BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Productivity and Technology

Dispersion in Dispersion: Measuring Establishment-Level Differences in Productivity

Cindy Cunningham, U.S. Bureau of Labor Statistics
Lucia Foster, Center for Economic Studies, U.S. Census Bureau
Cheryl Grim, Center for Economic Studies, U.S. Census Bureau
John Haltiwanger, Center for Economic Studies, U.S. Census Bureau
Sabrina Wulff Pabilonia, U.S. Bureau of Labor Statistics
Jay Stewart, U.S. Bureau of Labor Statistics
Zoltan Wolf, New Light Technologies

Working Paper 530
September 2020

Dispersion in Dispersion: Measuring Establishment-Level Differences in Productivity

Cindy Cunningham, Lucia Foster, Cheryl Grim, John Haltiwanger, Sabrina Wulff Pabilonia, Jay Stewart, and Zoltan Wolf*

September 2020

Abstract: We describe new *experimental* productivity statistics, Dispersion Statistics on Productivity (DiSP), jointly developed and published by the Bureau of Labor Statistics (BLS) and the Census Bureau. Productivity measures are critical for understanding economic performance. Official BLS productivity statistics, which are available for major sectors and detailed industries, provide information on the sources of aggregate productivity growth. A large body of research shows that *within-industry* variation in productivity provides important insights into productivity dynamics. This research reveals large and persistent productivity differences across businesses even within narrowly defined industries. These differences vary across industries and over time and are related to productivity-enhancing reallocation. Dispersion in productivity across businesses can provide information about the nature of competition and frictions within sectors, and about the sources of rising wage inequality across businesses. Because there were no official statistics providing this level of detail, BLS and the Census Bureau partnered to create measures of *within-industry* productivity dispersion. These measures complement official BLS aggregate and industry-level productivity growth statistics and thereby improve our understanding of the rich productivity dynamics in the U.S. economy. The underlying microdata for these measures are available for use by qualified researchers on approved projects in the Federal Statistical Research Data Center (FSRDC) network. These new statistics confirm the presence of large productivity differences and we hope that these new data products will encourage further research into understanding these differences.

*Cunningham, Pabilonia, and Stewart: Bureau of Labor Statistics; Foster and Grim: Center for Economic Studies, U.S. Census Bureau; Haltiwanger: University of Maryland; Wolf: New Light Technologies. John Haltiwanger was also a Schedule A part-time employee of the U.S. Census Bureau at the time of the writing of this paper. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau or the Bureau of Labor Statistics. All results have been reviewed to ensure that no confidential information is disclosed (DRB-FY19-393, DRB-FY19-526, CBDRB-FY20-410). We thank John Earle, Kevin Fox, Bernd Görgiz, Bart Hobijn, Mark Roberts, John Samuels, Chad Syverson, the Federal Economic Statistics Advisory Committee, the BLS Technical Advisory Committee, and participants at the Conference on Research on Income and Wealth and the 2015 Federal Statistical Research Data Center Conference for their helpful comments. Corresponding author: Cheryl.Ann.Grim@census.gov.

1. Introduction

Productivity measures are critical for understanding economic performance in the U.S. economy. The Bureau of Labor Statistics (BLS) produces the official labor and multifactor productivity growth statistics for major sectors and industries in the U.S. These statistics are constructed using aggregate industry-level data and can be thought of as changes in the first moment of establishment-level productivity (appropriately weighted). That is, these statistics show how productivity changes *on average* within sectors and industries, but they cannot provide insight into the variation in productivity levels across establishments within sectors or industries.[1]

To fill this void, BLS and the Census Bureau initiated the Collaborative Micro-productivity Project (CMP) to develop and publish experimental statistics on *within-industry* productivity dispersion (i.e., second-moment measures of establishment-level productivity) and to produce restricted-use research datasets. The public-use statistics developed in this project—Dispersion Statistics on Productivity (DiSP),[2] which were released for the first time in fall 2019—cover all 4-digit NAICS industries in the manufacturing sector and are published jointly by BLS and the Census Bureau. Restricted-use establishment-level data with micro-based estimates of productivity as well as its underlying components (e.g., output and input measures) are also available to qualified researchers on approved projects in secure Federal Statistical Research Data Centers (FSRDCs).[3]

Economic theory and recent empirical evidence suggest the second moments of productivity are informative on a number of important dimensions. One of the most important

[1] Although usually referred to as industry productivity growth or aggregate productivity growth, these statistics can be thought of as the weighted average of within-industry growth rates.
[2] These are available on both BLS and Census Bureau websites at: https://www.bls.gov/lpc/productivity-dispersion.htm and https://www.census.gov/disp.
[3] For more information on the FSRDCs: http://www.census.gov/fsrdc. An earlier version of this dataset was analyzed in Foster et al. (2016a).

findings in the literature on micro-level productivity is that large productivity differences across establishments exist even within narrowly defined industries.[4] For example, using data from the 1977 Census of Manufactures (CM), Syverson (2004a) found that establishments at the 90th percentile of the within-4-digit-SIC productivity distribution are about four times as productive as those at the 10th percentile.

Syverson's findings generated considerable interest in the causes and consequences of this dispersion. Possible market explanations include curvature of the profit function that prevents the most-productive business from taking over an industry, frictions in factor adjustments (such as costs of adjusting input factors), barriers to entry and exit, and distortions that inhibit the equalization of marginal products across businesses (such as the regulatory environment). Drivers of establishment-level productivity variation include differences in management skills, the quality of production factors, innovation, and investments in R&D.

Research has shown that the dispersion of establishment-level productivity varies across sectors, by geographic area, and over time. For example, Syverson (2004a, 2004b) shows that variation in dispersion across industries and geographic areas is related to product substitutability, market structure, and competition. Hsieh and Klenow (2009) argue that both cross-country variation and within-country variation in the dispersion of productivity are related to distortions that inhibit productivity-enhancing reallocation. Asker, Collard-Wexler, and De Loecker (2014) present evidence that the patterns of dispersion reflect the dynamic factor adjustment frictions within sectors. The findings in Foster et al. (2016b) suggest that productivity differences across establishments may be generated by differences in efficiency levels, demand shocks, frictions/distortions, or all of the above. Alternatively, Foster et al. (2017a) present evidence that industries experiencing a surge in innovation exhibit a burst of

[4] Syverson (2011) provides a survey of this literature.

firm entry, followed by an increase in productivity dispersion during an experimentation and shakeout phase, followed ultimately by an increase in industry-level productivity.

Establishment-level productivity differences are also correlated with important economic outcomes at the micro level, such as the survival and growth of establishments. There is a large literature on the connection between productivity, reallocation and growth (Baily, Hulten, and Campbell, 1992; Griliches and Regev, 1992; Foster, Haltiwanger and Krizan, 2001; Petrin, White and Reiter, 2001; Foster, Grim and Haltiwanger, 2016; Blackwood et. al., 2020; Decker et. al., 2020). These studies show that more-productive businesses are more to survive and grow. These findings contribute to the perspective that that reallocation—the process by which economic activity is allocated to its highest valued use—is an important contributor to aggregate productivity growth.

Productivity dispersion is also important for understanding rising wage inequality, which has been shown to be a between-firm phenomenon (Davis and Haltiwanger, 1991; Barth et al., 2016; Song et al., 2019; Haltiwanger and Spletzer, 2020). In addition, a number of studies have found that high-wage establishments are also highly productive and that rising between-establishment dispersion in wages is closely associated with rising between-establishment dispersion in productivity (e.g., Dunne et al., 2004). Economic theories of search and matching provide theoretical justification for the connection between productivity dispersion and wage dispersion (e.g., Burdett and Mortensen, 1998). Search and matching frictions create quasi-rents for worker-firm matches that make it optimal for high-productivity firms to pay high wages.

Our results using the DiSP experimental data confirm earlier findings about the large differences in productivity across establishments within industries. To preview our results, we find that on average, the manufacturing plant at the 75th percentile of the within-industry labor productivity distribution is more than twice as productive as the plant at the 25th percentile.[5] If we instead focus on multifactor productivity, we find that the 75th percentile plants are almost twice as productive as plants at the 25th percentile. Underlying these averages, we find substantial differences in dispersion across industries. For example, labor productivity dispersion in the industry at the 75th percentile of the dispersion distribution is about 1.4 times as big as the dispersion in the industry at the 25th percentile. The corresponding multiplier for multifactor productivity is 1.2.

The experimental statistics on productivity dispersion are intended to complement official BLS data, so it is crucial to understand the relationship between the dispersion of the productivity distribution derived from Census Bureau microdata and the statistics from BLS built from industry-level aggregates. Section 2 describes BLS productivity measures and productivity measures that we construct from Census microdata. Section 3 compares the two approaches to measuring inputs, output, and productivity for the manufacturing sector, and for 4-digit NAICS manufacturing industries. We also compare these measures to data from the NBER-CES Manufacturing Industry Database and examine a number of data and measurement issues such as imputation and weighting of the microdata. In section 4, we explore the variation in industry-level productivity dispersion measures across industries and over time. Section 5 summarizes our conclusions and describes plans for future work.

2. Measuring Productivity

Because our primary goal is to create statistics that provide insight to the official BLS industry-level productivity measures, it is useful to first describe how BLS constructs its measures from published aggregates, and then compare it to our measures that are constructed by aggregating Census microdata.

[5] We use the terms "establishment" and "plant" interchangeably throughout this paper.

2.1. BLS Industry-level Productivity

BLS publishes quarterly and annual measures of labor productivity for major sectors; annual measures of labor productivity for 199 3-digit and 4-digit NAICS industries; and annual measures of multifactor productivity for major sectors, 18 3-digit NAICS manufacturing industries, 86 4-digit NAICS manufacturing industries, the air transportation industry, and the line-haul railroad industry. Productivity growth is measured as the difference between percentage changes in indexes of output and inputs (labor and, in the case of multifactor productivity, capital and intermediate purchases). BLS does not publish industry productivity levels, although they are available on request.

BLS industry output is based on a sectoral concept, which measures the value of goods produced for sale outside the industry.[6] For manufacturing industries, BLS uses published Annual Survey of Manufactures (ASM) and Census of Manufactures (CM) data on the total value of shipments, which it adjusts to remove intrasectoral transactions and resales, and to account for changes in finished goods and work-in-process inventories.[7] This adjusted nominal output measure is then distributed to detailed categories of products and services using the mix of annual wherever-made product shipments from the ASM. Nominal output in each product category is deflated using the appropriate detailed producer price index from the BLS prices program. These real output measures are then Tornqvist-aggregated into industry output indexes. Self-employment revenues for manufacturing firms, which come from Internal Revenue Service data, are also added to these output measures.

BLS measures labor input as the total annual hours worked by all persons in an industry.

---

[6] Sectoral output is less than gross output, but greater than value-added output. In the most detailed industries, sectoral and gross output are the same or very close. However, going from very detailed industries to more aggregated industries, sectoral output moves closer to value-added output. In the limit, at the aggregate level, sectoral output is the same as value-added output, except for imported intermediate inputs.

[7] See https://www.census.gov/programs-surveys/asm.html and https://www.census.gov/programs-surveys/economic-census.html.

This measure is constructed by combining data from three BLS surveys: the Current Employment Statistics (CES) survey, the Current Population Survey (CPS), and the National Compensation Survey (NCS). The CES provides detailed information on the employment and average weekly hours *paid* for production and non-supervisory employees (henceforth referred to as production workers).[8] The NCS data are used to adjust CES average weekly hours from an hours-paid to an hours-worked basis by removing paid vacation accrued and sick leave taken.[9] To estimate nonproduction worker average weekly hours, BLS uses data from the CPS to calculate a ratio of nonproduction to production worker average weekly hours worked, which is then multiplied by the adjusted CES production worker hours (worked). Total nonproduction worker hours are estimated as:

$$\mathrm{TH}_{NP} = \mathrm{Emp}_{\mathrm{NP}}^{\mathrm{CES}} \times \mathrm{AWH}_{\mathrm{P}}^{\mathrm{CES}} \times hwhp_{P}^{NCS} \times \frac{\mathrm{AWH}_{\mathrm{NP}}^{\mathrm{CPS}}}{\mathrm{AWH}_{\mathrm{P}}^{\mathrm{CPS}}} \times 52 \tag{1}$$

where $\mathrm{Emp}_{\mathrm{NP}}^{\mathrm{CES}}$ is nonproduction worker employment from CES, $\mathrm{AWH}_{\mathrm{P}}^{\mathrm{CES}}$ is production worker average weekly hours paid from CES, $hwhp_{P}^{NCS}$ is the hours-worked-to-hours-paid ratio from NCS, and $\left(\frac{\mathrm{AWH}_{\mathrm{NP}}^{\mathrm{CPS}}}{\mathrm{AWH}_{\mathrm{P}}^{\mathrm{CPS}}}\right)$ is the CPS nonproduction/production hours ratio. CPS data are also used to directly obtain hours worked by self-employed and unpaid family workers (Eldridge et al., 2004).

For multifactor productivity at the 4-digit NAICS level, capital input is based on the flow of services from the productive stock of capital. BLS investment data for industries combine expenditures on structures and equipment from the ASM with data on investment in different types of assets by industry from BEA and the Annual Capital Expenditure Survey (ACES). Using a perpetual inventory method, BLS then computes industry-asset level capital

[8] Workers in goods-producing industries are referred to as being production or non-production workers and in the service-providing industries as nonsupervisory or supervisory workers.
[9] Note that this adjustment does not account for off-the-clock hours.

stocks from these investment flows. In BLS official multifactor productivity measures, these stocks are converted to capital services using industry-asset specific rental prices, and then aggregated to the industry level.

For intermediate purchases inputs, BLS combines quantities of materials, purchased business services, fuels, and electricity consumed by each industry. The nominal values of materials, fuels, and electricity are obtained from the CM and ASM. The values of purchased business services are estimated from BEA and Census Bureau data.

2.2 Establishment-level Productivity using Census Data

To measure establishment-level labor productivity, we combine establishment-level information from three Census Bureau restricted-use microdata files with public-use industry-level data from BLS. Given that one of the goals of our research is to shed light on BLS industry productivity statistics, we try to match BLS concepts and measures as closely as possible.

Our establishment-level microdata come from the CM, the ASM, and the Longitudinal Business Database (LBD). The CM is collected every 5 years in years ending in '2' and '7'. Data are collected from all manufacturing establishments except those that are very small. For these very small out-of-scope establishments, the Census Bureau imputes data using administrative records. The ASM sample is a 5-year panel of manufacturing establishments, updated every year for births, and data are collected annually. ASM panels begin in years ending in '4' and '9', and the probability of selection into the ASM sample is a function of both industry and size (generally employment or the value of shipments). Like the CM, the ASM does not collect data from very small establishments but accounts for them using administrative information. In CM years, ASM data are collected as part of the CM data collection, but for this analysis we use only the ASM establishments.[10] Data are imputed for establishments that do not respond or that fail to report some data elements (item non-response); we discuss this further in section 2.3. The LBD is a longitudinally linked version of the Census Bureau's Business Register that covers the non-agricultural employer universe of business establishments (see Jarmin and Miranda, 2002). The LBD provides us with both high-quality longitudinal links and information on the universe of manufacturing establishments, which we use to construct propensity-score weights that we use in our productivity calculations.

Ideally, we would construct an output measure that exactly matches the BLS measure. We start by using Census microdata to replicate the value of shipments as closely as possible. Specifically, we calculate plant-level real output as deflated revenues, adjusted for resales and changes in inventories.[11] But we cannot replicate the BLS sectoral output concept because the ASM does not collect the information needed to calculate intra-sectoral transactions. Instead, we add the value of intra-sectoral transactions back into BLS output measures to make the two measures comparable. Thus, we measure plant-level output as:

$$Q_{et} = (TVS_{et} + DF_{et} + DW_{et} - CR_{et})/PISHIP_{it} \tag{2}$$

where *TVS* = total value of shipments, $DF_{et} = FIE_{et} - FIB_{et}$ and $DW_{et} = WIE_{et} - WIB_{et}$ are the changes in finished-goods and work-in-process inventories respectively (*FIB, FIE* = beginning-of-year and end-of-year finished goods inventories and *WIB, WIE* = beginning-of-year and end-of-year work-in-process inventories), *CR* = cost of resales, *PISHIP* = deflator for the value of shipments,[12] and the *i*, *e*, and *t* subscripts index industries, establishments, and

[10] The microdata made available in the FSRDCs contains productivity measures for all CM establishments when productivity calculation is possible.
[11] In practice, subtracting resales does not make much difference because they are only a small fraction of revenue.
[12] The shipments deflator is constructed as part of NBER-CES data, using price indexes from BEA GDP-by-Industry data. In comparisons to BLS output data conducted in this paper, BLS industry implicit output price deflators (as described in 2.1) are used in place of the shipments deflator.

years.

We measure labor input as total hours worked. For each establishment, the ASM collects the total number of employees, the number of production workers, and the total number of hours worked by production workers. We calculate total annual hours worked by summing ASM production worker hours and an estimate of nonproduction worker hours, which we calculate using the essentially same methodology as BLS (equation 1) but substituting ASM data for CES:[13]

$$TH_{et} = PH_{et} + \left((TE_{et} - PW_{et}) \times \frac{PH_{et}}{PW_{et}} \times \left(\frac{\mathrm{AWH}_{\mathrm{NP}}^{\mathrm{CPS}}}{\mathrm{AWH}_{\mathrm{P}}^{\mathrm{CPS}}} \right) \right) \tag{3}$$

where PH = production worker hours, PW = average number of production workers, TE = total employment, and $\frac{\mathrm{AWH}_{\mathrm{NP}}^{\mathrm{CPS}}}{\mathrm{AWH}_{\mathrm{P}}^{\mathrm{CPS}}}$ = CPS non-production/production average weekly hours ratio.[14] We calculate establishment-level log labor productivity as:

$$lnLP_{et} = lnQ_{et} - lnTH_{et}. \tag{4}$$

Establishment-level multifactor productivity (MFP) in logs is measured as:

$$lnMFP_{et} = lnQ_{et} - \alpha_K lnK_{et} - \alpha_L lnTH_{et} - \alpha_M lnM_{et} - \alpha_E lnE_{et,} \tag{5}$$

where Q and TH are real output and total hours as defined above in (2) and (3), K denotes real productive capital stock, and M and E denote deflated values of expenditures on materials and energy.[15] The productive capital stock is constructed using the perpetual inventory method for equipment and structures separately.[16] The value of M is calculated as the deflated sum of cost

[13] Note that it is not necessary to apply the hours-worked-to-hours-paid ratio because establishments are requested to report hours worked.
[14] These ratios are calculated at the 4-digit industry level.
[15] BLS constructs aggregate MFP using capital services rather than productive stock.
[16] We do not include rented capital due to its irregular collection in the ASM. Pre-1986 and post-2006, this information is collected annually on the ASM. In the intervening years, this information was only collected in the Economic Census. Exploratory analysis for the years when this is available shows rented capital is small and does not make much difference to plant-level capital measures. We plan on exploring this further in future research.

of materials, the cost of resales and the cost of contract work done for the establishment by others.[17] The nominal value of E is obtained as the sum of the cost of electricity and fuels. For the CMP data ultimately included in the DiSP data, the two expenditures are deflated using the appropriate deflators from the NBER-CES database. We measure the factor elasticities, α_K, α_L, α_{E}, and α_{M}, using the share of expenditures of the corresponding input in total cost in each 6-digit NAICS industry.[18] For the BLS data comparisons, we use published BLS industry cost shares; for the DiSP statistics, we use shares calculated from CMP data.

DiSP does not include value-added-based productivity dispersion for conceptual reasons. While there is a market for final demand, a market for value-added does not exist. This consideration is irrelevant for the overall economy because value added equals the value of final demand at that level of aggregation. However, the relationship between value added and final demand is only an approximation at lower levels of aggregation and in order to make inference about final goods produced by a plant or industry, not only are the output and input values necessary but also the relevant input-output linkages.[19] A related issue is that the existence of a value-added production function at the establishment-level requires very strong functional form assumptions that are likely violated (see Basu and Fernald, 1997).

2.3 Missing Data and Imputation

[17] Published ASM measures of value-added are based on the difference between the value of output and a composite operating expenses measure inclusive of materials and energy expenditures. We break out the cost of materials and energy separately in our MFP measure. The inclusion of contract work implies some aspects of purchased services are included in the materials expenditures. However, since 2006, the ASM survey has included questions on other operating expenses including leased employees and additional purchased services not included in the cost of contract work. We are actively exploring the inclusion of those operating expenses for a supplemental multifactor productivity dispersion measure commencing in 2006. Challenges for inclusion of these variables are the short time series, item non-response rates, and the treatment of establishments of single-unit versus multi-unit establishment firms. Establishments of the latter are less likely to have such additional operating expenses because the headquarters establishment of the parent firm may be providing and/or purchasing those services.

[18] See web appendix B in Foster et al. (2016a) for more details. Procedures for extending the output and input price deflators are described in this appendix.

[19] An additional consideration follows when using the logarithmic transformation to stabilize the variance of empirical distributions. This transformation truncates distributions at zero, which leads to biased inference if the probability of negative value added is correlated with plant-level characteristics. In other words, negative value-added may be plausible, for example in intermediate-intensive industries during periods of recession or crises.

As noted above, the ASM microdata are subject to item non-response, and these missing values are imputed by the Census Bureau. The Census Bureau's imputation methods are designed to yield accurate published aggregates but do not necessarily preserve the distribution or adequately reflect the variability of the underlying microdata. There is evidence that certain imputation methods may affect microdata analyses. However, there are techniques available to mitigate the effects of imputation on dispersion measures. For example, White, Petrin, and Reiter (2018) analyze dispersion statistics using classification and regression-tree methods. Foster et al. (2017b) follow a different approach and address imputation by dropping observations with imputed data and reweighting the remaining observations. The results from these studies suggest that imputation yields lower measured dispersion relative to the case when imputation is corrected for. For the purposes of this paper, we consider the entire set of observations in the sample and leave further analysis of these issues for future work.

3. Comparing Micro-Aggregated Data to Published Industry Data

In this section, we compare our micro-aggregated estimates to the official data published by BLS, covering the 1997–2016 period. Based on earlier work comparing similar business data across the two government agencies, we expect that there will be some systematic differences between these measures (Elvery et al., 2006). Even though differences in the levels of the micro and published first moments do not directly affect our conclusions about dispersion (because we sweep out industry-year effects), it is useful to determine how far apart the two sets of estimates are. If the first moments are close, then it is more reasonable to think of micro-based second moments as measuring variation around the published first moments. As an additional check, we compare these two sets of estimates to data from the NBER-CES database. The NBER-CES database is used only for comparisons and should be thought of as equivalent to the official published ASM and CM statistics upon which it is based.[20] We start by comparing input and output measures, and then we compare productivity measures.

3.1. Input and Output Measures

Figure 1 shows the total number of employees in the manufacturing sector from the different series. The first thing to note is that employment levels based on ASM microdata (using ASM sample weights) are significantly lower than the published ASM and BLS estimates because they exclude the "non-mail" stratum—small establishments that are not sampled by the ASM. The published ASM series includes adjustments for the non-mail stratum and is much closer to the BLS estimates.

To account for these small out-of-scope establishments, we construct an alternative set of weights. As noted, the weighted sample total calculated from the ASM (using ASM sample weights) is by design not equal to the published total because there are additional adjustments in the latter for the non-mail cases. Fortunately, there are some very small establishments in the ASM sample each year that are below the thresholds for non-mail cases. This occurs because, among the smaller establishments that were selected for the ASM (that is, establishments with employment above the threshold), some had fallen below the size threshold by the time that they provided data. This implies that there is coverage for all business sizes in the ASM sample (e.g., there are ASM establishments in any given year and industry with 1–4 employees even though this is typically below the ASM sample threshold). To create an alternative set of weights, we use the LBD. Specifically, we define the manufacturing universe using the LBD, and use LBD data to estimate the probability that an establishment is included in the ASM sample, and then use these probabilities to construct inverse propensity score weights (see the Appendix for a full discussion of the weighting procedures). This procedure increases the

[20] For more information on the NBER-CES Manufacturing Industry Database, see http://www.nber.org/nberces/. The NBER-CES series was last updated through 2011 as of September 15, 2020.

weights assigned to the “non-mail” establishments so that the propensity-score-weighted (PSW hereafter) employment totals are consistent with the LBD.[21] Moreover, as can be seen in Figure 1, the micro-aggregated employment series using PSW yields totals that align with the published BLS and ASM more closely than those using ASM weights.[22] This aligns with our objective in using PSW as these weights correct for the contribution of the “non-mail” establishments in a manner that the ASM weights are not designed to address.

We next compare total manufacturing output and input growth between the BLS series, the CMP–PSW series, and the NBER series (Figure 2).[23] For all three series, we use the BLS price deflators for these comparisons, with the exception of capital because BLS does not have separate deflators for equipment and structures. We also adjust the BLS output series to make it comparable to what we can construct from the ASM by adding the value of intrasectoral transactions back into the BLS output series. The BLS and NBER output series track each other very closely, while the CMP series deviates from the other two series in some years but exhibits the same pattern of growth rates, see Panel (a). Panel (b) compares hours growth rates: the growth rates of hours exhibit similar dynamics, except for during the 2005–2007 period, when the NBER series diverges. Panel (c) shows capital stock growth rates. All three series exhibit a slight downward trend in the late 1990s and are essentially flat starting in the early 2000s. The greater year-to-year variation in the CMP series is a due largely to the difficulty of implementing the PIM at the establishment level (see below for further discussion). Panels (d) and (e) show that for energy and material inputs the three series track very closely.[24]

[21] In addition, unreported results suggest that PSW do a good job matching the industry/year-specific size and age distributions of the LBD.
[22] We explored the possibility of benchmarking CMP employment (based on the manufacturing universe in the Census Business Register) to BLS employment (based on the manufacturing universe in the BLS Quarterly Census of Employment and Wages). While this benchmarking, by definition, improves the correlation between labor (employment and hours) measures between BLS and CMP, it actually decreases the correlation between BLS and CMP output and measures of other inputs, which are both based on the manufacturing universe in the Census Business Register.
[23] See Appendix A Table A1 for further details about the construction of these series.
[24] The cost of purchased services and resales are not included in the materials comparisons.

Table 1 shows correlations between the three data sources for inputs and output for the total manufacturing sector. The correlations in the top panel are based on total manufacturing aggregate time series, while the bottom panel shows the average of the within-industry correlations for 4-digit NAICS industries, calculated over the 19 years of the sample. The top panel of the table indicates that hours, energy, materials, and output, both in levels and growth rates, are highly correlated across the data series (the correlations range from 0.88 to 0.99). Average industry-level correlations, shown in the bottom panel, are lower than for total manufacturing, but they are still reasonably high for these variables, both in levels and growth rates.

The correlation between the capital series is significantly lower than the correlations described above. There are several possible explanations. First, there is a fundamental difference in the underlying data. In particular, BLS investment data for 4-digit industries combine expenditures on structures and equipment from the ASM with data on investment in different types of assets by industry from BEA and the Annual Capital Expenditure Survey (ACES). Using a perpetual inventory method, BLS then computes industry capital stocks from these investment flows. In contrast, our approach takes investment flows directly from the establishment and uses these flows with the perpetual inventory method at the plant level to generate capital stocks. Second, given the difference in source data, the BLS investment series covers a longer period of time than the micro-aggregated series. The BLS capital stock is built up from investment flows that stretch back to 1958 (and longer for some assets). For CMP, we initialize new plants using book value, and the earliest book value and investment data that we use dates back to 1972. Despite differences in data sources and methodologies, we can conclude that the micro-aggregated data are largely consistent with published aggregate data.

3.2. Productivity Growth

We calculate productivity growth as the change in the log-productivity for output per hour and multifactor productivity.[25] Figure 3(a) shows that output-per-hour growth rates for the manufacturing sector are broadly similar, with some greater discrepancies in various subperiods (e.g., 2003–2009). These differences can be attributed to the differences in data sources and methodologies, some of which were illustrated in Figure 2. Despite underlying differences, multifactor productivity growth shows remarkable similarity across these data sources, see Figure 3(b). Table 2 echoes these findings: the correlations between the series of different data sources are highest for multifactor productivity growth.

This comparison of inputs, output, and productivity serves as an important backdrop to our new experimental statistics on within-industry dispersion. Although there are some differences between the BLS aggregates and the micro-aggregated series, they are close enough to each other to allow us to make meaningful inferences about the relationship between within-industry dispersion and BLS published estimates of industry productivity growth.

4. Productivity Dispersion

For our analysis of productivity dispersion, we focus on levels rather than growth rates. Since we are interested in comparing within-industry dispersion of productivity across industries and over time, it is necessary to account for industry differences in average productivity. To do this, we calculate establishment-level productivity as the deviation from average productivity in that establishment's 4-digit industry in each year.[26] The interpretation of normalized productivity levels is intuitive: they tell us how far above or below the mean the establishment sits in the productivity distribution.

[25] The official BLS productivity series are calculated using percentage changes in the index, and thus the BLS series that we refer to here differs from the published series. Additionally, the official total manufacturing productivity series is published by the BLS Division of Major Sector Productivity, whereas the data here are aggregated from industry data provided by the BLS Division of Industry Productivity Statistics.

[26] These are weighted averages (using our propensity score weights, see appendix) where establishment-level productivity is expressed as a deviation from average productivity in that establishment's 4-digit industry. This normalizes productivity across industries and thereby accounts for industry differences in average productivity.

We use the interquartile range (IQR) as our primary measure of dispersion, because it is intuitive and easy to interpret. The IQR shows how much more productive an establishment at the 75th percentile of the productivity distribution is than an establishment at the 25th percentile of the productivity distribution. The standard deviation may seem like an obvious alternative to the IQR, but, in addition to being harder to interpret, it is known to be more sensitive to outliers than quantile-based dispersion measures. We also report the 90-10 differential as well as the 10-1 and 99-90 differentials.

Table 3 shows the descriptive statistics of the distribution of dispersion measures.[27] The first entry in the table (0.898) says that in the average industry and year, establishments at the 75th percentile are about ($e^{0.898}\approx$) 2.45 times as productive as those at the 25th percentile. Establishments in the 90th percentile are about 5.9 times as productive as those at the 10th percentile. Average dispersion in multifactor productivity is lower.[28] Still, establishments at the 75th percentile are about 1.7 times as productive as those at the 25th percentile using multifactor productivity measures. This still implies substantial differences in a core measure of business performance at the establishment-level within narrowly-defined industries.

Many factors may underlie the observed dispersion in measured productivity across establishments in the same industry.[29] We define a "wedge" as any mechanism that prevents equalization of marginal revenue products across producers. Because the measures of productivity dispersion reported here are revenue-based, the presence of widespread dispersion is consistent with the presence of some form of wedges. One form is adjustment frictions that inhibit businesses from adjusting their scale of operations and specific inputs to changing economic conditions. These adjustment frictions may be related to the costs of adopting new

[27] We present standard deviations in Table 3 but do not discuss these results.
[28] The range for output per hour is somewhat larger than those found by Syverson (2004a)—he found a multiplier of about 1.9—but our findings are generally in line with his results.
[29] See Syverson (2011) for more detailed discussion of these issues and Blackwood et al (2020) with discussion more closely related to the new data product.

technologies or business practices; thus, dispersion in an industry may reflect the gap between the frontier establishments and other producers. Additional sources of wedges are market distortions such as differences in markups across producers or financial constraints in the same industry. Complicating matters is that in the presence of wedges that are correlated with fundamentals, the variation in the dispersion will also reflect differences in business fundamentals such as technical efficiency and product appeal across businesses, see Blackwood et al. (2020). Adjustment frictions are one source of wedges that yields such a correlation. For example, an increase in the dispersion in product appeal across producers in the presence of adjustment frictions will yield an increase in dispersion of revenue productivity across producers (even if the adjustment frictions remain constant). In a similar fashion, dispersion may reflect unmeasured inputs. These could include production methods, management practices, and the mix of worker types (labor quality).

It is well beyond the scope of the current paper to determine the relative importance of each of these alternative factors. Instead, the objective here is to describe the DiSP data product and point to its potential for investigating these alternative determinants of dispersion. One of the strengths of the new data product is that dispersion measures are provided at a detailed level of aggregation by year. Figure 4a summarizes how within-industry dispersion in output per hour—measured as the IQR—varies across industries and over time. The mean and median IQR are fairly close to each other, but the large differences between the 25th and 75th percentiles indicate that there is a lot of variation in the IQR across industries. For example, in 2002, the productivity difference between establishments at the 75th and 25th percentiles is about 100 log points in the industry at the 75th percentile of the IQR distribution, while this difference is approximately 70 log points in the industry at the 25th percentile of the IQR distribution. These findings are confirmed by multifactor productivity dispersion, with IQRs of approximately 60 and 40 log points at the 75th and 25th percentiles, respectively (see Figure 4b).

The differences in IQRs suggest that there are factors such as those discussed above that generate "dispersion in dispersion," including differences in shocks, adjustment costs, distortions, technology, and distributions of capital intensities. In addition, dispersion is rising during the period under investigation, more so for multifactor productivity than for labor productivity. The rising trend suggests that wedges, and the dispersion in business fundamentals underlying the observed dispersion, are changing in systematic ways over time.[30] We can also see from Figure 4 that although the volatility of (the mean) dispersion is non-trivial, it is dwarfed by the variation across industries. Table 4 shows that there is substantial churning in the ranking of industries in terms of their dispersion. This finding highlights that not only is there dispersion in dispersion but the cross-sectional variation across industries varies over time.

For the remainder of this section, we consider two extensions to our analysis to illustrate further the nature of the dispersion. We first examine how our results change when we weight establishments using activity weights. Activity weights are generated by multiplying our PSWs by an activity measure such as employment or hours. Activity weighting paints a potentially different picture of dispersion because there may be differences between the dispersion of different size groups.[31] Our second extension is to examine the tails of the productivity distribution. There has been great interest in the finding that a substantial portion of wage inequality is driven by the upper tail of the distribution and by increasing between-establishment wage differentials. Investigating the upper tail of the productivity distribution is analogously interesting, as theory and evidence show that the productivity and earnings distributions are related.

[30] There is an ongoing debate about the source of the rising dispersion in revenue productivity measures. See, e.g., Bils et al. (2020), Blackwood et al (2020) and Decker et al. (2020). We do not seek to directly address that debate here but note that Decker et al. (2020) find that rising (revenue-based) labor productivity dispersion in manufacturing is present in both the ASM survey data used here and in administrative data from the Business Register. This finding suggests that rising measurement error is not driving the rising dispersion.

[31] For example, employment-weighted distributions tell us about the productivity of a worker at a given percentile, where the productivity of a worker is the average productivity of the establishment where he or she works.

4.1. Activity-Weighted Dispersion Measures

Figure 5 replicates the dispersion measures in Figure 4 using propensity score weights that have been reweighted using activity weights, which we will refer to as "activity-weighted" or simply "weighted" distributions. For output per hour, the activity weights are defined by hours shares (the share of a given plant's hours of the total hours in its industry) while for multifactor productivity they are defined by composite input shares. The main difference relative to previous results is that activity-weighted dispersion is smaller and exhibits less year-to-year variation. These findings suggest that although dispersion seems smaller among large establishments, it has been rising over the period under investigation, and is likely to be the main driver behind the increase in the mean IQR.

4.2. Dispersion in the Tails

Turning to the tails of the productivity distributions, a distinctive feature of the within-industry productivity distribution is that mean and median dispersion in the right tail (the 99-to-90 ratio) are about the same order of magnitude as the mean and median dispersion in the center of the support as measured by the IQR (see Figure 6). This is remarkable given that each tail covers only one-fifth as many establishments as the IQR. Comparing Figure 6a to Figure 4a, output per hour differences among the most productive establishments are slightly lower than differences among those around the average. In contrast, multifactor productivity differences among the most productive establishments are larger relative to those in the center of the support. In addition, these differences in the right tail of multifactor productivity are rising faster: the mean indicates that dispersion in the right tail rose by about 40 log points between 1997 and 2016 (compare Figures 6b and 4b). The weighted dispersion measures (Figure 7) show generally similar patterns but a smaller absolute increase. Also, it is notable that dispersion in dispersion is substantial in the right tail.

In contrast to the right tail, the left tail (the 10-to-1 ratio) exhibits lower dispersion relative to the center. In particular, mean output-per-hour differences among the least productive establishments are 20–30 log-points smaller relative to the set of establishments in the middle, though exhibit similar volatility (see Figure 8a). Mean multifactor productivity differences in the left tail are similarly smaller initially than in the center and exhibit no positive trend, see Figure 8b.[32] The weighted dispersion measures (Figure 9) tell a similar story.

These findings highlight the importance of looking at the entire distribution. The IQR is a convenient measure that covers half of the distribution. However, there is generally comparable dispersion in the upper and lower tails as there is in the middle. We also see that weighting matters: accounting for size tends to reduce both productivity dispersion and its volatility.

*4.3 Establishment and Firm Characteristics and Within-Industry Dispersi*on

Many factors may underlie the substantial dispersion in productivity measures across establishments within the same industry as well as the variation in the dispersion measures between measures over time. To provide more guidance on the potential driving forces, we examine the relationship between productivity and observable establishment characteristics. For this purpose, we examine spatial variation (by state), establishment size and age.

Table 5 shows R^2 and p-values from F-tests from regressions of establishment-level productivity on individual observable establishment characteristics. The first entry of the table shows that the variation across 4-digit NAICS industries explains approximately 34 percent of the overall variance in establishment-level log output per hour. The explanatory power of the other characteristics is similarly significant, but an order of magnitude smaller. The implications

[32] The spike in mean dispersion in 1998 is due to transitory changes in the following 4-digit NAICS industries: 3344, 3345, 3341, 3342 and 3351. In these industries, the least productive establishments shifted to the left in 1998 and then back to right in 1999. The significant changes in production technologies in these industries (factoryless manufacturing and offshoring) may explain these transitory dynamics in these years.

for MFP are similar, see the lower panel of Table 5. In the right panel of Table 5, we control for industry and time variation by removing industry and year effects from the productivity measures and find similar results. Our results suggest that differences in size, age, location, time or even industry classification account for only a fraction of the observed differences in the productivity levels across establishments. Further research is necessary to uncover the contribution of other factors and our hope is that the DiSP data product will prove useful in this regard.

4.4 Description of the Dispersion Statistics on Productivity (DiSP) Data Product

The new data product, Dispersion Statistics on Productivity (DiSP), which will be updated by the two agencies on an annual basis,[33] contains a balanced panel of productivity statistics summarizing the within-industry distributions of output per hour and multifactor productivity.[34] Dispersion statistics include standard deviations, interquartile and interdecile ranges of the within-industry distributions of plant-level productivity. All data moments are frequency weighted; see sections 2.2 and Appendix B.2. In addition, the dataset includes activity-weighted versions of dispersion measures. The 99-to-90 and 10-to-1 ranges are currently under consideration for future releases given the interesting patterns in the right and left tails highlighted above.

The data product will be useful in analyzing the relationships between productivity dynamics at the plant-level, industry-level and for the entire manufacturing sector. As discussed above, many factors may underlie the cross-industry and time-series variation in dispersion. We anticipate that this new data product will facilitate our understanding of the connection between

[33] The timeliness of the data depends on the release of establishment- and firm-level information. In non-Census years, the ASM is available in the fall of the subsequent year, while the LBD becomes available in spring of the year after. In Census years, microdata become available later. The productivity dataset can be created approximately 2–3 months after the underlying microdata becomes available.
[34] See both BLS and Census Bureau websites: https://www.bls.gov/lpc/productivity-dispersion.htm and https://www.census.gov/disp.

micro- and macro-level productivity. A key benefit of making these data available will be to allow researchers without access to the confidential microdata to explore the various possible causes—and effects—of the differences in within-industry dispersion across industries and over time.

5. Concluding Remarks

A growing literature uses micro-level data to examine establishment-level productivity dynamics and finds substantial within-industry productivity dispersion. This paper provides an overview of a new data product, Dispersion Statistics on Productivity (DiSP), released jointly by BLS and the Census Bureau. This new data product provides measures of productivity dispersion within narrowly defined industries by year.

Much of the paper discusses the measurement methodology used to produce this data product. We compare inputs and output aggregated from micro-level data, to BLS aggregates at the industry and manufacturing-wide level that are part of the official industry and manufacturing-level productivity statistics produced and released by BLS. Not surprisingly, we find some differences between BLS industry-level data and micro-aggregated ASM data; however, in general, we find high correlations between BLS and micro-aggregated outputs and inputs (for example, at the total manufacturing level the correlation between the BLS published series and the micro-aggregated data for output and hours growth are both about 0.9).

Using measures of inputs and output, we develop measures of labor productivity (output per hour) and multifactor productivity and examine some of their properties. Correlations between BLS and micro-aggregated labor productivity growth are also reasonably high and especially high for multifactor productivity growth (e.g., at the total manufacturing level the MFP growth correlation is 0.94).

Illustrating the properties of the new data product, we find large within-industry dispersion in labor productivity: establishments at the 75th percentile are about 2.4 times as productive as those at the 25th percentile on average. For multifactor productivity, we find that the analogous ratio is 1.7. These patterns indicate enormous differences in measures of business performance across plants in the same narrowly defined industry and year. Differences might stem from many factors but they highlight both great potential for growth (e.g., if the gaps between high- and low-productivity businesses could be reduced) and also possible sources of frictions or distortions that are impeding a more efficient allocation of resources.

A core feature of the DiSP data product is the release of measures of within-industry productivity dispersion by detailed industry. We find significant dispersion in within-industry dispersion across industries. For the top quartile of industries, the ratio of multifactor productivity across plants implied by the IQR exceeds 1.7 while for the bottom quartile of industries the ratio is lower than 1.4. Dispersion in dispersion over time is small by comparison, but it is likely still important. There is rising dispersion in both labor productivity and multifactor productivity, which is the source of ongoing debate in the academic literature.

Our results also indicate that average dispersion depends on where we measure it: average dispersion is greater as we move further away from the center of the support of the within-industry productivity distribution. Specifically, average productivity differences across establishments are largest in the right tail. Similar to what we find for average dispersion, the dynamics of these measures depend on where we measure productivity differences. We find evidence that dispersion among the most productive establishments has been increasing during our sample period, while differences in the left tail do not show these patterns. This suggests that positively trending dispersion found in earlier studies may be a consequence of the dynamics of the most productive establishments. Our analysis suggests these patterns are sensitive to how dispersion is measured. We find that activity weights generally imply smaller, less volatile productivity differences among establishments in the entire distribution. We also find that, on average, weighted dispersion among more productive establishments shows a more pronounced positive trend.

In future work, we plan to explore extending the data product in a number of directions. As noted, one active area of exploration is releasing statistics on the tails of the productivity distribution. Another area of exploration is to release statistics by additional characteristics such as firm age and firm size. Research has shown that young businesses exhibit especially high productivity dispersion. This may reflect greater experimentation by young businesses as well as greater challenges that young businesses face in changing their scale of operations.

References

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. "Demand Fluctuations in the Ready-Mix Concrete Industry," *Journal of Political Economy*, 122(5), 1013–63.

Baily, Martin N., Charles Hulten, and David Campbell. 1992. "Productivity Dynamics in Manufacturing Plants," *Brookings Papers on Economic Activity. Microeconomics*, 187–267.

Barth, Erling, Alex Bryson, James C. Davis, and Richard Freeman. 2016. "It's Where You Work: Increases in Earnings Dispersion across Establishments and Individuals in the United States," *Journal of Labor Economics,* 34(2), S67–S97.

Basu, Susanto and John G. Fernald. 1997. "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, 105(2), 249–283.

Bils, Mark, Peter J. Klenow, and Cian Ruane. 2020. "Misallocation or Mismeasurement?" NBER Working Paper No. 26711, National Bureau of Economic Research, Inc.

Blackwood, G. Jacob, Lucia S. Foster, Cheryl A. Grim, John Haltiwanger, and Zoltan Wolf. (2020). "Macro and Micro Dynamics of Productivity: From Devilish Details to Insights," *American Economic Journal: Macroeconomics*, (forthcoming).

Burdett, Kenneth, and Dale T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment," *International Economic Review,* 39(2), 257–73.

Davis, Steven J., and John Haltiwanger. 1991. "Wage Dispersion Between and Within U.S. Manufacturing Plants, 1963-1986," NBER Working Paper No. 3722, National Bureau of Economic Research, Inc.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction*. Cambridge MA: MIT Press.

Decker, Ryan, John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2020. "Changing Business Dynamism and Productivity: Shocks vs. Responsiveness," *American Economic Review* (forthcoming).

Dunne, Timothy, Lucia Foster, John Haltiwanger, and Kenneth R. Troske. 2004. "Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment," *Journal of Labor Economics*, 22(2), 397–430.

Eldridge, Lucy P., Sabrina Wulff Pabilonia, and Jay Stewart. 2019. "Improving Estimates of Hours Worked for U.S. Productivity Measurement," Unpublished manuscript.

Eldridge, Lucy P., Marilyn E. Manser, and Phyllis F. Otto. 2004. "Alternative Measures of Supervisory Employee Hours and Productivity Growth," *Monthly Labor Review*, 27(4), 9–28.

Elvery, Joel, Lucia Foster, C.J. Krizan, and David Talan. 2006. "Preliminary Micro Data Results from the Business List Comparison Project," 2006 Proceedings of the American Statistical

Associations, Business and Economics Statistics Section [CD-ROM], Alexandria, VA: American Statistical Association.

Fisher, Sylvia, Karen Goldenberg, Eileen O'Brien, Clyde Tucker, and Diane Willimack. 2001. "Measuring Employee Hours in Government Surveys," Paper presented to the Federal Economic Statistics Advisory Committee (FESAC), June 7.

Fort, Teresa. 2013. "NAICS 2002 Code Assignments to the Longitudinal Business Database." Mimeo, Center for Economic Studies.

Foster, Lucia, John Haltiwanger, and C.J. Krizan. 2001. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence," in New Directions in Productivity Analysis (eds. Edward Dean, Michael Harper and Charles Hulten), University of Chicago Press, 303-372.

Foster, Lucia, Cheryl Grim, and John Haltiwanger. 2016a. "Reallocation in the Great Recession: Cleansing or Not?" *Journal of Labor Economics,* 34(S1), S293−S331.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2016b. "Firm-Level Dispersion in Productivity: Is the Devil in the Details?" *American Economic Review,* 106(5), 95–98.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2017a. "Innovation, Productivity Growth and Productivity Dispersion," prepared for the 2017 CRIW.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2017b. "Micro and Macro Dynamics of Productivity: From Devilish Details to Insights," NBER Working Paper No. 23666, National Bureau of Economic Research, Inc.

Frazis, Harley and Jay Stewart. 2004. "What Can Time Use Data Tell Us About Hours of Work?" *Monthly Labor Review,* 127(12), December, 3−9.

Goldenberg, Karen and Diane Willimack. 2003. "Measurement Differences in Key Economic Indicators" JSM Papers and Proceedings.

Griliches, Zvi, and Haim Regev. 1992. "Productivity and Firm Turnover in Israeli Industry: 1979-1988," NBER Working Paper No. 4059, National Bureau of Economic Research, Inc.

Haltiwanger, John and James Spletzer. 2020. "Between Firm Changes in Earnings Inequality: The Role of Industry and Occupation Effects." NBER Working Paper No. 26786, National Bureau of Economic Research, Inc.

Hseih, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics,* 124(4), 1403−1448.

Hseih, Chang-Tai, and Peter J. Klenow. 2014. "The Life-Cycle of Manufacturing Plants in India and Mexico," *Quarterly Journal of Economics*, 129(3), 1035−1084.

Jarmin, Ron S., and Javier Miranda. 2002. "The Longitudinal Business Database." Center for Economic Studies Discussion Paper, No. 02-17.

Petrin, Amil, T. Kirk White, and Jerome P Reiter. 2011. "The Impact of Plant-level Resource Reallocations and Technical Progress on US Macroeconomic Growth," *Review of Economic Dynamics,* 14(1), 3−26.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, Till von Wachter. 2019. "Firming Up Inequality," *The Quarterly Journal of Economics*, 134(1), 1–50.

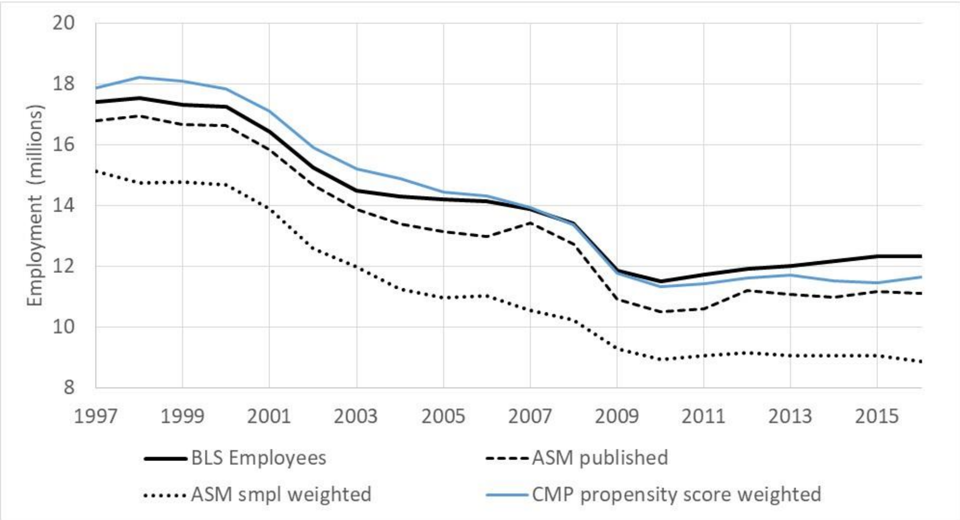
Syverson, Chad. 2004a. "Product Substitutability and Productivity Dispersion." *The Review of Economics and Statistics*, 86(2), 534−550.

Syverson, Chad. 2004b. "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy*, 112(6), 1181−222.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49(2), 326−365.

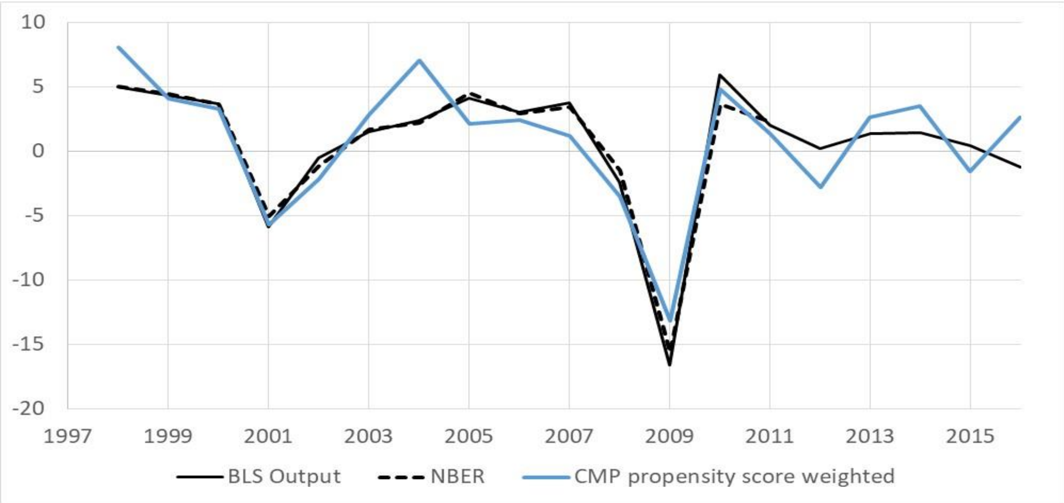
White, T. Kirk, Jerome P. Reiter, and Amil Petrin. 2018. "Imputation in U.S. Manufacturing Data and Its Implications for Productivity Dispersion," *Review of Economics and Statistics*, 100(3), 502-509.

Figure 1. Manufacturing Employment Levels, 1997–2016

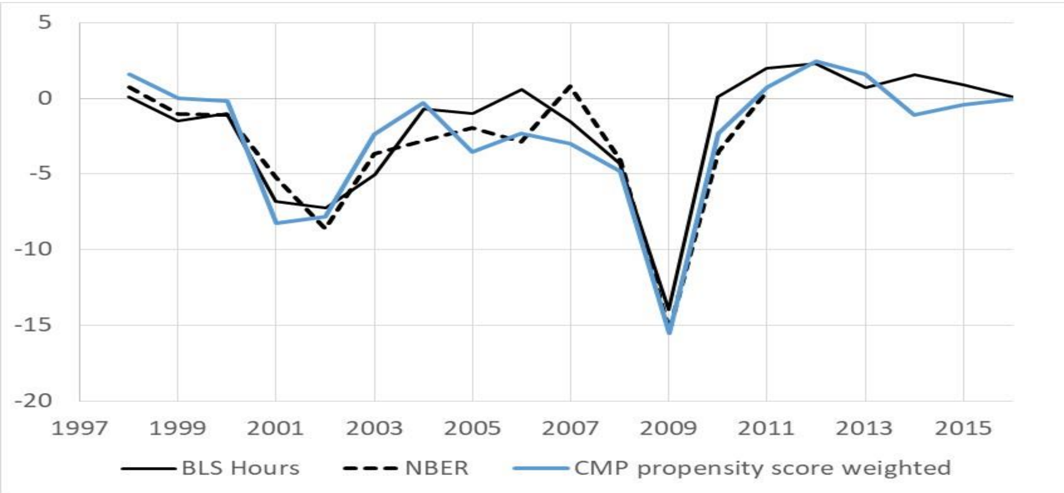


Source: "BLS-Employees" is the annual average of the not seasonally adjusted employment in manufacturing [CEU3000000001, Current Employment Statistics program]. "ASM Published" is the published aggregate employment series from the ASM. CMP denotes micro-aggregated series using the ASM: CMP-ASM Sample Weighted total employment is calculated using ASM sample weights. CMP-Propensity Score Weighted total employment is calculated using our estimated inverse propensity score weights.

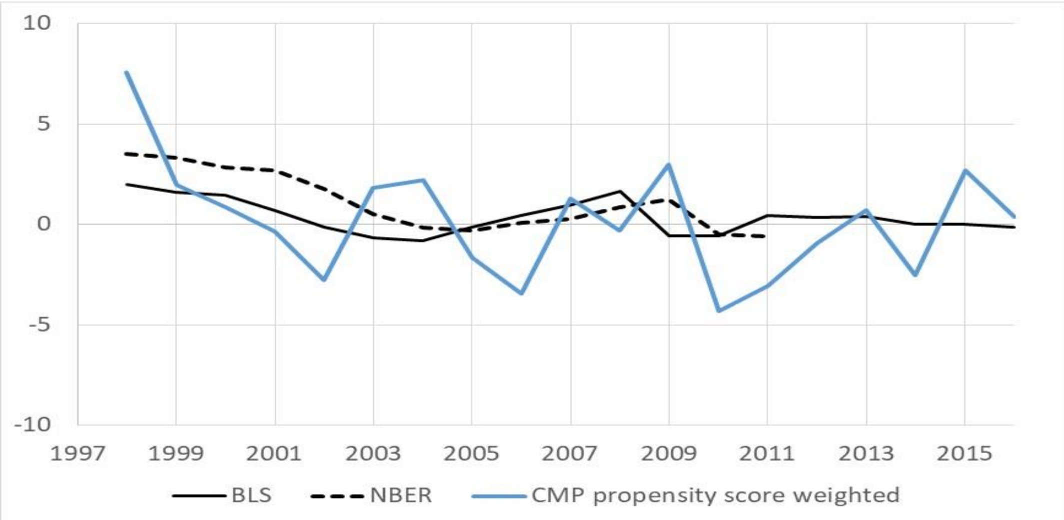
Figure 2. Manufacturing Output and Inputs, 1998–2016



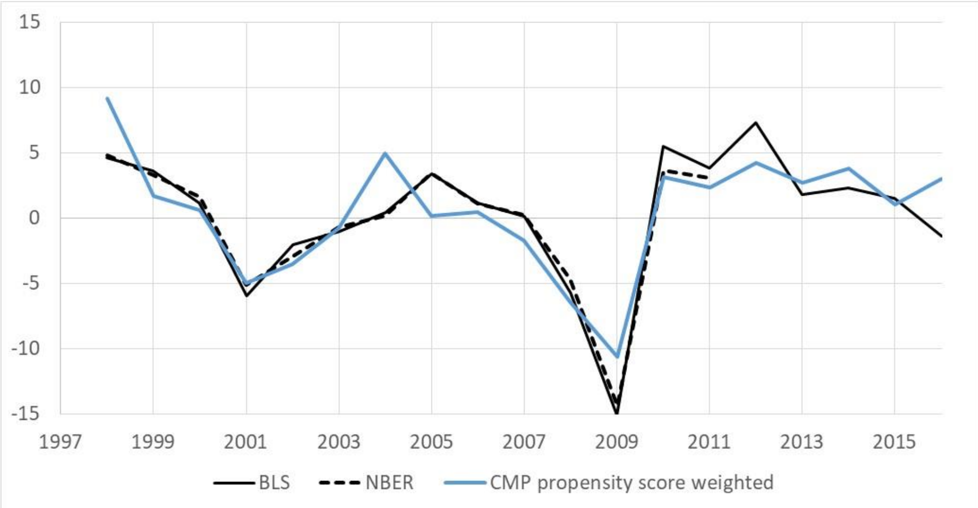
a) Output



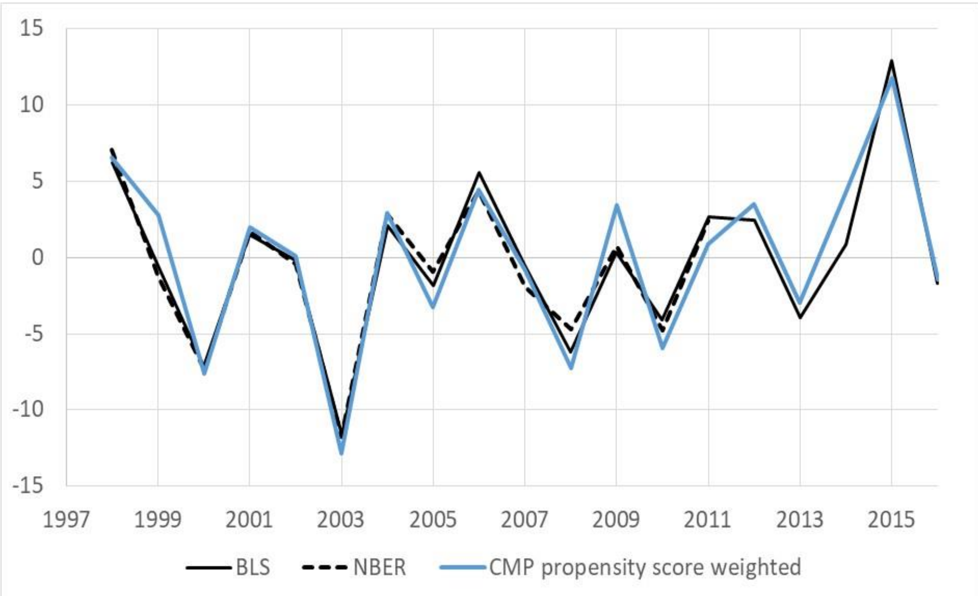
b) Hours



c) Capital



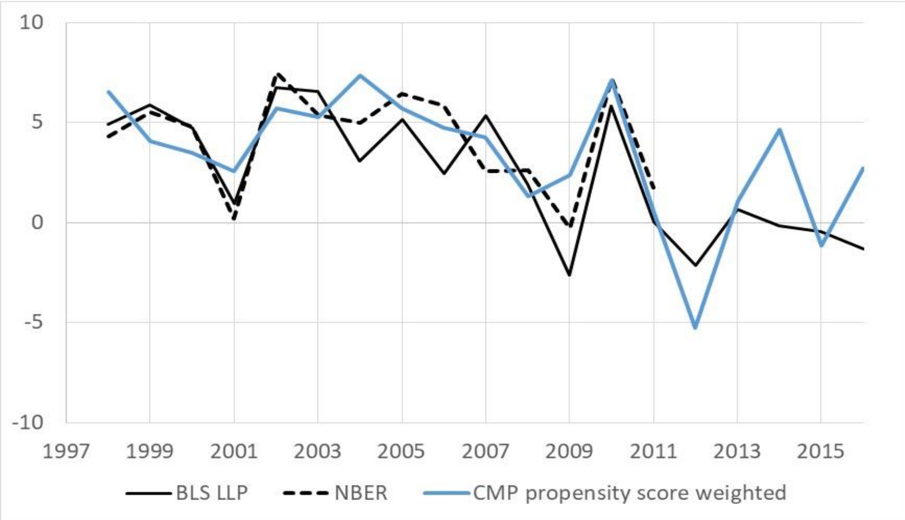
d) Materials



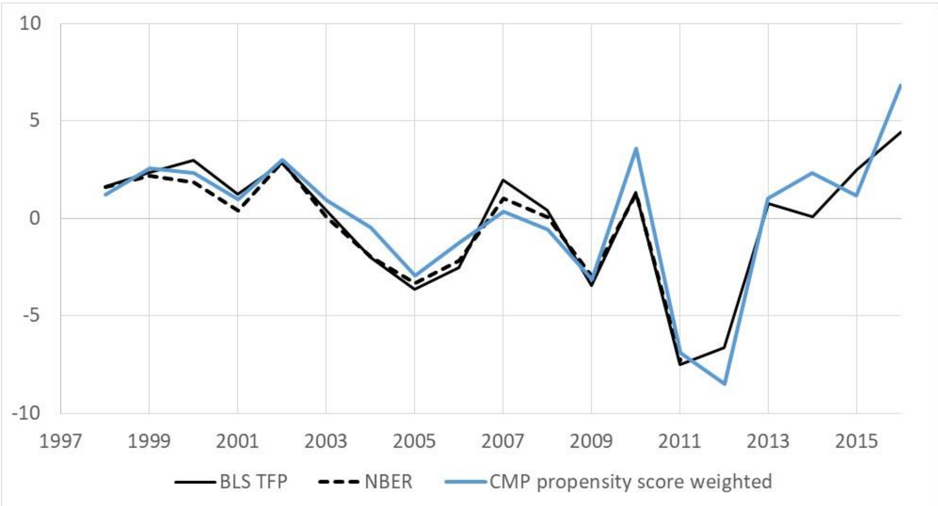
e) Energy

Source: "BLS" is the authors' calculations from Industry Productivity Program data. "CMP – Propensity Score Weight" is the authors' calculations on the ASM. "NBER" is the authors' calculations on the NBER database

Figure 3. Productivity Growth over Sources and Measures, 1998–2016



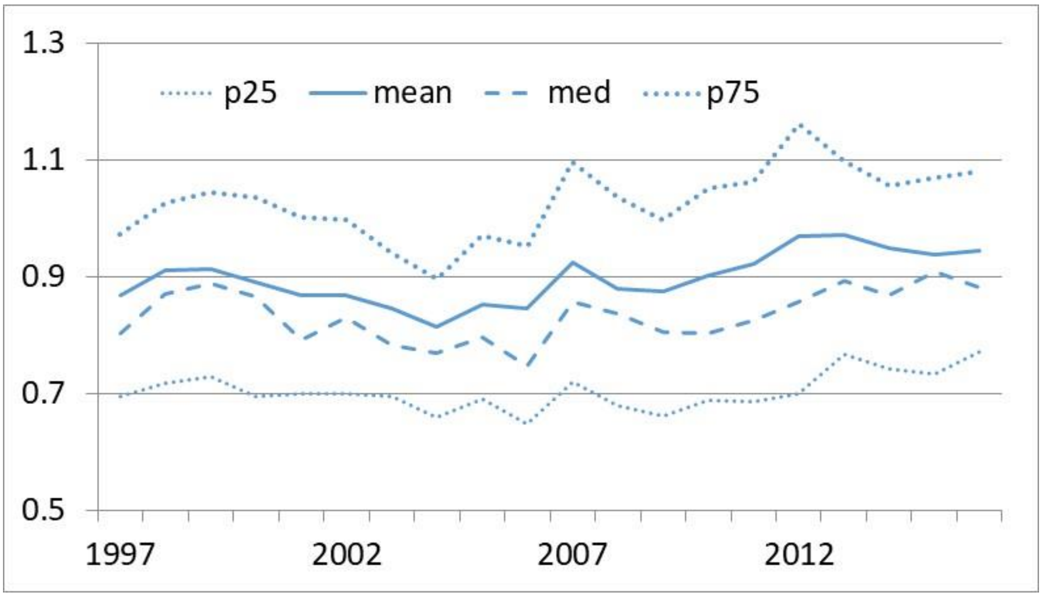
a) Output per hour



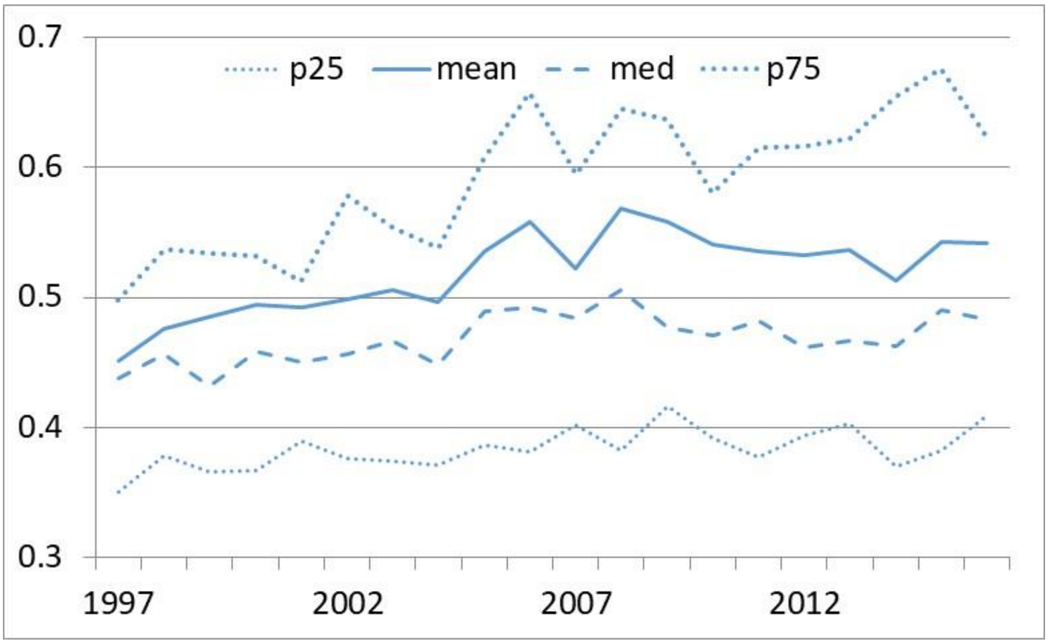
b) Multifactor productivity

Source: See the notes to Figures 2.

Figure 4. Distribution of IQR of Productivity, 1997–2016



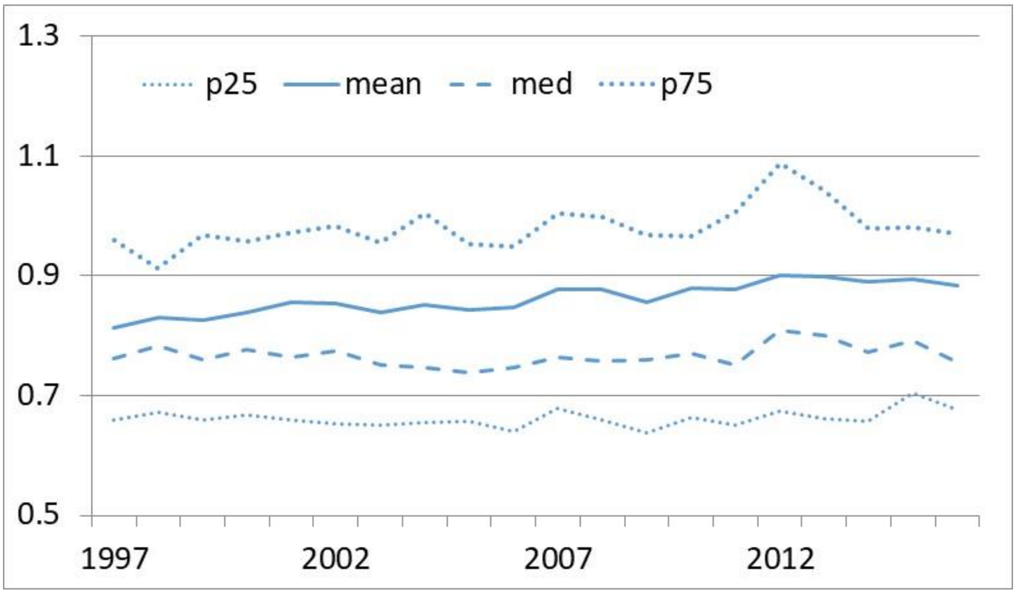
(a) Output per hour



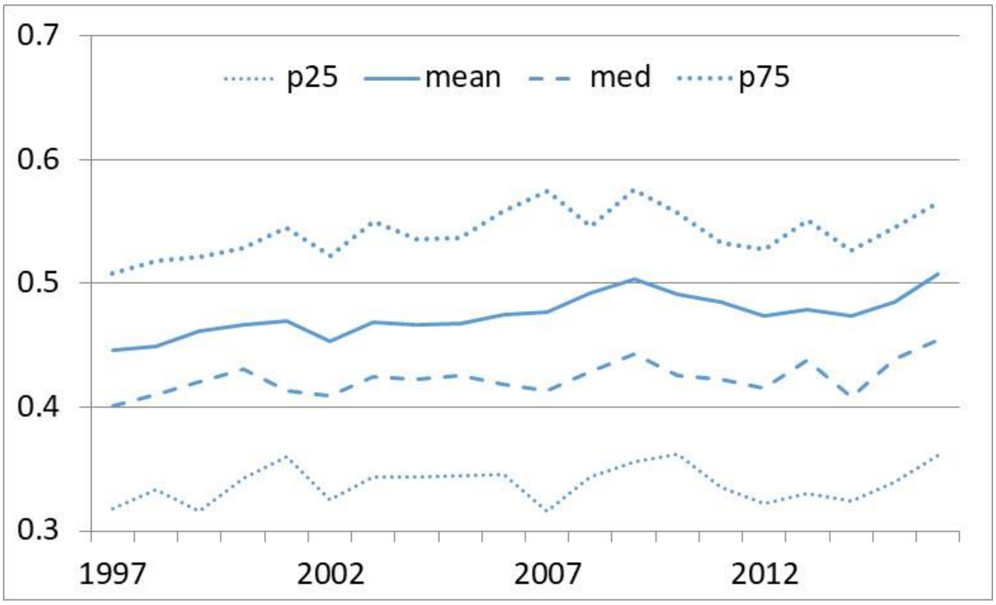
(b) Multifactor productivity

Source: Authors' calculations on the ASM. Notes: Within-industry productivity moments are created at the 4-digit NAICS level, weighted by our propensity score weight. Annual descriptive statistics of industry dispersion are unweighted.

Figure 5. Distribution of *Weighted* IQR of Productivity, 1997–2016



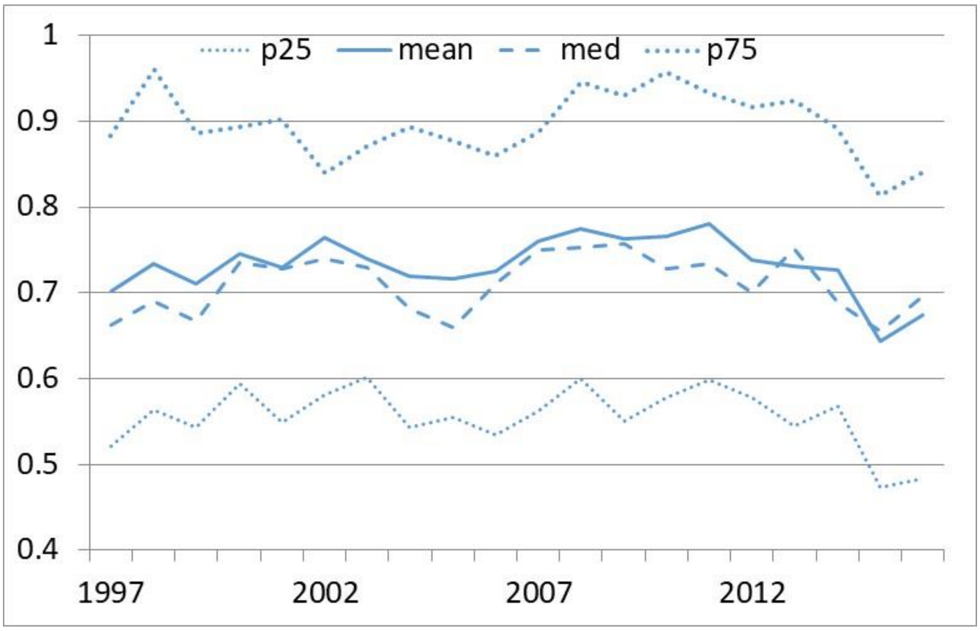
(a) Output per hour (hours-weighted)



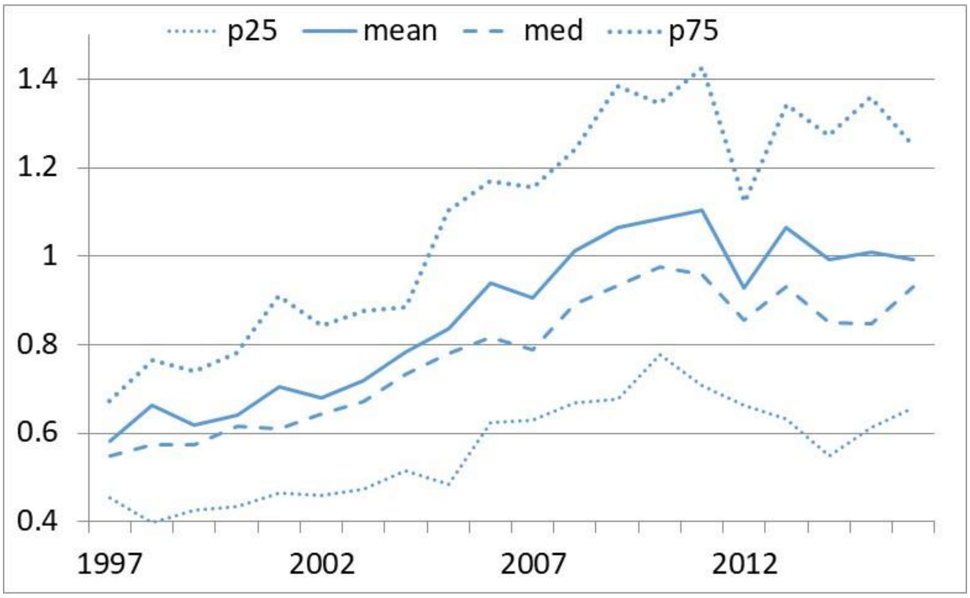
(b) Multifactor productivity (composite-input-weighted)

Source: See notes to Figure 4.

Figure 6. Distribution of 99-90 Difference of Productivity, 1997–2016



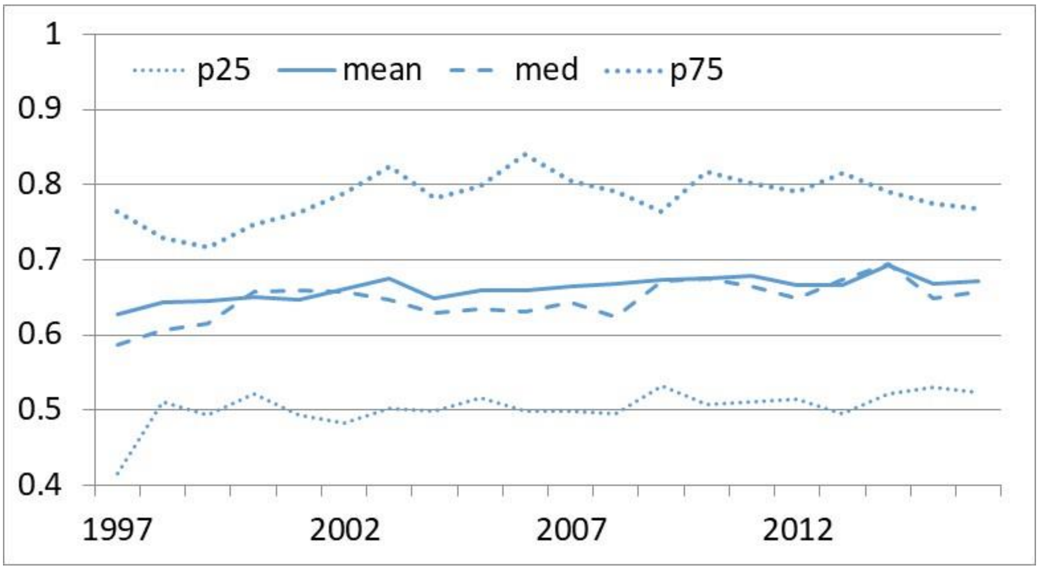
(a) Output per hour



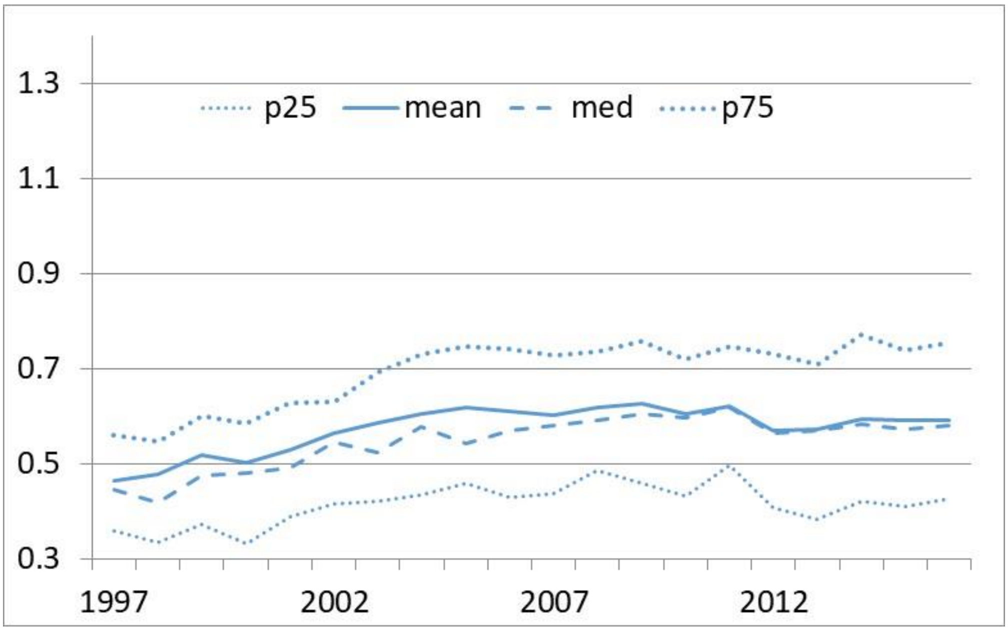
(b) Multifactor productivity

Source: See notes to Figure 4.

Figure 7. Distribution of *Weighted* 99-90 Difference of Productivity, 1997–2016



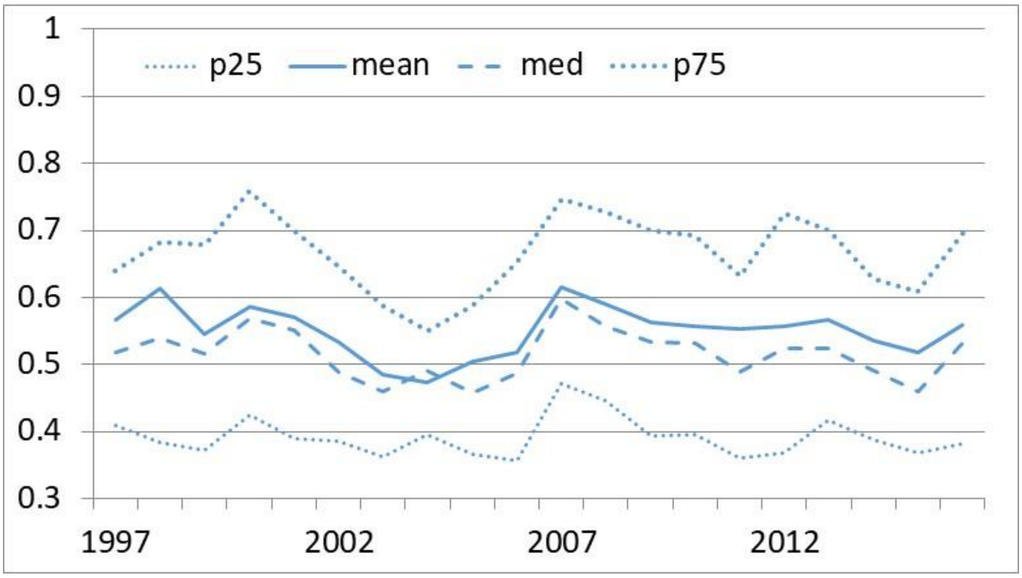
(a) Output per hour (hours-weighted)



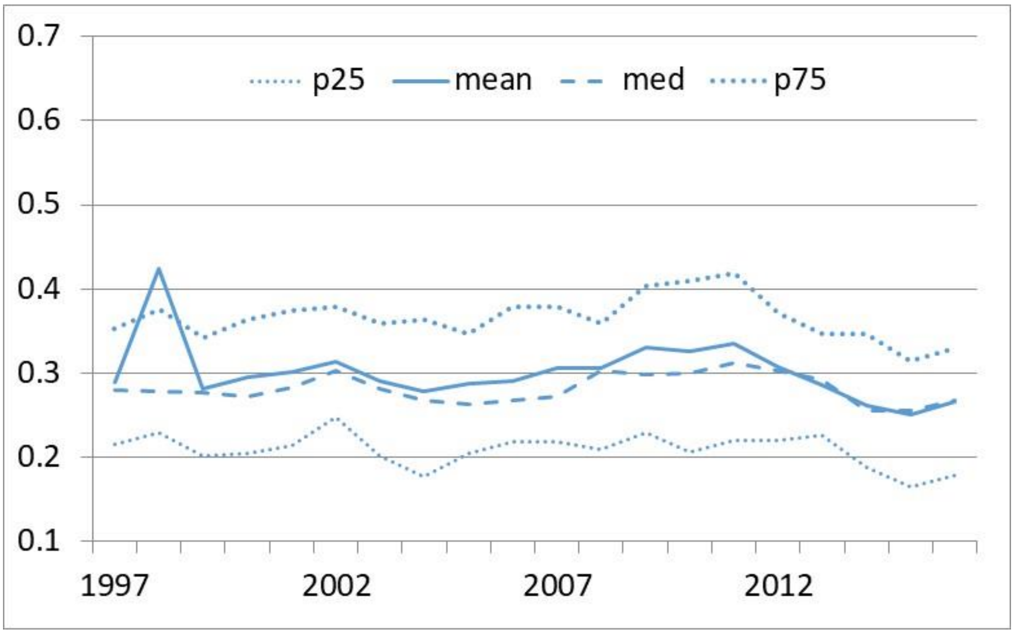
(b) Multifactor productivity (composite-input-weighted)

Source: See notes to Figure 4.

Figure 8. Distribution of 10-1 Difference of Productivity, 1997–2016



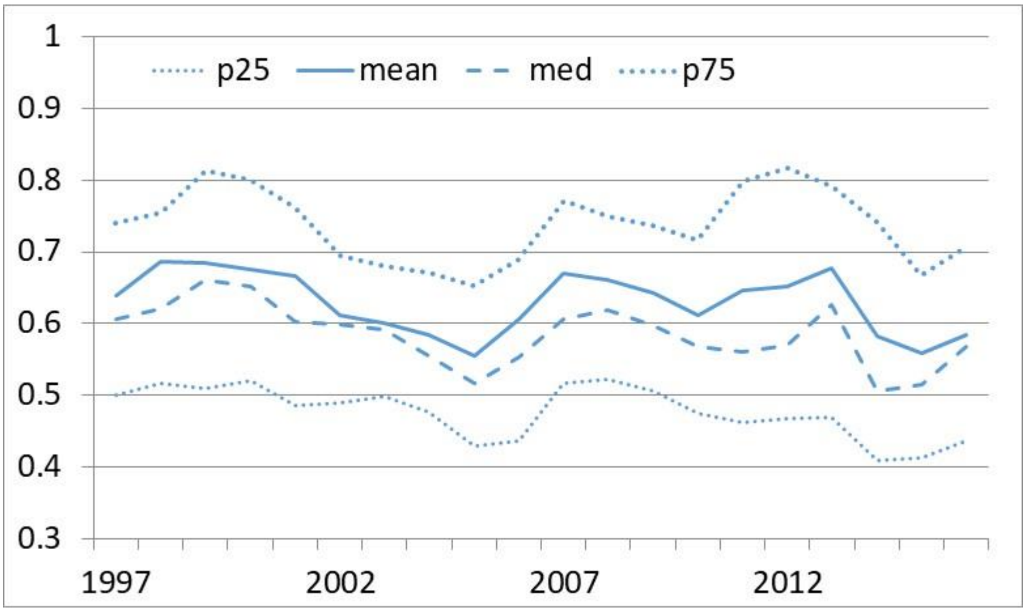
(a) Output per hour



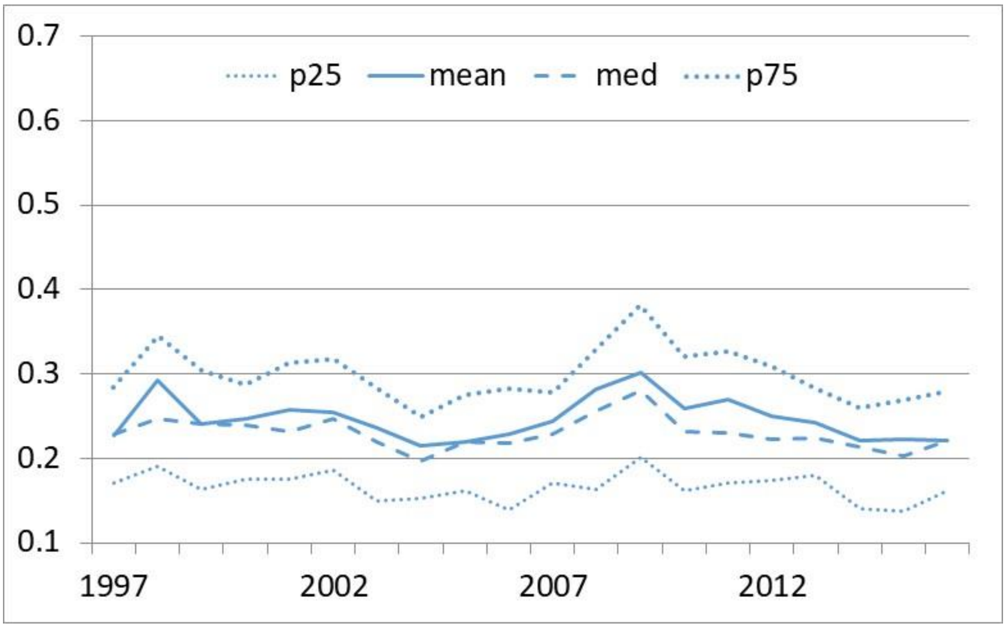
(b) Multifactor productivity

Source: See notes to Figure 4.

Figure 9. Distribution of *Weighted* 10-1 Difference of Productivity, 1997–2016



(a) Output per hour (hours-weighted)



(b) Multifactor productivity (composite-input-weighted)

Source: See notes to Figure 4.

Table 1. Input and Output Correlations between BLS, CMP, and NBER (1997–2016)

	BLS/CMP	BLS/NBER	CMP/NBER
Total Manufacturing			
Hours worked, levels	0.986	0.994	0.996
Hours worked, growth	0.930	0.909	0.923
Capital, levels	-0.184	0.880	0.328
Capital, growth	0.330	0.643	0.585
Energy, levels	0.977	0.997	0.978
Energy, growth	0.958	0.985	0.948
Materials, levels	0.931	0.996	0.963
Materials, growth	0.858	0.992	0.880
Output, levels	0.945	0.996	0.961
Output, growth	0.894	0.993	0.926
Average of 4-Digit NAICS			
Hours worked, levels	0.803	0.889	0.883
Hours worked, growth	0.457	0.632	0.599
Capital, levels	0.521	0.714	0.489
Capital, growth	0.265	0.565	0.263
Energy, levels	0.861	0.986	0.841
Energy, growth	0.714	0.714	0.709
Materials, levels	0.837	0.962	0.843
Materials, growth	0.659	0.937	0.661
Output, levels	0.838	0.985	0.850
Output, growth	0.676	0.951	0.675

Source: Authors' calculations on the ASM.

Table 2. Productivity Growth Correlations between BLS, CMP, and NBER (1997–2016)

	BLS/CMP	BLS/NBER	CMP/NBER
Labor productivity (Total Manufacturing)	0.705	0.818	0.735
Labor productivity (Average of 4-Digit NAICS)	0.465	0.619	0.660
Multifactor productivity (Total Manufacturing)	0.935	0.991	0.960
Multifactor productivity (Average of 4-Digit NAICS)	0.786	0.896	0.810

Source: Authors' calculations on the ASM.

Table 3. Summary of Within-Industry Productivity Distributions (1997–2016)

Within-Industry productivity moment	Mean	Standard deviation	IQR
Labor Productivity			
IQR	0.898	0.290	0.322
90-10 differential	1.773	0.476	0.613
Standard deviation	0.684	0.167	0.222
99-90 differential	0.732	0.279	0.333
10-1 differential	0.550	0.267	0.275
Multifactor productivity			
IQR	0.520	0.222	0.205
90-10 differential	1.078	0.393	0.371
Standard deviation	0.460	0.152	0.161
99-90 differential	0.866	0.512	0.546
10-1 differential	0.301	0.181	0.153

Source: Authors' calculations on the ASM.
Notes: Log labor productivity is calculated as log (output/hours) where hours are BLS-adjusted total hours. The 4-digit NAICS industry mean log LP is subtracted off establishment-level log LP. Within-industry productivity moments are created at the 4-digit NAICS level using propensity score weights. Annual summary statistics of these industry statistics are then created weighting each industry equally. The numbers shown are means of the annual summary statistic values from 1997–2016 weighting each year equally.

Table 4. Probability of Transition across Quintiles of the Cross-Industry Distribution of Dispersion (annual averages between 1997 and 2016)

	IQR						IQR (Activity weighted)				
Output per hour											
	1	2	3	4	5		1	2	3	4	5
1	0.63	0.24	0.08	0.03	0.01	1	0.70	0.22	0.06	0.01	0.00
2	0.23	0.41	0.22	0.10	0.02	2	0.22	0.49	0.24	0.04	0.00
3	0.10	0.25	0.41	0.25	0.03	3	0.06	0.25	0.56	0.15	0.01
4	0.03	0.08	0.25	0.44	0.19	4	0.01	0.04	0.14	0.65	0.15
5	0.01	0.03	0.03	0.18	0.74	5	0.00	0.00	0.01	0.15	0.84
Multifactor productivity											
	1	2	3	4	5		1	2	3	4	5
1	0.64	0.23	0.07	0.04	0.01	1	0.75	0.19	0.03	0.01	0.01
2	0.22	0.42	0.28	0.05	0.02	2	0.19	0.53	0.21	0.05	0.00
3	0.08	0.26	0.40	0.26	0.04	3	0.03	0.24	0.53	0.20	0.03
4	0.04	0.08	0.20	0.47	0.18	4	0.02	0.02	0.20	0.56	0.17
5	0.01	0.02	0.05	0.16	0.75	5	0.01	0.01	0.03	0.17	0.79

Source: Authors' calculations on the ASM. Rows index quintiles in *t-1*, columns index quintiles in *t*. Probabilities in each table are normalized by column sums, i.e. column elements sum to one, apart from rounding.

Table 5. Relationship between Productivity and Establishment Characteristics, 1997–2016

Characteristic	R^2	Pr > F	R^2	Pr > F
	LogLP		LogLP (demeaned)	
naics4	0.340	<.0001		
year	0.030	<.0001		
fipsst	0.010	<.0001	0.0032	<.0001
sizeclass 1	0.020	<.0001	0.0117	<.0001
sizeclass 2	0.022	<.0001	0.0131	<.0001
sizeclass 3	0.023	<.0001	0.0139	<.0001
ageclass	0.013	<.0001	0.0073	<.0001
	LogMFP		LogMFP (demeaned)	
naics4	0.370	<.0001		
year	0.002	<.0001		
fipsst	0.020	<.0001	0.0062	<.0001
sizeclass1	0.010	<.0001	0.0042	<.0001
sizeclass2	0.012	<.0001	0.0045	<.0001
sizeclass3	0.011	<.0001	0.0044	<.0001
ageclass	0.001	<.0001	0.0009	<.0001

Source: Authors' calculations on the ASM. LogLP: log labor productivity, LogLP (demeaned): log labor productivity, industry and year effects are removed, LogMFP: log multifactor productivity, LogMFP (demeaned): log multifactor productivity, industry and year effects are removed, naics4: 4-digit NAICS code, year: time identifier, fipsst: Federal Information Processing Standard state code, sizeclass 1: employment size class with thresholds 25, 50, 100, and 150, sizeclass 2: employment sizeclass based on deciles, sizeclass 3: employment sizeclass with thresholds 20, 50, 100, 250, and 500, ageclass: establishment age class with thresholds 1, 2, 3, 4, 5, 10, and 15.

Appendix A

Table A1. Summary of Variables Used in Selected Tables and Figures

	Table 1 Correlations	Table 2 Correlations	Table 3 Dispersion	Figure 1 Employment	Figure 2 Hours	Figure 2 Output	Figure 3 Productivity
				Comparisons			
BLS implicit price deflator used for all estimates (except for capital)	yes	yes				yes	yes
Shipments deflator used to deflate output			yes				
Cost of resales (CR) removed from CMP	yes	yes	yes			yes	yes
Employees only			yes	yes		N/A	
Include BLS employees and self-employed (SE) and unpaid family workers (UFW) in BLS data only	yes	yes			yes	N/A	yes
CPS nonproduction/production hours ratio (even for NBER hours)	yes	yes	yes		yes		yes
BLS intrasectorals included	yes	yes				yes	yes

Appendix B

B.1. Properties of ASM samples

The ASM is a 5-year panel of roughly 50,000–70,000 manufacturing establishments. It is a sample of establishments drawn from the manufacturing portion of the Census Bureau's Business Register using a probability proportional to size sampling scheme.[35] The largest establishments are sampled with certainty and are included in every panel.[36] Smaller establishments are sampled with a probability less than 1, where the probability increases with establishment size (measured by shipments). The smallest single-unit establishments, which are part of the "non-mail" stratum, are not mailed a form but they are included in the estimates. Due to the desire to reduce reporting burden, the Census Bureau uses administrative records to impute payroll, employment, industry and location from the administrative data for the smallest single-unit establishments, while total value of shipments are imputed using industry averages.[37]

The ASM sample is refreshed every 5 years. New ASM panels are drawn from the Economic Census and begin 2 years after the Census from which it was drawn (years ending in 4 and 9). The sample is also updated annually to include new establishments which are identified on the Census Bureau's Business Register. The Business Register is updated with information from the Economic Census as well as administrative records from the IRS and the Census Bureau's annual Company Organization Survey.

Data for the ASM are collected in all years except for years ending in 2 and 7, when the

[35] More information about the ASM: http://www.census.gov/manufacturing/asm/.
[36] Prior to 1999, certainty units were establishments with 250 or more employees. In 1999, the cutoff was increased to 500 employees, and in 2004, it was increased again to 1,000 employees. Currently, the 10 largest establishments in an industry are also sampled with certainty. In addition to establishment size, certainty criteria include other characteristics such as industry, cell size, or energy use. For example, Computers, Flat-glass, Sugar, and Small industries (with less than 20 establishments), or establishments with large inventories, assets, fuel/electric expenditures are also sampled with certainty.
[37] Non-mail cases are included in the official estimates and have a weight of one. The survey is designed to tabulate cases from the mail and the non-mail component. The mail component was not designed to estimate the total population.

ASM data are collected as part of the Economic Census. Data on payroll, employment, industry, and geography for establishments in the non-mail stratum are obtained from administrative records.[38]

The ASM sample is designed to estimate unbiased national level estimates of a skewed population. For example, in the 2014 ASM panel, large establishments sampled with certainty account for approximately 72% of the total value of shipments in the 2012 Census of Manufactures; non-certainty establishments are sampled with probabilities from 0.05 to 1.00.[39] This sample design implies that the establishment counts in various size bins may not reflect those calculated from the LBD.

The ASM sample weights, which are inversely proportional to a shipments-based establishment size measure could, in principle, be used to correct for the effects of the ASM sample design. However, the sample design implies that the weighted sum of shipments from the mail stratum only will not match published totals.[40]

Another important aspect of the sample design is that the composition of establishments changes over time and between sample selections. Any weighting procedure aiming at creating unbiased estimates of productivity dispersion should account for the fact that the sampling probabilities, and therefore the composition of the ASM, change every 5 years. In addition, sampling and non-mail stratum thresholds vary across years.

B.2. Establishment Characteristics and the Probability of Selection into the ASM

The ASM's sample design has important implications for our analysis. For example, the sum of the ASM sample weighted employment or sales might equal total employment or total

[38] Federal regulations require the Census Bureau to limit survey response burden.
[39] Source - ASM Methodology website, https://www.census.gov/programs-surveys/asm/technical-documentation/methodology.html (accessed September 16, 2020).
[40] As mentioned above, only the mail component together with the adjustment for the non-mail stratum yields unbiased estimates of the total population. See Davis, Haltiwanger and Schuh (1996) for more details.

sales. However, it is not clear that the ASM sample weights are appropriate for our analysis. This section is devoted to describing our weighting procedure.

To address the effects of the ASM's sample design, we construct propensity score weights using the Longitudinal Business Database (LBD). See Jarmin and Miranda (2002) for more details about the LBD. The propensity score weights are constructed from a logistic regression in which we model the relationship between plant characteristics and the probability that an establishment is selected into the ASM. We start by matching establishments in the ASM to LBD establishments by year and "LBD Number."[41] Our dependent variable is a dummy variable that equals one if the establishment is in both the ASM and the LBD for that year and zero if the establishment is only in the LBD. For establishments in the non-mail stratum, the dummy variable is set to 0.

The set of regressors consists of dummy variables that classify each establishment based on its employment and payroll size class, whether the establishment is part of a multi-unit entity, the establishment's industry code, and the interaction between industry and employment size effects. Including industry-size interactions allows us to estimate industry-specific size distributions. These variables are obvious candidates for our logistic regressions because the probability of selection into the ASM sample and the cutoff for the non-mail stratum in the ASM vary by industry and size.

When determining weights, we define industries at the 3-digit NAICS level because the interaction of size indicators and more narrowly defined industry codes leads to empty cells in smaller industries. Empty-size bins imply that the size distribution cannot be estimated in these industries.[42] When the size distribution cannot be estimated for an industry, propensity scores

[41] The LBD Number is an establishment identifier that is consistently defined across both datasets. Although linking the datasets by LBD Number is straightforward, a small percentage of establishment-year observations do not match due to timing issues between the ASM and the LBD.
[42] The size distribution cannot be estimated if all establishments are in the same size bin.

cannot be calculated because maximum-likelihood estimates of the size effects do not exist. Empty cells can, in principle, be avoided by collapsing size bins, combining similar narrowly defined industries, or allowing bin definitions to vary across industries. We experimented with the number and definition of the size bins and the level of industry aggregation and found that using 3-digit industry codes together with 4 size bins allows us to estimate the size distribution in every industry and year. Allowing for more heterogeneity by using either industry-specific size bins or more narrowly defined industries leads to feasibility problems with the logistic regression.

We defined the size bins so that the resulting distribution allows the lowest size bins to vary over time. That is, in every year and every industry, the 50th percentile of establishments with fewer than 50 employees is used to define bins 1 and 2. For larger establishments, the following bins are defined: 50–99, 100–199, and 200+.[43] There are 21 3-digit NAICS industries in the 2002 classification system, which results in 105 industry specific size distributions. We include a continuous size measure in order to allow the weights to vary within these cells. This is necessary to account for possible within-cell compositional changes. Adding 5 payroll classes and 2 groups related to multi-unit status increases the number of cells to 113.[44]

The 2002 change in the industry classification system resulted in missing NAICS-2002 codes for a nontrivial number of establishments in the LBD between 1997 and 2001. For example, the NAICS code is missing if an establishment exited prior to 2002. For these observations, we used imputed NAICS codes.[45] From 2002 on, NAICS codes are available for all establishments in the LBD.

[43] The payroll size classes are 0–200, 201–500, 501–1000, 1001–5000, 5001+.

[44] If we were to use 4-digit industry, the number of cells would increase significantly. There are 86 4-digit NAICS industries implying 86 different size distributions and 430 industry-size cells. Such an increase in the number of cells yields empty size bins in several industries.

[45] NAICS codes are imputed using a method described in Fort (2013).

Our inverse propensity score weights generate employment counts that do a good job of matching the trends and cyclical variation in BLS manufacturing employment, but they do not match BLS levels.

B.3 Comparison of Hours Measures

In this study, we use hours data from ASM, augmented with data from the CPS. However, for official estimates of productivity growth, BLS uses the CES as its main source of hours data. Although the CES and ASM are establishment surveys, the two surveys differ in which hours data they collect and how they collect it. The best information on these differences comes from studies completed in the early 2000s (Goldenberg and Willimack, 2003; and Fisher et al., 2001). These studies do a nice job of summarizing the differences between the two surveys and how those differences affect estimates of hours worked. In this appendix, we summarize that research and discuss the implications for comparing our estimates to published BLS estimates.

There are some general differences between the two surveys that are worth noting. First, the ASM is an annual survey, whereas the CES is conducted monthly. As a result, the reference periods of the two surveys differ. The reference period for the CES is the pay period that includes the 12^{th} of the month. The CES collects data on the total number of employees, hours for all employees since 2006, the number of production workers, production worker payroll and production worker hours.

In contrast, the ASM has different reference periods for different data elements. For production worker employment, the ASM reference period is the pay period that includes the 12^{th} of the month in the months of March, May, August, and November. These quarterly reports are then averaged into an annual number. The ASM collects employment data for Other Employees only for the pay period that includes March 12^{th}. The implicit assumption is that non-production worker employment does not vary much over the year. Total employment is not collected directly, but rather is equal to the sum of non-production worker employment in March and the annual average of quarterly production worker employment.

Annual total employment in the two surveys can differ if there are seasonal patterns in production worker employment that are missed in the ASM's quarterly reports or if there is a seasonal pattern to non-production worker employment. We examined this issue using monthly CES data. Specifically, we calculated the average employment for each quarter using CES data, and then calculated the ratio of average quarterly employment to CES employment in the ASM reference months (March, May, August, and November). The ratios were very close to 1, indicating that estimates of average annual employment are the same whether we use four quarterly reports or 12 monthly reports.

There are greater differences in the hours data collected in the two surveys. First, the two surveys use different concepts. The ASM asks employers to report hours *worked*, whereas the CES collects hours *paid*. The main difference is that the CES hours data include holidays, annual leave, and sick leave that were paid but not worked. Thus, we would expect total annual hours reported in the ASM to exceed total annual hours in the CES. For productivity measurement, hours worked is the correct concept, which is why BLS adjusts the CES data using hours-worked-to-hours-paid ratios from the NCS.

The two surveys also differ in how they ask respondents to report hours. The ASM asks respondents to report total annual production worker hours. The CES asks respondents to report total production workers for the pay period that includes the 12^{th} of the month. The hours reports are converted to a weekly number using conversion factors that vary with the number of workdays in the month. Apart from the difference in concept, these two approaches to collecting hours data could result in different estimates of total annual hours. Research by Frazis and Stewart (2004) has shown that people work longer hours during the week that includes the 12^{th}

of the month.[46] This would lead to annual hours in the CES being higher than in the ASM and would offset some of the difference due to the difference in concept. Neither survey collects hours data for non-production workers. As noted in the text, nonproduction worker hours are estimated using data from the CPS.

[46] Their research examined the accuracy of CPS hours reports by comparing the CPS hours data to hours data from the American Time Use Survey (ATUS). Subsequent research by Eldridge et al. (2019) found differences in some years, but not in others.