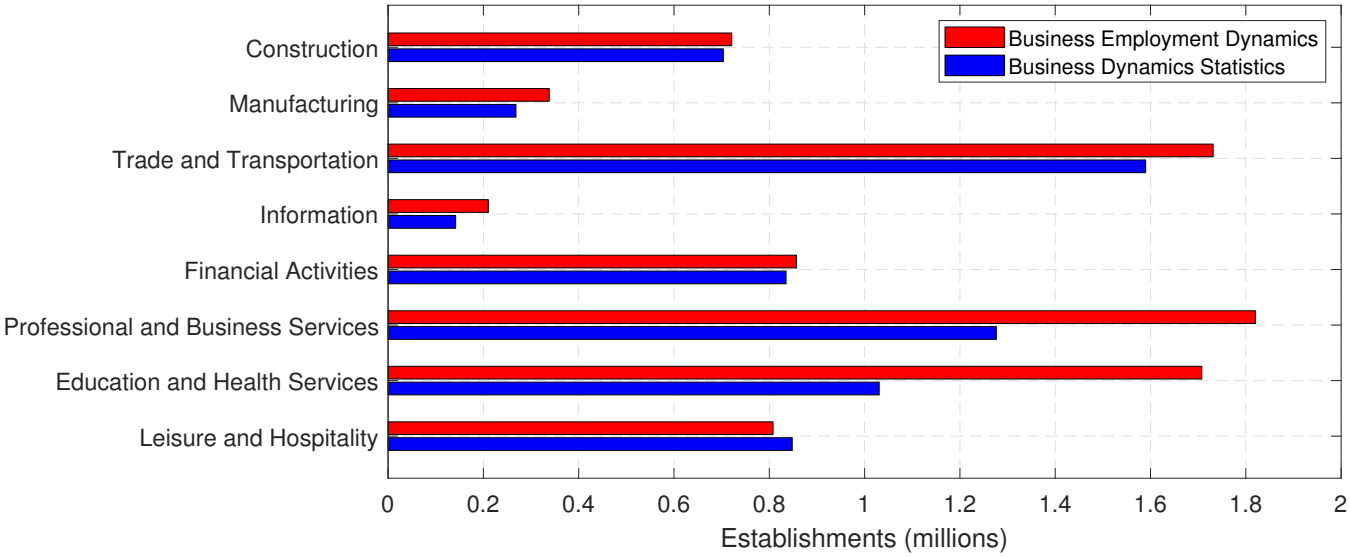
establishments than the BDM. In the early 2000s, the two datasets agreed on the number of establishments at about six and a half million. However, in the late 2000s, the establishment total in the BDS approached seven million before falling during and after the 2007-2009 recession and showed only modest growth in the decade that followed. In contrast, the BDM indicates that the U.S. gained more than one million establishments during the 2010s.

This widening divergence in the number of establishments between the BDS and BDM is not reflected in the employment data from each source. The employment counts are compared in Appendix B.1. In contrast to the number of establishments, BDS reports greater total employment than the BDM, by as much as five percent, and the disparity is persistent. These findings suggest that differences in published aggregates may be related to how workers are allocated across establishments in each data source.

3.2 Cross-industry differences

We now consider differences in the total number of establishments by industry in the BDS and BDM. Results at the NAICS supersector level are shown in Figure 2 for 2023. The greatest difference is found in the number of establishments in Education and Health Services (61-62), where the BDM reports more than 1.7 million establishments, compared with about 1.0 million in the BDS. The second-

Figure 2: U.S. private sector establishments, by industry, March 2023.



Notes: The Business Employment Dynamics (BDM) is derived from the Quarterly Census of Employment and Wages (QCEW) and published by the Bureau of Labor Statistics. The Business Dynamics Statistics (BDS) is derived from the Longitudinal Business Database (LBD) and published by the Census Bureau.

largest difference is found in Professional and Business Services (54-56), in which the BDM reports about 1.7 million establishments, compared with over 1.2 million in the BDS. These two supersectors alone account for about two-thirds of the difference in the number of establishments in the BDS and BDM. Differences in the remaining six supersectors are smaller but favor the BDM, with the exception of Leisure and Hospitality (71-72). These considerable cross-industry differences motivate our subsequent analysis of the microdata that underlies each of these data sources.

4 What drives these differences?

Motivated by these facts, we conduct further analysis using the QCEW and BDS microdata. We obtained access to establishment-level QCEW microdata that is part of the LEHD ECF for all fifty U.S. states plus the District of Columbia for the years 2004-2016, which we have linked with the LBD.

Broadly speaking, our linked dataset allows us to distinguish differential coverage from organizational structure as recorded in each data source. By coverage, we mean that certain firms may be lacking in one or the other data source. By organizational structure, we refer to the establishment composition of a particular firm, which may be recorded differently in the LBD and QCEW.

4.1 A QCEW-LBD matched dataset

We utilize two microdata sources: the LBD and the LEHD ECF, the latter of which contains establishment-level QCEW data. The fundamental data bridge between the ECF and LBD was developed by Haltiwanger et al. (2014), and is based on a business's Employer Identification Number (EIN). Each ECF establishment that links to the LBD is assigned its firm identifier from the Census Bureau BR. For ECF establishments that do not link to the LBD, the EIN is assigned as a substitute. Each dataset is restricted to the private sector, and agriculture is removed from each data series. For the LBD, establishments are only included if they are in scope for the CBP. For the ECF, we only reference the March employment, for comparability with the LBD.

The analysis dataset includes characteristics of each establishment, including its firm identifier, total employment, industry, as well as state and county (geography)—as measured separately in the ECF and LBD. Our analysis dataset utilizes the LBD industry codes developed by Fort and Klimek (2016). The establishment-level data is aggregated to the firm level by summing the total number of establishments, as well as summing employment in each establishment. For comparability with published totals, industry-specific analyses restrict to the set of establishments operating in a particular industry before aggregating to the firm level.

4.2 Data analyses

4.2.1 Comparisons among particular firms

We start with the question of whether the divergence in the growth in the number of establishments between ECF and LBD is driven by firms that are missing in the LBD, and firms that appear with more establishments in the ECF than the LBD. Our initial results for 2004 and 2016, which pool all industries together, are shown in Table 1. Columns (1) through (3) report results for firms that appear in both the ECF and LBD and agree on their single-establishment vs. multi-establishment status. Matched firms associated with a single establishment in both the ECF and LBD do not contribute to the divergence by definition, and there were 4.1 million such firms in 2004 and 4.2 million in 2016. Firms with multiple establishments in both the LBD and ECF exhibited greater growth in the latter data source. In 2004, the LBD had 61 thousand more establishments associated with these firms, while in 2016 the ECF had 75 thousand more.

Table 1: Establishment counts (thousands), by single- vs. multi-unit agreement for particular firms between the LBD and ECF, by year

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Single-unit	Multi-unit in both		Multi-unit in ECF		Multi-unit in LBD		Only in one source	
	in both	ECF	LBD	ECF	LBD	ECF	LBD	ECF	LBD
				Agreement in 2004					
Total	4,160	1,351	1,412	423	119	61	186	788	790
Difference	0		-61		304		-126		-2
				Agreement in 2016					
Total	4,223	1,828	1,753	621	167	52	164	1,162	822
Difference	0		75		454		-112		340
				Change from 2004 to 2016					
Difference	0		136		150		14		342
Share	0.0%		21.2%		23.4%		2.1%		53.3%

Notes: Authors' tabulations of the LEHD ECF and the LBD. The ECF statistics are derived from the March employment counts reported as part of the QCEW. Counts by category may not sum to the total due to rounding. "Share" in the last row is the percentage of the entire penulatimate row represented by its corresponding "Difference" value.

The remainder of the columns (4) through (9) of Table 1 report results for firms that appear differently in the ECF and LBD. Columns (4) through (7) report results for firms that appear in both the ECF and LBD but disagree on whether they have multiple establishments vs. a single one. The LBD has more so-defined single establishment firms (119 thousand in 2004, 167 thousand in 2016) than the ECF (61 thousand in 2004, 52 thousand in 2016). There is substantial growth in the number of establishments associated with so-defined multi-establishment ECF firms: from 423 thousand in 2004 to 621 thousand in 2016. An increasing number of firms appear with multiple establishments in the ECF but with a single establishment in the LBD, and such firms make a substantial contribution to the disagreement in the number of establishments in the ECF and LBD.

The final columns (8) and (9) of Table 1 report results for firms that appear only in one data source. In 2004, the number of establishments associated with such firms was similar in each data source: 788 thousand in the ECF, and 790 thousand in the LBD. The number of establishments of such firms was increasing faster in the ECF, which had about 1.2 million such establishments in 2016, while LBD had only 822 thousand.

This exercise yields insights into the nature of the increasing disagreement in the number of establishments between the ECF and LBD. Differences in the set of firms explain slightly more than half (53.3%) of the divergence between the ECF and LBD between 2004 and 2016. In other words, about half of the divergence between the ECF and LBD is due to firms that only appear in the ECF. The remainder of the divergence is due to firms that appear in both data sources but have more establishments in the ECF. Almost one-quarter (23.4%) of the increase is attributed to firms that are single establishment in the LBD but multi-establishment in the ECF. About one-fifth (21.2%) is due to the ECF having more establishments among firms that are multi-establishment in both data sources. The declining number of firms that are single-establishment in the ECF but multi-establishment in the LBD explains a small (2.1%) amount of this divergence.

4.2.2 Industry details

The analysis of published BDS and BDM aggregates highlighted disagreements in particular industries, especially two supersectors: Education and Health Services (NAICS 61-62), as well as Professional and Business Services (NAICS 54-56). Our dataset allows us to apply the previous accounting exercise that distinguishes between different firms vs. differently-appearing firms, at the industry

level. It also allows us to drill down into more detailed industries and highlight the detailed industries that might account for a disproportionate share of the disagreement between the BDS and BDM. Below, industry-specific analyses assign a dominant (employment-weighted) industry to each firm. Table 2 describes the results for two of our ten supersectors and three detailed (6-digit NAICS) industries. These two supersectors are the Professional and Business Services (54-56) and Education and Health Services (61-62), where the greatest differences between the ECF and LBD are found. Table B4 in Appendix B.3 records all ten supersectors. Below, we focus on these two supersectors.

Education and Health Services Consider the Education and Health Services (61-62) supersector first (Table 2). The ECF has a greater number of establishments in firms in this sector, and this is entirely driven by differences among missing firms. The ECF has 607 thousand establishments in firms in this supersector that do not appear in the LBD. The LBD, by contrast has only 100 thousand establishments that do not appear in the ECF. This net differential of about 508 thousand accounts for all of the difference between the ECF and LBD. Among firms that are present in the ECF and LBD, the LBD has 15 ($= 31 + 21 - 37$) thousand more establishments.

Our analysis of detailed industries in Table 2 sheds further light on the difference in the Education and Health Services (61-62) supersector. There is one 6-digit NAICS industry that accounts for all of the differential between the ECF and LBD: Services for the Elderly and Persons with Disabilities (624120). There are 539 thousand establishments associated with firms that are present in the ECF but missing in the LBD, while there are only 6 thousand establishments for which the reverse is true. This differential of 532 thousand (after rounding) is even greater than that of its broader supersector, which is only 508. This industry consists of establishments that provide "day care, non-medical home care or homemaker services, social activities, group support, and companionship."[4] As the U.S. population ages, there is tremendous growth in the number of firms that provide services to the elderly and those with disabilities outside of a residential facility such as a nursing home. This growth drives much of the establishment count in the ECF but is mostly absent from the LBD.

While it is well known that the U.S. population is ageing and that the number of residential establishments for the elderly (e.g. nursing homes) is increasing rapidly, nursing homes do not drive the difference in the number of establishments between the ECF and LBD. Instead, it is the increasing

[4]The quotation is from the U.S. Census Bureau website https://www.census.gov/naics/?input=624120&year=2017&details=624120 (accessed: March 21, 2024).

number of firms that offer at-home care for the elderly, or day-care services, which appear as a rapidly growing industry in the ECF but a marginal one in the LBD. This highly concentrated phenomenon accounts for on the order of half of the divergence between the ECF and LBD. This difference is likely due to how employment by private households enters (or does not enter) into the establishment frames maintained by BLS and the Census Bureau. The core of the Census Bureau establishment frame is IRS Form 941, whose official name is "Employer's QUARTERLY Federal Tax Return." Private households that employ home health aides do not file IRS Form 941 but instead file Schedule H "Household Employment Taxes" along with IRS Form 1040 "U.S. Individual Income Tax Return." When a home health aide is employed by a private household, they may be covered by unemployment insurance and therefore enter into the BLS establishment frame. There appear to have been about a half million such employer establishments in 2016.[5]

Professional and Business Services We now consider the Professional and Business Services (54-56) supersector (Table 2). There are a number of firms that appear in only one data source but not the other. In 2016, 99 thousand firms appear in only the ECF, while 139 thousand appear only in the LBD. This moderate net difference of 40 thousand favors the LBD.

Larger differences appear in the number of establishments associated with firms present in both data sources, and the ECF tends to have more establishments in such firms than does the LBD. Firms with multiple establishments in both data sources have 283 thousand establishments in the ECF but only 178 thousand in the LBD. There are furthermore 57 thousand single-establishment firms in the LBD that have a total of 233 thousand establishments in the ECF. There are also 5 thousand firms that have a single establishment in the ECF but have a total of 16 thousand establishments in the LBD. This net differential of 271 ($= 176 + 105 - 10$) thousand establishments favors the ECF.

This phenomenon is not driven by a single 6-digit NAICS industry. Table 2 presents detailed statistics for two 6-digit NAICS industries that, together, account for about one-third of the total differential in the Professional and Business Services (54-56) supersector: Custom Computer Programming Services (541511) and Computer Systems Design Services (541512). Differences between the ECF and LBD for these two industries are driven almost entirely by firms that appear in both data sources (columns (8) and (9) contribute almost nothing to the differential). There are also virtually no

[5] In principle, a home health aide may be an employee for multiple private households concurrently, in which case the ECF may represent an overcount of such employer establishments relative to other service-providing industries in which the establishment record coincides with the service-provider instead of the consumer.

Table 2: Establishment counts (thousands), by single- vs. multi-unit agreement for particular firms between the LBD and ECF, 2016, selected industries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Single-unit	Multi-unit in both		Multi-unit in ECF		Multi-unit in LBD		Only in one source	
	in both	ECF	LBD	ECF	LBD	ECF	LBD	ECF	LBD
Education and Health Services (61-62)									
Total	568	205	226	47	11	11	42	607	100
Difference	0	-21		37		-31		508	
Services for the Elderly and Persons with Disabilities (624120)									
Total	25	7	9	3	0	0	2	539	6
Difference	0	-2		3		-1		532	
Professional and Business Services (54-56)									
Total	793	283	178	233	57	5	16	99	139
Difference	0	105		176		-10		-40	
Custom Computer Programming Services (541511)									
Total	29	25	9	37	8	0	0	8	9
Difference	0	17		29		0		-1	
Computer Systems Design Services (541512)									
Total	32	29	9	36	7	0	0	6	6
Difference	0	20		29		0		0	

Notes: Authors' tabulations of the LEHD ECF and the LBD. The ECF statistics are derived from the March employment counts reported as part of the QCEW. Counts by category may not sum to the total due to rounding.

firms that appear as a single establishment in the ECF but with multiple establishments in the LBD, as shown in columns (6) and (7). In Custom Computer Programming Services (541511), firms that appear with multiple establishments in the ECF have 48 ($= 17 + 29$) thousand more establishments in the ECF than the LBD. In Computer Systems Design Services (541512), there are 49 ($= 20 + 29$) thousand more such establishments. These two 6-digit NAICS industries therefore account for about 35.8% ($\approx (49 + 48)/271$) of the differential in the Professional and Business Services (54-56) supersector. We therefore conclude that differences in the number of establishments in the ECF versus LBD are spread diffusely throughout this supersector. The contribution of these two high-tech, high-growth sectors to the divergence in the number of establishments between the ECF and LBD is also noteworthy.

5 Implications for economic analysis

The differences in how the Census Bureau and the BLS capture business units lead to two distinct characterizations of the U.S. economy. In this section, we explore the implications for the measured prevalence of multi-unit firms and the resulting establishment size distributions across industries.

5.1 The distribution of the number of establishments across firms

The differences in how the Census Bureau and the BLS link establishments to their parent organizations lead to two distinct characterizations of the U.S. business landscape. The choice of data source fundamentally changes the measured prevalence of multi-unit firms. This distinction is critical for economic analysis, as the boundary of the firm dictates our understanding of production technology, market concentration, labor market resilience and the spatial placement of establishments (Hsieh and Rossi-Hansberg, 2023; Brynjolfsson et al., 2023; Oberfield et al., 2024; McElheran et al., 2026; Argente et al., 2025).

Across nearly every industry supersector, the ECF identifies a significantly higher share of firms as multi-unit than the LBD.[6] From an industrial organization perspective, this result suggests that the U.S. economy may be more concentrated than Census-based records imply. If these units are categorized as independent "single-unit" firms, researchers may overstate the degree of competitive

[6]Table B5 in Appendix B.3 compares the share of multi-unit firms in firm count, establishment count, and employment for ECF and LBD.

entry; however, if they are recognized as branches, as in the ECF, the narrative shifts toward one of corporate scaling and the dominance of integrated organizations.

This discrepancy also has implications for how we characterize production technology. Multi-unit firms often leverage non-rival internal inputs—such as proprietary software, centralized logistics, and brand capital—that can be deployed across many locations at low marginal cost. In high-growth, service-oriented industries where the ECF shows the largest multi-unit advantage, such as Information (117.5% difference) and Professional and Business Services (119.3% difference), the data suggest an economy that has aggressively adopted information technologies to manage decentralized production. By identifying these links, the ECF allows for a more accurate assessment of economies of scale that are invisible when establishments are viewed in isolation.

Furthermore, the higher share of multi-unit employment in the ECF alters the interpretation of labor market dynamics. Multi-unit firms often operate internal labor markets that can shield workers from local economic shocks through cross-location reallocation. If the LBD undercounts these multi-unit structures, it likely overestimates the fragility of the employment relationship by mislabeling protected "branch" employees as vulnerable workers at independent firms.

5.2 The size distribution of the BDS vs. BDM

The discrepancies in coverage and organizational linkage culminate in a significant divergence in the measured size distribution of U.S. establishments. As documented in Table 3, the BDM and BDS provide conflicting accounts of where economic activity is concentrated, with the gap between the two frames following a distinct and systematic pattern across size classes.

The most striking divergence occurs at the bottom of the size distribution. For establishments with 1–19 employees, the BDM reports 7.1 million establishments across the eight supersectors in 2023, representing a 23.8% surplus over the 5.6 million establishments recorded in the BDS. This "micro-establishment surge" is nearly universal across service industries, being most pronounced in Education and Health Services (a 58.2% proportionate difference) and Information (46.0%). From a technological perspective, this surplus in the BDM suggests an economy that is rapidly decentralizing, perhaps reflecting a shift toward smaller, more agile work sites or the proliferation of specialized service providers that fall outside the Census Bureau's IRS-based frame.

However, as establishment size increases, this relationship reverses. While the BDM captures

Table 3: Total establishments (thousands), by establishment size (number of employees), 2023

		1-19		20-99		100-499		500+	
		BDM	BDS	BDM	BDS	BDM	BDS	BDM	BDS
Construction (23)	Total	644.2	631.4	67.6	62.7	8.7	8.6	0.5	0.7
	Difference	*2.0%*		*7.4%*		*1.2%*		*-28.9%*	
Manufacturing (31-33)	Total	241.1	179.2	71.0	63.3	22.9	22.6	3.1	3.3
	Difference	*29.5%*		*11.5%*		*1.4%*		*-5.4%*	
Trade and Transportation (42, 44-45, 48-49)	Total	1,475.7	1,331.9	209.2	226.3	44.0	47.7	2.6	3.3
	Difference	*10.2%*		*-7.9%*		*-8.2%*		*-25.4%*	
Information (51)	Total	188.2	117.8	17.7	17.9	3.7	5.0	0.7	1.0
	Difference	*46.0%*		*-1.2%*		*-28.5%*		*-36.1%*	
Financial Activities (52-53)	Total	798.4	778.1	48.7	46.0	8.4	9.1	1.5	2.0
	Difference	*2.6%*		*5.7%*		*-7.0%*		*-27.4%*	
Professional and Business Services (54-56)	Total	1,640.8	1,110.5	144.3	126.7	30.8	32.9	4.0	6.2
	Difference	*38.5%*		*12.9%*		*-6.4%*		*-42.7%*	
Education and Health Services (61-62)	Total	1,530.8	840.9	144.9	155.1	27.5	29.4	4.4	5.2
	Difference	*58.2%*		*-6.8%*		*-6.7%*		*-17.6%*	
Leisure and Hospitality (71-72)	Total	571.1	590.5	223.6	243.4	12.4	13.3	0.7	0.8
	Difference	*-3.3%*		*-8.5%*		*-6.8%*		*-15.4%*	
Total, eight supersectors	Total	7,090.2	5,580.3	926.8	941.4	158.4	168.5	17.5	22.5
	Difference	*23.8%*		*-1.6%*		*-6.1%*		*-25.1%*	

Notes: The Business Employment Dynamics (BDM) is derived from the Quarterly Census of Employment and Wages (QCEW) and published by the Bureau of Labor Statistics. The Business Dynamics Statistics (BDS) is derived from the Longitudinal Business Database (LBD) and published by the Census Bureau.

significantly more small establishments, the BDS consistently identifies a more robust presence of large-scale infrastructure. For the largest category (500+ employees), the BDS reports 25.1% more establishments than the BDM. This deficit in the BDM is most severe in Professional and Business Services (-42.7%) and Information (-36.1%).

This creates a pervasive distributional skew in 2023: the BDM characterizes the economy as a vast sea of micro-establishments, while the BDS emphasizes the continued importance of large-scale employment centers. Researchers must be mindful that the choice of data source may fundamentally shift the measured balance of the economy. One source identifies a surge in tiny, perhaps decentralized establishments, while the other identifies a landscape still dominated by significantly larger work sites.

5.3 Implications for quantitative policy analysis

Establishment size distributions have been extensively used in the quantitative evaluations of economic policies. To demonstrate how the measurement of establishments matter for the economic analysis, we perform a quantitative evaluation of a firing tax, similarly to Hopenhayn and Rogerson (1993). The purpose of the experiment is to show that the measurement of establishments matters for drawing quantitative conclusions from a model experiment, rather than to perform a novel model analysis, and as such, our model departs only minimally from Hopenhayn and Rogerson (1993). The details of the model description and the results are described in Appendix C.

The key to our analysis is the model calibration. We compare two quantitative implementations of the model with different calibrations. We assume the log establishment total factor productivity to follow an AR(1) process, as in Hopenhayn and Rogerson (1993), and estimate the values of the persistence and shock dispersion parameters using the information from establishment-size distribution. In one version of the model (BDM calibration), we use the establishment-size distribution from the BDM, drawn from the "Total" row in Table 3. In the other version (BDS calibration), we use the establishment-size distribution from the BDS.

As is described in detail in Appendix C, the outcome of the firing-tax exercise is meaningfully different across the two calibrations. We consider a firing tax equal to one year's wages. In the BDM version, aggregate output declines by 5.0%, aggregate labor declines by 3.6%, and labor productivity declines by 1.4%. In the BDS version, the corresponding numbers are 3.7%, 2.8%, and 0.9%. By using the BDS calibration instead of the BDM calibration, the policy effects on these variables are

underestimated by about one quarter to one third.[7] Thus, by employing the establishment-size distributions from different data sources, substantially different quantitative conclusions can be drawn for a policy evaluation.

6 Conclusion

We have considered the widening divergence between the number of establishments in the BLS and Census Bureau establishment frames in recent decades. In 2023, published aggregates that use the Census Bureau frame reported 7.4 million business establishments but those that use the BLS frame reported 9.2 million. We considered this divergence using both published aggregates and a unique set of linked LBD-ECF microdata.

Our analysis of published aggregates from the BDS and BDM pointed to segments of the U.S. economy where the number of establishments is especially different. Two supersectors stood out: Professional and Business Services (54-56) and Education and Health Services (61-62). We also found the most difference in the number of establishments at the extremes of the establishment size distribution. The smallest (1-19) employee establishments are more numerous in the BDM while the largest (500+) are more common in the BDS. The reasons for divergence in our two leading supersectors are quite different. Education and Health Services (61-62) is driven by a relatively narrow industry: firms that provide at-home or day care services for the elderly. This industry is growing rapidly in the ECF but its employers are largely absent from the LBD. The Professional and Business Services (54-56) supersector largely agrees on which firms are operating, but the ECF shows these firms with far more establishments than in the LBD. Most industries with a substantial divergence between the ECF and LBD are more similar to Professional and Business Services (54-56), with disagreement driven by a set of firms that is common to the two datasets.

These discrepancies have significant implications for the analysis of the U.S. aggregate economy. First, the two data sources present very different views on firm-level complexity. Across all supersectors, the ECF identifies a much higher share of firms as multi-unit than the LBD. Second, the discrepancies lead to a distinct distributional skew at the establishment level. While the BLS-based BDM reports a surge in micro-unit activity (1–19 employees), the Census-based BDS continues to

[7] The relative and absolute losses are comparable to those reported in Berlingieri et al. (2025), who evaluate policy effects in the Hopenhayn and Rogerson (1993) model under alternative specifications of the firm growth distributions (e.g, parameteric versus non-parametric).

report a higher prevalence of large-scale establishments (500+ employees). In addition, we demonstrated that the choice of dataset matters for quantitative policy analysis.

These findings suggest that the choice of data source is not neutral. Understanding these differences is essential for accurately characterizing the evolution of the U.S. business landscape and for evaluating policies related to market concentration and labor market dynamics.

References

Abowd, J., B. Stephens, L. Vilhuber, F. Andersson, K. McKinney, M. Roemer, and S. Woodcock (2009). The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators. In T. Dunne, J. B. Jensen, and M. J. Roberts (Eds.), *Producer Dynamics: New Evidence from Micro Data*, Volume 68, Chicago, IL, pp. 149–230. Studies in Income and Wealth: University of Chicago Press.

Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2023). A Theory of Falling Growth and Rising Rents. *The Review of Economic Studies 90*(6), 2675–2702.

Argente, D., S. Moreira, E. Oberfield, and V. Venkateswaran (2025). Scalable Expertise: How Standardization Drives Scale and Scope. Working Paper 34160, National Bureau of Economic Research.

Autor, D., C. Patterson, and J. Van Reenen (2023, April). Local and National Concentration Trends in Jobs and Sales: The Role of Structural Transformation. Working Paper 31130, National Bureau of Economic Research.

Barnatchez, K., L. D. Crane, and R. A. E. Decker (2017). An Assessment of the National Establishment Time Series (NETS) Database. Working Paper.

Becker, R., J. Elvery, L. Foster, C. Krizan, S. Nguyen, and D. Talan (2005). A Comparison of the Business Registers Used by the Bureau of Labor Statistics and the Bureau of the Census. Working Paper.

Berlingieri, G., N. Jaimovich, S. Terry, and N. Vincent (2025). The Empirical Distribution of Firm Dynamics and Its Macro Implications. Working Paper.

Brynjolfsson, E., W. Jin, G. Petropoulos, and X. Wang (2023). Information Technology, Firm Size, and Industrial Concentration. Working Paper 31065, National Bureau of Economic Research.

Cao, D., H. Hyatt, T. Mukoyama, and E. Sager (2020). Firm Growth through New Establishments. GCER Working Paper 20-20-02.

Chow, M. C., T. C. Fort, C. Goetz, N. Goldschlag, J. Lawrence, E. R. Perlman, M. Stinson, and T. K. White (2021). Redesigning the Longitudinal Business Database. Working Paper 28839, National Bureau of Economic Research.

Decker, R. and J. Haltiwanger (2023). Surging Business Formation in the Pandemic: Causes and Consequences? *Brookings Papers on Economic Activity*, 249–302.

Fairman, K., L. Foster, C. Krizan, and I. Rucker (2008). An Analysis of Key Differences in Micro Data: Results from the Business List Comparison Project. Working Paper.

Fort, T. and S. Klimek (2016). The Effects of Industry Classification Changes on US Employment Composition. Working Paper.

Giroud, X. and H. M. Mueller (2019). Firms' Internal Networks and Local Economic Shocks. *American Economic Review 109*(10), 3617–49.

Haltiwanger, J., H. Hyatt, E. McEntarfer, L. Sousa, and S. Tibbets (2014). Firm Age and Size in the Longitudinal Employer-Household Dynamics Data. Working Paper.

Hopenhayn, H. and R. Rogerson (1993). Job Turnover and Policy Evaluation: A General Equilibrium Analysis. *Journal of Political Economy 101*(5), 915–938.

Hsieh, C.-T. and E. Rossi-Hansberg (2023). The Industrial Revolution in Services. *Journal of Political Economy Macroeconomics 1*(1), 3–42.

Kleinman, B. (2023). Wage Inequality and the Spatial Expansion of Firms. Working Paper.

McElheran, K., M.-J. Yang, Z. Kroff, and E. Brynjolfsson (2026). The Adoption of Industrial AI in America. *AEA Papers and Proceedings 116*, 20–25.

Oberfield, E., E. Rossi-Hansberg, P.-D. Sarte, and N. Trachter (2024). Plants in space. *Journal of Political Economy 132*(3), 867–909.

Rossi-Hansberg, E., P.-D. Sarte, and N. Trachter (2021). Diverging Trends in National and Local Concentration. *NBER Macroeconomics Annual 35*, 115–150.

U.S. Census Bureau (2023). Statistics of U.S. Businesses Methodology. https://www.census.gov/programs-surveys/susb/technical-documentation/methodology.html. Accessed: 2024-11-06.

Appendix

A Details of two establishment frames

A.1 Census Bureau

A.1.1 Construction of the Census Bureau Business Register

The Census Bureau Business Register (BR) is derived from administrative records data from the IRS supplemented with survey, census, and administrative records data.[8] The most important use of the Census BR is to serve as the frame for the Economic Census: a major data collection effort that occurs every five years to ensure accurate measurement of total output. The main data sources include the IRS, the Social Security Administration (SSA), and BLS. IRS data relevant for the BR include business names, mailing addresses, physical addresses, industry codes, and payroll tax data. The SSA provides monthly updates of industry codes for newly assigned EINs, and BLS provides quarterly updates of industry codes from unemployment insurance records. Updates to legal form of organization (i.e., corporation, partnership, sole proprietorship, etc.) can come from IRS, BLS, or SSA. Administrative records for employer businesses are provided at the EIN level. For single-establishment firms, this is equivalent to the establishment and the firm. However, since multiple-establishment firms can report for multiple establishments under one EIN, establishment-level updates to MU data must come either from survey responses or imputation.

At the core of these records is IRS Form 941 which reports employment and payroll for each employer business with an Employer Identification Number (EIN). Employers who do not file IRS Form 941 are beyond the scope of the Census Bureau establishment frame. Such employers include farms, who file IRS Form 943. More importantly in what follows, the Census Bureau establishment frame excludes private households who pay UI payroll taxes and issue W-2s. These include households who employ nannies and home health aides. These households do not file IRS Form 941 but instead file IRS Form 1040 Schedule H to record their payroll expenses. In order to hire an employee and pay necessary payroll taxes to the federal government, an employer must have an EIN.

A key enhancement to these IRS records is provided by what has historically been known as

[8] For additional details, see the discussion at https://www.census.gov/programs-surveys/bds/about.html and Chow et al. (2021).

the Company Organization Survey (COS), which has collected information on how businesses are structured.[9] The COS is administered to larger businesses with greater frequency, and businesses with 500 or more employees are sampled with certainty (U.S. Census Bureau, 2023). The BR defines firms on the basis of operational control, determined primarily through the COS, and groups establishments belonging to the same firm in a particular year under a common firm (alpha) identifier. The COS also allows the Census Bureau to aggregate firms whose activities cover multiple EINs into a single organization.

A.1.2 Establishment-level Census microdata

The LBD is an enhanced version of the Census BR, a longitudinal database of business establishments and firms with coverage starting in 1976. The LBD is constructed by linking annual cross-sectional files from the Census BR to provide a longitudinal history for each establishment. The linkage process makes use of numeric establishment identifiers as well as probabilistic name and address matching. The linkage process allows the tracking of net employment changes at the establishment level, which in turn allows the estimation of jobs gained at opening and expanding establishments and jobs lost at closing and contracting establishments. The microdata includes data on total employment (in March of a particular year), as well as total annual payroll. Each establishment is associated with a particular industry and geography.

Access to the LBD is restricted to researchers with Special Sworn Status through the Federal Statistical Research Data Center (FSRDC) network. The FSRDCs are secure facilities managed by the Census Bureau that allow for the analysis of sensitive, non-public microdata while ensuring strict adherence to the confidentiality requirements of Title 13 and Title 26 of the U.S. Code. Within this environment, researchers can access the LBD to track individual establishments over time using a unique establishment-level identifier. Because the LBD records every employer establishment in the BR, it also allows researchers to identify the parent firm associated with each establishment, facilitating the study of firm-level dynamics and multi-unit organizational structures.

[9]The COS was replaced by the Report of Organization, and, at the time of this writing, the historic functions of the COS have been incorporated into the Annual Integrated Economic Survey (AIES).

A.1.3 Published tabulations of the Census Bureau establishment frame

The Census Bureau regularly produces several public-use datasets derived from the LBD, most notably the Business Dynamics Statistics (BDS). The BDS provides annual measures of establishment openings and closings, as well as stock counts of establishments and employment, longitudinalized to allow for analysis by firm and establishment characteristics. For the purposes of this paper, the BDS is a critical resource as it provides the official published aggregates for establishment counts by industry and size class. Specifically, the BDS categorizes establishments into size bins based on their current employment (e.g., 1–4, 5–9, up to 500+ employees), which we utilize in Section 5 to evaluate the shifting establishment size distribution. Because the BDS is built from the LBD, it inherits the frame's reliance on IRS administrative records and the annual Company Organization Survey, reflecting the Census Bureau's specific methodology for identifying business locations and linking them to parent firms.

A.2 BLS

A.2.1 Construction of the BLS establishment frame

The BLS establishment frame is the result of a partnership with U.S. states, territories, and the District of Columbia. When a private sector business hires employees in a covered industry (which includes nearly all employees) in a given state, it must notify the state government of its new hires and start paying unemployment insurance taxes on them. These governments in turn provide these reports of employment and wage bills, as well as each establishment's location and industry, to the BLS each quarter. These QCEW microdata are the basis for the QCEW report and are derived from the UI accounting system in that state.

The states receive a Quarterly Contributions Report (QCR) from all private sector employers, as well as from state and local governments covered under the UI program. Federal government employers provide statistical reports via the Report of Federal Employment and Wages; these reports contain only employment and wages data, for each employer's installations within each state. Normally, private sector employers submit one contribution report covering all of their economic activities conducted in a given state. For employers having only a single physical location or worksite in a state (i.e., single-unit employers), and thus operating under a single assigned industry and geographic code,

the data from the accounting file are sufficient for statistical purposes.

However, for employers with multiple establishments in a given state or for employers engaged in different industrial activities within a state, the employer's QCR reflects only statewide employment and wages, and is not disaggregated by establishment or worksite. With some exceptions, all employers with multiple establishments and at least ten employees must submit a Multiple Worksite Report (MWR) to the government of the state or territory in which they operate, which lists the total number of employees as well as total payroll at each establishment. These governments, in turn, provide these responses to the BLS. This improves the quality of local economic data by more accurately reporting the location and type of economic activity.

BLS maintains its establishment frame on a 3-year cycle. During this cycle, BLS identifies which establishments have changed from one type of business to another and reclassifies them under a different NAICS code. Approximately one-third of all private sector businesses (with more than three employees) are contacted annually by the Annual Refiling Survey (ARS). The ARS collects specific information on the employer's industrial activities and geographic locations. The information is used to ensure that each establishment is assigned to the correct industry and that each address geocodes the correct geographic location of the establishment. The ARS also asks employers to identify the locations of new worksites they have established in the state. If these employers meet QCEW program reporting criteria, then an MWR requesting information on employment and wages for each worksite each quarter is sent to the employer. Thus, the ARS is also used to identify new potential MWR-eligible employers.

A.2.2 Establishment-level QCEW and LEHD microdata

Establishment-level QCEW microdata is restricted to protect the confidentiality of responding businesses. Researchers interested in using these records directly may apply for access through the Bureau of Labor Statistics' Research Data Assistance (RDA) program (https://www.bls.gov/rda/) and the Standard Application Process (SAP, https://www.researchdatagov.org/). Information regarding the available data and the specific requirements for the SAP can be found on the BLS website (https://www.bls.gov/rda/data/restricted-data.htm). Projects must be approved by BLS and requires agreement from the particular U.S. state that collects the data (https://www.bls.gov/rda/data/qcew-institution-transfer-policy.htm). Approved projects typically require re-

searchers to access the data through secure virtual or physical enclaves maintained by the agency.

A secondary and increasingly common path for researchers to analyze QCEW-based records in conjunction with Census data is through the Longitudinal Employer-Household Dynamics (LEHD) program. As described by Abowd et al. (2009), the LEHD program is a partnership between the Census Bureau and participating states to integrate administrative records for the purpose of producing new labor market statistics. Under this partnership, the Census Bureau receives QCEW responses from state agencies, which are then used to build the LEHD's Employer Characteristics File (ECF).

The ECF serves as the employer establishment frame for the LEHD system, containing longitudinal records of establishments including their Unemployment Insurance (UI) account numbers, industry codes, and physical locations. Because the LEHD program is housed within the Census Bureau, it performs its own internal crosswalk to link these state-reported UI records to the Census Bureau's Business Register. This unique infrastructure allows researchers within the FSRDC environment to compare the BLS-derived establishment counts in the ECF directly against the Census-derived counts in the LBD for the same underlying firms.

A.2.3 Published tabulations of the BLS establishment frame

The QCEW program publishes a wide array of high-frequency data, but for researchers studying business entry and exit, the primary public-use product is the Business Employment Dynamics (BDM). The BDM is generated by linking QCEW records over time to track the lifecycle of individual establishments. It provides the official BLS counts of establishment "births" and "deaths," as well as the total stock of active establishments. Like the BDS, the BDM publishes these totals at various levels of aggregation, including by NAICS supersector and by establishment size class. Because the BDM is rooted in state UI filings and the MWR, it provides a different lens on the economy than the BDS. In the following sections, we compare these BDM aggregates directly against BDS aggregates to document a widening gap in the U.S. establishment count and to show how the BLS's unique frame construction leads to a disproportionate surge in the measured number of small-scale establishments.

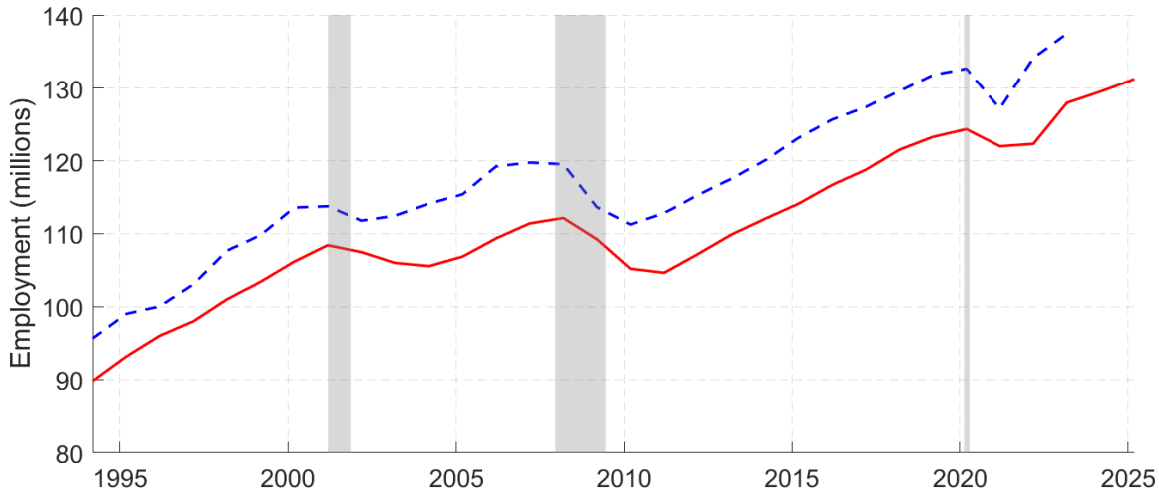
B Supplemental tables and figures

This appendix provides additional empirical results that complement the analysis in the main text. The material is organized into two subsections. Section B.1 presents supplemental figures regarding employment trends derived from the published BDS and BDM aggregates. Section B.2 contains detailed tabulations from the linked ECF-LBD microdata exercise that offer a more granular view of the establishment discrepancies discussed in Section 4.

B.1 Employment from published aggregates

While the primary focus of this paper is the divergence in establishment counts, the underlying employment data also exhibit systematic differences. The figures in this section display total employment levels both overall and across industry supersectors.

Figure B1: Total U.S. employment, March, 1994-2024



Notes: The Business Employment Dynamics (BDM, solid) is derived from the Quarterly Census of Employment and Wages (QCEW) and published by the Bureau of Labor Statistics. The Business Dynamics Statistics (BDS) is derived from the Longitudinal Business Database (LBD, dashed) and published by the Census Bureau. Shaded regions indicate recessions.

Appendix Figure B1 displays the total employment levels from the published aggregates of the BDM and the BDS from 1994 through the most recent available data. Both series exhibit a long-term upward trajectory, interrupted by significant cyclical downturns corresponding to the 2001 recession, the 2008-2009 Global Financial Crisis, and the COVID-19 pandemic in 2020. Despite these shared trends, a persistent level difference exists between the two series. In 1994, BDM employment stood at

approximately 89.8 million compared to 95.6 million in the BDS; by 2023, these totals grew to 127.9 million and 137.5 million, respectively. This consistent gap suggests that while both frames track the aggregate business cycle closely, they maintain different coverage or measurement baselines for total workforce size.

The recent data from 2020 to 2024 highlights the volatility of the pandemic and subsequent recovery period. Both datasets record the sharp contraction in employment during 2021 (reflecting the March 12 reference date for that fiscal year) and a robust rebound thereafter. Notably, the gap between the two series fluctuated during this period; the difference narrowed to approximately 5.2 million in 2021 before widening significantly to 11.8 million in 2022. By 2023, the gap remained substantial at 9.5 million. The BDM series extends to 2024, reaching a total of 129.4 million employees, while the BDS series ends in 2023 due to the standard reporting lag in Census administrative data. These diverging levels in the final years of the sample underscore the importance of understanding the underlying frame differences when interpreting labor market dynamism.

The persistent gap between the two series is further characterized by a consistent proportionate difference. Throughout most of the sample period from 1994 to 2019, the BDS employment level was approximately 6.5% to 7.5% higher than the corresponding BDM level. For instance, in 1994, the BDS reported 6.5% more employees than the BED; by the eve of the pandemic in 2019, this relative difference remained nearly unchanged at 6.9%. This stability suggests that the two frames, while operating at different scales due to the inclusion or exclusion of specific administrative records, historically captured the trend and cycle of the U.S. labor market in a highly synchronized manner.

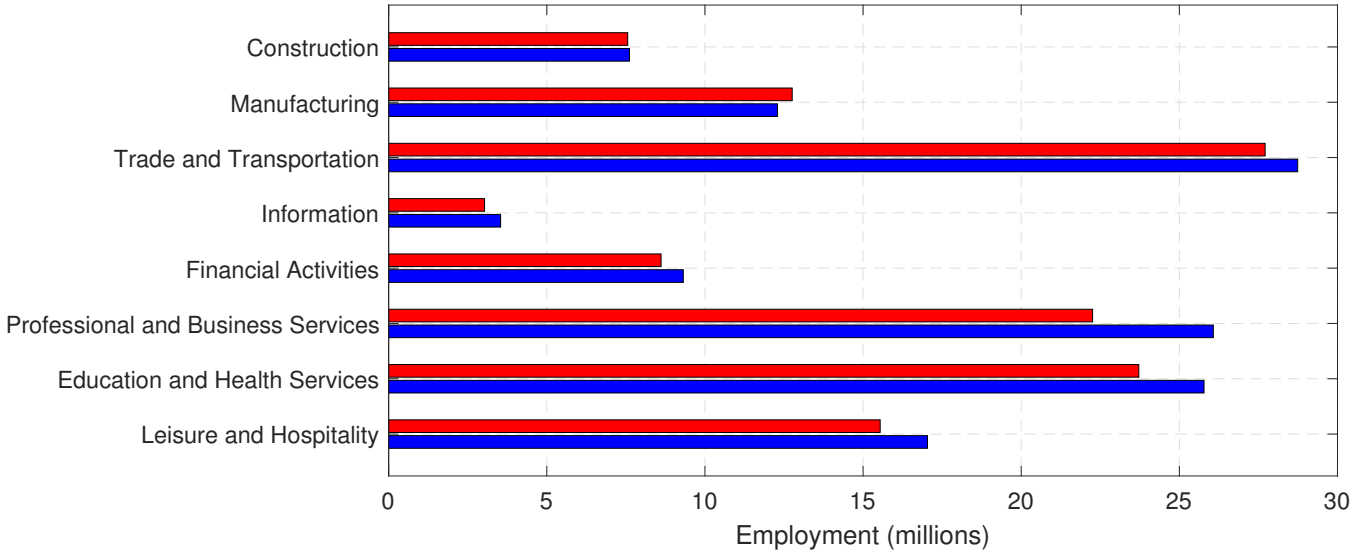
However, the period following the COVID-19 shock introduced a rare break in this relative stability. The proportionate difference narrowed to 4.3% in 2021, the lowest level in the thirty-year series, before widening to a record high of 9.6% in 2022. By 2023, the proportionate difference stood at 7.4%, returning closer to its historical average. This recent volatility in the relative gap underscores that the "establishment surge" and subsequent labor market reallocation have not been recorded uniformly across the two frames, suggesting that the divergence in establishment counts documented in this paper is accompanied by a non-trivial, though less extreme, divergence in the measurement of total employment.

This contrast—higher aggregate employment in the BDS despite a lower establishment count—is one of the most compelling methodological puzzles revealed by the comparison of the two frames. It suggests that the Census Bureau effectively assigns a larger volume of employees to a more con-

solidated set of physical locations than the BLS. This observation aligns with the distributional skew documented in Section 5.2, where the BDS was found to identify significantly more large-scale establishments (500+ employees) while the BDM identified a massive surplus of micro-establishments (1–19 employees).

From an analytical standpoint, this implies that the two agencies are providing distinct lenses on the organizational density of the U.S. workforce. The Census-based BDS appears to prioritize the capture of high-employment hubs and the consolidated workforce within established multi-unit firms, likely due to its reliance on IRS-based administrative records. Conversely, the BLS-based BDM identifies a more geographically dispersed and fragmented workforce. The fact that the BDS reports nearly 10 million more employees while missing millions of the micro-establishments found in the BDM suggests that the "establishment surge" of the 2020s is driven by a class of small-scale units that, while numerous, account for a relatively small fraction of total aggregate employment.

Figure B2: U.S. employment, by industry, March 2023.



Notes: The Business Employment Dynamics (BDM) is derived from the Quarterly Census of Employment and Wages (QCEW) and published by the Bureau of Labor Statistics. The Business Dynamics Statistics (BDS) is derived from the Longitudinal Business Database (LBD) and published by the Census Bureau.

Appendix Figure B2 disaggregates the employment gap by industry supersector for 2023, revealing that the aggregate lead maintained by the BDS is a near-universal feature of the industrial landscape. In every major supersector except Manufacturing, the Census BDS identifies a higher volume of employment than the BLS BDM. This divergence is most significant in Professional and Business Services, where the BDS reports 26.1 million employees compared to 22.3 million in the BDM (a 14.6% difference), and in Information, where the BDS lead stands at 14.3%. These results

suggest that the difference in the establishment size distribution is not merely a matter of counting buildings; the Census Bureau effectively assigns a much larger workforce to its establishment frame, particularly in high-value service sectors.

The data for Education and Health Services provides a particularly illuminating case study of these frame differences. In this sector, the BDS reports 25.8 million employees, approximately 8.0% more than the 23.7 million recorded in the BDM. This employment lead is notable when contrasted with the establishment-level findings in Section 5.2, which showed that the BLS identifies significantly more physical locations in this industry. This pattern—more people assigned to fewer sites—likely reflects the Census Bureau's reliance on IRS Form 941 business records, which aggregate employment into larger corporate hubs.

Furthermore, the employment gap in Education and Health Services persists despite a known coverage limitation in the Census frame: home-based healthcare providers are often reported via IRS Form 1040, Schedule H (Household Employment Taxes), which falls outside the scope of the Census Bureau's employer business universe. Because the BLS frame captures these workers through state Unemployment Insurance records, the BDM likely has a more comprehensive view of this specific, decentralized workforce. The fact that the BDS still maintains a 2-million-person employment lead in this sector suggests that its tendency to consolidate employment within traditional employer-firm boundaries is massive enough to overwhelm the exclusion of the "Schedule H" workforce. Ultimately, Figure B2 confirms that while the BLS identifies a surge of micro-establishments, the Census Bureau continues to characterize the U.S. economy as one dominated by much larger, more labor-intensive employment centers.

B.2 Supplemental linked microdata tabulations

This section provides a series of detailed tabulations derived from the linked ECF-LBD microdata. While the results presented in the main text focus on the primary drivers of the establishment discrepancy, these supplemental tables offer a comprehensive view across a broader range of industrial classifications and firm types. By extending the analysis to more granular NAICS levels, these tables allow for a more precise identification of the specific organizational structures—such as specialized professional service providers and home-based healthcare entities—that define the divergence between the Census Bureau and BLS frames.

A central theme of these supplemental results is the documentation of "firm-level disagreement." The tables contained herein disaggregate the linked sample into categories based on whether the LBD and ECF agree on a firm's multi-unit status. These results demonstrate that the majority of the "missing" establishments in the LBD are not simply omitted firms, but are rather additional work sites belonging to firms already present in both databases. This granular evidence is essential for understanding the distributional skew discussed in Section 5.2, as it reveals how the internal organization of established firms is captured differently by the two statistical agencies.

Table B1 provides a longitudinal perspective on the employment discrepancies between the ECF and LBD, categorized by the organizational agreement between the two frames. The data reveal that the most substantial employment gap occurs among firms that both datasets correctly identify as multi-unit entities. In 2016, for firms labeled as multi-unit in both sources, the LBD reported 69.5 million employees compared to 63.7 million in the ECF. This 5.8-million-person difference underscores that even when there is consensus on a firm's complexity, the Census Bureau's administrative model assigns significantly more internal employment to these multi-unit structures than the BLS-based frame.

The table also highlights the stability of these organizational disagreements over time. Between 2004 and 2016, the total employment difference for firms where both sources agree on single-unit status actually narrowed from -2.6 million to -1.7 million. However, the persistent gap in the "Multi-unit in both" columns (3) and (4) remains the dominant driver of the aggregate employment lead held by the LBD. Interestingly, the employment associated with firms present in only one source is relatively small and stable, totaling approximately 6.8 million in 2016. This confirms that the broader "employment puzzle"—the LBD's ability to identify more employees despite having fewer establishments—is primarily driven by how common, established firms are represented within each agency's respective infrastructure rather than by a total lack of coverage of the firms themselves.

Table B2 provides a comprehensive industry-level breakdown of firm counts, categorized by the organizational agreement between the LBD and ECF. Across most supersectors, there is a striking degree of consensus regarding the identity of established firms. For example, in Professional and Business Services, the two frames agree on approximately 793,000 single-unit firms and 14,000 multi-unit firms. This high level of agreement for firms common to both sources confirms that the discrepancies documented in the main text are not driven by a failure to identify the same parent organizations, but rather by how those organizations are internally structured across physical work sites.

Table B1: Employment counts (millions), by single- vs. multi-unit agreement for particular firms between the LBD and QCEW, by year

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Single-unit		Multi-unit in both		Multi-unit in QCEW		Multi-unit in LBD		Only in one source	
	ECF	LBD	ECF	LBD	ECF	LBD	ECF	LBD	ECF	LBD
Agreement in 2004										
Total	35.6	38.2	53.0	58.5	7.2	7.5	2.9	3.4	6.8	6.8
Difference	-2.6		-5.5		-0.3		0.0		-0.5	
Agreement in 2016										
Total	36.0	37.8	63.7	69.5	8.4	9.2	2.5	3.0	6.8	6.6
Difference	-1.7		-5.8		-0.7		-0.6		0.2	
Change from 2004 to 2016										
Difference	0.8		-0.3		-0.5		0.0		0.6	

Notes: Authors' tabulations of the LEHD ECF and the LBD. The ECF statistics are derived from the March employment counts reported as part of the QCEW. Counts by category may not sum to the total due to rounding.

Table B2: Firm counts (thousands), by single- vs. multi-unit agreement for particular firms between the LBD and QCEW, 2016, by industry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Single-unit	Multi-unit in both		Multi-unit in QCEW		Multi-unit in LBD		Only in one source	
	in both	QCEW	LBD	QCEW	LBD	QCEW	LBD	QCEW	LBD
Natural Resources and Mining (11, 21)									
Total	15	1	1	1	1	0	0	0	2
Difference	0	0		0		0		-2	
Construction (23)									
Total	501	3	3	15	15	1	1	15	78
Difference	0	0		0		0		-64	
Manufacturing (31-33)									
Total	197	11	11	15	15	2	2	8	18
Difference	0	0		0		0		-10	
Trade, Transportation, and Utilities (42, 44-45, 48-49, 22)									
Total	828	26	26	35	35	16	16	39	103
Difference	0	0		0		0		-64	
Information (51)									
Total	46	2	2	7	7	1	1	3	9
Difference	0	0		0		0		-6	

Notes: Authors' tabulations of the LEHD ECF and the LBD. The ECF statistics are derived from the March employment counts reported as part of the QCEW. Counts by category may not sum to the total due to rounding.

Table B2: Firm counts (thousands), by single- vs. multi-unit agreement for particular firms between the LBD and QCEW, 2016, by industry (cntd.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Single-unit	Multi-unit in both		Multi-unit in QCEW		Multi-unit in LBD		Only in one source	
	in both	QCEW	LBD	QCEW	LBD	QCEW	LBD	QCEW	LBD
Financial Activities (52-53)									
Total	372	11	11	14	14	6	6	17	56
Difference	0	0		0		0		-39	
Professional and Business Services (54-56)									
Total	793	14	14	57	57	5	5	37	132
Difference	0	0		0		0		-96	
Education and Health Services (61-62)									
Total	568	14	14	11	11	11	11	498	86
Difference	0	0		0		0		412	
Leisure and Hospitality (71-72)									
Total	490	12	12	6	6	6	6	18	56
Difference	0	0		0		0		-39	
Other Services (81)									
Total	413	4	4	7	7	4	4	104	219
Difference	0	0		0		0		-115	

Notes: Authors' tabulations of the LEHD ECF and the LBD. The ECF statistics are derived from the March employment counts reported as part of the QCEW. Counts by category may not sum to the total due to rounding.

The most notable outlier in the table is Education and Health Services, which displays a unique pattern in the "Only in one source" columns (8) and (9). In this sector, the ECF identifies 498,000 firms that are entirely absent from the LBD, compared to just 86,000 firms that are unique to the LBD. This net surplus of 412,000 firms in the ECF represents a fundamental coverage difference that is not seen in other industries. As discussed in the context of at-home care providers, this massive influx of ECF-only firms reflects the BLS frame's ability to capture smaller, decentralized employers—likely reported through state unemployment insurance records—that fall outside the Census Bureau's IRS-based employer business frame.

For most other supersectors, columns (8) and (9) show the opposite relationship, with the LBD typically identifying more unique firms than the ECF. In Professional and Business Services, for instance, the LBD reports 132,000 unique firms while the ECF reports only 37,000. These results suggest that outside of the specialized healthcare and home-service domains, the Census Bureau's administrative model is generally more expansive in identifying individual firm entities, even as the BLS model identifies a higher number of individual establishments. This reinforces our other findings on the establishment size distribution: the LBD is robust at identifying a larger number of distinct firm entities and high-employment hubs, while the ECF is uniquely sensitive to the micro-scale granularity of the modern service economy.

Table B3 provides the industrial counterpart to the employment agreement trends discussed in Table B1, showing that the Census Bureau's employment lead is deeply embedded in the "Multi-unit in both" columns (3) and (4) across nearly every supersector. In the two sectors with the largest establishment discrepancies—Professional and Business Services and Education and Health Services—the LBD identifies a surplus of 1.1 million and 1.3 million employees, respectively, within firms that both datasets agree are multi-unit. This confirms that for the most complex organizations in the economy, the Census Bureau consistently assigns a larger internal workforce to its frame than the BLS.

The table also provides critical context for the employment puzzle we document in service industries. In Professional and Business Services, for instance, even for firms that the QCEW identifies as multi-unit but the LBD identifies as single-unit (Column 6), the LBD reports significantly more employment (3.13 million vs. 2.74 million). This specific comparison illustrates the aggregation effect at the heart of differences in the establishment size distribution: the Census Bureau's administrative model appears to consolidate a larger volume of activity into fewer, larger reporting units, whereas the BLS frame disaggregates that same activity across a more granular establishment network.

Table B3: Employment counts (thousands), by single- vs. multi-unit agreement for particular firms between the LBD and QCEW, 2016, by industry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Single-unit		Multi-unit in both		Multi-unit in QCEW		Multi-unit in LBD		Only in one source	
	QCEW	LBD	QCEW	LBD	QCEW	LBD	QCEW	LBD	QCEW	LBD
Natural Resources and Mining (11, 21)										
Total	128	132	449	483	59	61	6	8	20	11
Difference	-4		-34		-120		-2		10	
Construction (23)										
Total	3,865	3,988	1,335	1,434	784	790	48	52	277	256
Difference	-123		-99		-1,574		-4		22	
Manufacturing (31-33)										
Total	2,835	2,909	8,879	9,251	1,109	1,144	131	160	505	205
Difference	-74		-372		-2,253		-29		300	
Trade, Transportation, and Utilities (42, 44-45, 48-49, 22)										
Total	6,031	6,250	17,500	18,790	1,095	1,150	490	546	1496	693
Difference	-219		-1,290		-2,245		-56		803	
Information (51)										
Total	302	319	2,228	2,408	272	292	18	22	147	66
Difference	-17		-180		-564		-4		82	

Notes: Authors' tabulations of the LEHD ECF and the LBD. The ECF statistics are derived from the March employment counts reported as part of the QCEW. Counts by category may not sum to the total due to rounding.

Table B3: Employment counts (thousands), by single- vs. multi-unit agreement for particular firms between the LBD and QCEW, 2016, by industry (cntd.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Single-unit		Multi-unit in both		Multi-unit in QCEW		Multi-unit in LBD		Only in one source	
	QCEW	LBD	QCEW	LBD	QCEW	LBD	QCEW	LBD	QCEW	LBD
Financial Activities (52-53)										
Total	1,540	1,595	5,621	6,049	510	535	152	169	389	207
Difference	-55		-428		-1,045		-17		182	
Professional and Business Services (54-56)										
Total	4,897	5,177	8943	10,050	2,741	3,128	230	443	1,101	709
Difference	-280		-1107		-5,869		-213		-392	
Education and Health Services (61-62)										
Total	7,099	7,575	11,040	12,320	1,200	1,351	855	1,001	1,953	2,005
Difference	-476		-1,280		-2,551		-146		-52	
Leisure and Hospitality (71-72)										
Total	6,958	7,331	6,900	7,626	418	425	405	6	18	56
Difference	-373		-726		-843		-77		165	
Other Services (81)										
Total	2,367	2,482	932	1,037	267	310	147	160	401	1,626
Difference	-115		-105		-577		-13		-1225	

Notes: Authors' tabulations of the LEHD ECF and the LBD. The ECF statistics are derived from the March employment counts reported as part of the QCEW. Counts by category may not sum to the total due to rounding.

Finally, the comparison of employment in the "Only in one source" columns (9) and (10) reinforces the conclusion that the BLS "establishment surge" is a phenomenon of micro-scale activity. While the previous analysis of firm counts (Table B2) showed that the ECF identifies over 400,000 more unique firms in Education and Health Services than the LBD, Table B3 shows that the total employment associated with those ECF-exclusive firms (1.95 million) is actually slightly lower than the employment in the LBD-exclusive firms (2.01 million). This implies that the hundreds of thousands of additional firms and establishments captured by the BLS in this sector are, on average, extremely small. In contrast, the firms unique to the Census frame, though far fewer in number, represent significantly larger employment centers. This data collectively illustrates an economy that looks increasingly fragmented and decentralized in the BLS frame, yet remains anchored by large-scale, high-density employment hubs in the Census frame.

B.3 Additional supersector results

This section presents Table B4, which represents the same information as Table 2 for all ten supersectors. We have seen that two supersectors represented in Table 2 play dominant role in explaining the difference between LBD and ECF, and thus this section's analysis is supplemental.

It is noteworthy that in nearly all supersectors, among matched firms, single-establishment firms in the LBD outnumber those in the ECF. This can be seen by comparing the sum of columns (1) and (5) to the sum of columns (1) and (6): column (5) is always greater than or equal to column (6). In nearly all supersectors, the number of establishments associated with matched firms is greater in the ECF than the LBD. The exceptions are Education and Health Services (61-62), discussed previously, and Financial Activities (52-53). While Professional and Business Services (54-56) has the greatest difference among matched firms, with the ECF having 271 thousand more establishments (discussed above), there are a few supersectors that also have substantial differentials. Manufacturing has the second highest differential, of 99 ($= 65 + 37 - 3$) thousand. Construction (23) has a differential of 35 ($= 29 + 7 - 1$) thousand favoring the ECF; Trade, Transportation, and Utilities (42, 44-45, 48-49, 22) has one of 34 ($= 81 - 29 - 18$) thousand; and Information (51) has one of 15 ($= 21 - 5 - 1$) thousand. We conclude that, among firms present in both the ECF and LBD, they tend to have more establishments in the ECF, and that this phenomenon is present in most industries if not spread uniformly through them.

As mentioned above, firms only present in one data source provide, on net, 508 thousand more establishments to the ECF than the LBD in the Education and Health Services (61-62) supersector. Few of the other supersectors have more missing firms in the ECF than the LBD. None have as many as the Education and Health Services (61-62) supersector - it has an order of magnitude more than any other. The Manufacturing (31-33) supersector has 15 thousand more such firms, Leisure and Hospitality (71-72) has 13 thousand, and Information (51) has two thousand. In all six other supersectors, the LBD has more establishments in firms that are absent from the ECF than the reverse. We therefore conclude that differences in firm composition between the ECF and LBD are mostly concentrated in the Education and Health Services (61-62) supersector, and, indeed in a single 6-digit NAICS industry, Services for the Elderly and Persons with Disabilities (624120).

Table B5 compares the share of multi-unit firms in firm count, establishment count, and employment for ECF and LBD. The contents of this table is analyzed in Section 5.1 in the main text.

Table B4: Establishment counts (thousands), by single- vs. multi-unit agreement for particular firms between the LBD and ECF, 2016, by industry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Single-unit	Multi-unit in both		Multi-unit in ECF		Multi-unit in LBD		Only in one source	
	in both	ECF	LBD	ECF	LBD	ECF	LBD	ECF	LBD
Natural Resources and Mining (11, 21)									
Total	15	10	9	4	1	0	0	2	2
Difference	0	1		2		0		-1	
Construction (23)									
Total	501	30	23	44	15	1	2	37	79
Difference	0	7		29		-1		-42	
Manufacturing (31-33)									
Total	197	171	106	52	15	2	5	34	19
Difference	0	65		37		-3		15	
Trade, Transportation, and Utilities (42,44-45,48-49,22)									
Total	828	551	569	115	35	16	45	113	121
Difference	0	-18		81		-29		-8	
Information (51)									
Total	46	58	63	28	7	1	2	12	10
Difference	0	-5		21		-1		2	

Notes: Authors' tabulations of the LEHD ECF and the LBD. The ECF statistics are derived from the March employment counts reported as part of the QCEW. Counts by category may not sum to the total due to rounding.

Table B4: Establishment counts (thousands), by single- vs. multi-unit agreement for particular firms between the LBD and ECF, 2016, by industry (cntd.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Single-unit	Multi-unit in both		Multi-unit in ECF		Multi-unit in LBD		Only in one source	
	in both	ECF	LBD	ECF	LBD	ECF	LBD	ECF	LBD
Financial Activities (52-53)									
Total	372	280	331	50	14	6	20	43	62
Difference	0	-51		36		-14		-19	
Professional and Business Services (54-56)									
Total	793	283	178	233	57	5	16	99	139
Difference	0	105		176		-10		-40	
Education and Health Services (61-62)									
Total	568	205	226	47	11	11	42	607	100
Difference	0	-21		37		-31		508	
Leisure and Hospitality (71-72)									
Total	490	194	193	23	6	6	20	81	68
Difference	0	1		17		-13		13	
Other Services (81)									
Total	413	51	57	28	7	4	14	170	223
Difference	0	-6		20		-10		-53	

Notes: Authors' tabulations of the LEHD ECF and the LBD. The ECF statistics are derived from the March employment counts reported as part of the QCEW. Counts by category may not sum to the total due to rounding.

Table B5: Share of firms with multiple establishments, as well as the share of establishments and employment at such firms: ECF vs. LBD

Industry		Firms		Establishments		Employment	
		ECF	LBD	ECF	LBD	ECF	LBD
Natural Resources and Mining (11, 21)	Share	13.3%	5.3%	50.4%	34.3%	79.4%	71.3%
	Difference	*86.0%*		*38.0%*		*10.8%*	
Construction (23)	Share	3.9%	0.7%	16.1%	4.1%	37.0%	23.5%
	Difference	*139.1%*		*118.8%*		*44.6%*	
Manufacturing (31-33)	Share	11.9%	5.6%	55.2%	33.3%	77.3%	69.6%
	Difference	*72.0%*		*49.5%*		*10.5%*	
Trade, Transportation, and Utilities (42, 44-45, 48-49, 22)	Share	7.3%	4.8%	46.1%	40.0%	74.4%	71.7%
	Difference	*41.3%*		*14.2%*		*3.7%*	
Information (51)	Share	17.7%	4.6%	66.3%	51.7%	88.7%	79.3%
	Difference	*117.5%*		*24.7%*		*11.2%*	
Financial Activities (52-53)	Share	6.6%	3.9%	47.9%	44.8%	78.4%	73.6%
	Difference	*51.4%*		*6.7%*		*6.3%*	
Professional and Business Services (54-56)	Share	8.7%	2.2%	41.5%	17.2%	70.2%	55.3%
	Difference	*119.3%*		*82.8%*		*23.7%*	
Education and Health Services (61-62)	Share	4.9%	4.3%	27.2%	30.3%	60.1%	57.3%
	Difference	*13.0%*		*-10.8%*		*4.8%*	
Leisure and Hospitality (71-72)	Share	4.2%	3.8%	35.8%	29.3%	51.9%	50.9%
	Difference	*10.0%*		*20.0%*		*1.9%*	
Other Services (81)	Share	7.9%	1.6%	26.4%	10.8%	34.9%	22.6%
	Difference	*132.6%*		*83.9%*		*42.8%*	
Overall	Share	6.7%	3.3%	37.9%	28.1%	65.7%	59.0%
	Difference	*68.0%*		*29.7%*		*10.7%*	

Notes: Authors' tabulations of the LEHD ECF and the LBD. The ECF statistics are derived from the March employment counts reported as part of the QCEW. Proportionate differences are calculated as $(B-A)/((B+A)/2)$.

C Details of the model analysis in Section 5.3

Our model follows Hopenhayn and Rogerson (1993) closely. The purpose of the model analysis is to show that the different measurement of the establishment size distribution yields quantitatively different policy predictions.

C.1 Setting

The time is discrete with infinite horizon. The representative consumer maximizes the utility

$$\sum_{t=0}^{\infty} \beta^t \log(c_t) - AL_t^s$$

subject to

$$p_t c_t \leq L_t^s + \Pi_t + R_t,$$

where c_t is consumption, L_t^s is labor supply, p_t is price, Π_t is profit of all firms, and R_t is transfer from the government. The wage rate is the numeraire. The parameter $\beta \in (0,1)$ and $A > 0$. We focus on the steady state and thus remove the subscript t. The first-order condition:

$$A = \frac{1}{L^s + \Pi + R}. \tag{1}$$

We consider heterogeneous establishments. All establishments are owned by the household, and each establishment maximizes the present value of profit, appropriately discounted. When the aggregate economy is in the steady state, an establishment solves the dynamic programming problem

$$W(s,n,c_f) = \max\{W^o(s,n,c_f), W^e(s,n)\},$$

where

$$W^o(s,n,c_f) = \max_{n'} \; pf(n',s) - n' - pc_f - g(n',n) + \beta \mathbb{E}_{s',c_f'}[W(s',n',c_f')]$$

and

$$W^e(s,n) = \max_{n'} \; pf(n',s) - n' - g(n',n) - \beta g(0,n').$$

The timing within the period is (i) the establishment enters the period with the state variable s (the

idiosyncratic productivity, which follows a Markov process) and n (the employment last period); (ii) the establishment receives the i.i.d. cost shock c_f, which indicates the continuation cost (a fixed cost if it continues operating) for the current period; (iii) the establishment decides whether to exit after operating the current period; (iv) after hiring workers n', the establishment pays taxes and operates; (v) if it exits, the exit occurs in the beginning of the following period. Here, $f(n',s)$ is the production function and $g(n',n)$ is the firing tax when the current employment is n' and the past employment is n. The first equation is the exit choice. $W(\cdot)$ is the value function in the beginning of the period, $W^o(\cdot)$ is the value if the operation continues into the next period, and $W^e(\cdot)$ is the value if the establishment exits after the current period. The second and the third equations are Bellman equations for both cases.

Here, the departure from Hopenhayn and Rogerson (1993) is the inclusion of the c_f shock. In Hopenhayn and Rogerson (1993), c_f is fixed and certain. Under that formulation, in the baseline without the firing tax, small establishments (corresponding to unproductive establishments) below a certain threshold exit after one period of operation. With reasonable calibrations, it is impossible to match the prevalence of small establishments in the data. All other parts of the model, including the timing, follow Hopenhayn and Rogerson (1993).

We denote the exit decision by $X(s,n,c_f)$ and the employment decision by $N(s,n,c_f)$. The beginning-of-the-period (after s and c_f are revealed) incumbent measure is denoted $\mu(s,n,c_f)$. The entrant distribution, when the total mass of entrants is normalized to one, is $\nu(s,c_f)$. The total mass of entrants is M. The total labor demand sums up the individual labor demand:

$$L^d(\mu,M) = \int N(s,n,c_f)d\mu(s,n,c_f) + M\int N(s,0,c_f)d\nu(s,c_f).$$

We assume free entry with entry cost c_e. Thus the condition

$$c_e = \int W(s,0,c_f)d\nu(s,c_f)$$

has to hold in a stationary equilibrium with positive entry. Other equilibrium objects R and Π can be computed as

$$R(\mu,M) = \int r(s,n,c_f)d\mu(s,n,c_f) + M\int r(s,0,c_f)d\nu(s,c_f).$$

where

$$r(s,n,c_f) = \begin{array}{l}(1-X(s,n,c_f))\int g(N(s',n',c_f'),n')dF(s,s')dG(c_f') \\ +X(s,n,c_f)g(0,n'),\end{array}$$

where $n' = N(s,n,c_f)$, and

$$\Pi(\mu,M) = pY(\mu,M) - L^d(\mu,M) - R(\mu,M) - Mpc_e.$$

Thus

$$L^d + \Pi + R = pY(\mu,M) - Mpc_e.$$

This equation, together with the labor supply condition (1), implies

$$A = \frac{1}{pM(\hat{Y} - c_e)}$$

or alternatively

$$M = \frac{1}{pA(\hat{Y} - c_e)},$$

where $\hat{Y}$ is the aggregate output corresponding to unit entry.

C.2 Calibration

For calibration, we follow Hopenhayn and Rogerson (1993) where we can. For the production function, we assume a Cobb-Douglas function

$$f(n,s) = sn^{\theta}.$$

The productivity follows an AR(1) process

$$\log(s_t) = a + \rho \log(s_{t-1}) + \varepsilon, \text{ where } \varepsilon \sim N(0,\sigma^2).$$

The AR(1) process is discretized into a Markov chain using the Tauchen method. The firing tax is formulated as

$$g(n',n) = \tau \max\{0, n - n'\}.$$

When we calibrate, we assume that $\tau = 0$ in the baseline.

Table B6 lists the parameters whose values are set identically to Hopenhayn and Rogerson (1993). As in Hopenhayn and Rogerson (1993), the maximum of $\log(s)$ is set so that the optimal employment at that value is $5,000$. With the minimum value of $\log(s)$ being zero, we set 50 equally-spaced grids on s. As in Hopenhayn and Rogerson (1993), we set 250 grids on n. The grid points are set so that $\log(n)$ is equally spaced.

Table B6: Preset parameter values

parameter	value
θ	0.64
β	0.8

The probability of c_f being zero is set at 0.5. The parameter $a = 0.30$. The value of c_f when nonzero is set so that the exit rate is similar to Hopenhayn and Rogerson's (1993) target of 37%. We set $c_f = 15$. The exit rate for BDM calibration is 37%, and for BDS calibration, it is 34%. The number of grids for s is 100, with the same upper and lower bounds as above, and the entrants' s have the uniform distribution on the grids with the upper bound of the first 45 grids (value of $\log(s) = 1.56$) out of a total of 100 grids.

The values of ρ and σ explicitly target the share of 100+ establishments and 20-99 establishments, computed from the "Total" row of Table 3. The shares in the data are in Table B7 below. The model

Table B7: Data establishment distribution

	BDM (BLS)	BDS (Census)
500+	0.0021	0.0034
100-499	0.0193	0.0251
20-99	0.1131	0.1402
1-19	0.8654	0.8313

matches the two targets exactly, and the fit is also good when we divide the $100+$ share into $100-499$ and $500+$, as we can see from Table B8. The estimated parameter values are as follows. From (B9), the BDM version of the calibration implies a less persistent and more volatile productivity process compared to the BDS version. The cost of entry, c_e, and the preference parameter A are set so that $p = 1$ and $L = 0.6$ when $\tau = 0$.

Table B8: Model establishment distribution

	BDM (BLS)	BDS (Census)
500+	0.0020	0.0028
100-499	0.0194	0.0257
20-99	0.1131	0.1402
1-19	0.8654	0.8313

Table B9: Estimated parameter values

	BDM (BLS)	BDS (Census)
ρ	0.902	0.923
σ	0.319	0.283

C.3 Results

The following are the results when we impose $\tau = 0.2$. Note that one period is five years and the wage rate (that is, five-year earnings) is normalized to one. Thus $\tau = 0.2$ corresponds to one year of earnings. Table B10 summarizes the results. The fourth column divides the third column by the loss implied by the second column (100 minus the value in the second column).

Table B10: Model outcome with $\tau = 0.2$, normalizing the $\tau = 0$ values of Y, L, and Y/L to 100.

	BDM (BLS)	BDS (Census)	Difference (BDS−BDM)	BDM-BDS Loss Ratios
p	1.043	1.038	0.005	–
Y	95.0	96.3	1.3	33.6%
L	96.4	97.2	0.8	28.5%
Y/L	98.6	99.1	0.5	50.8%

Under the BDM calibration, the output declines by 5.0%, the employment declines by 3.6%, and the labor productivity declines by 1.4%.[10] When our target is based on BDS, the policy prediction is substantially different. Now the output is predicted to decline by 3.7%, employment declines by 2.8%, and labor productivity declines by 0.9%. Thus we underpredict the effect of the firing tax if we base the calibration on the Census/BDS establishment-size distribution.

[10] In Hopenhayn and Rogerson (1993), the corresponding output decline is 4.6%, employment decline is 2.5%, and productivity decline is 2.1%.