University of Toronto
Department of Economics

Working Paper 654

The Role of Nonemployers in Business Dynamism and Aggregate Productivity

By Pedro Bento and Diego Restuccia

January 31, 2020

The Role of Nonemployers in Business Dynamism and Aggregate Productivity†

Pedro Bento
Texas A&M University*

Diego Restuccia
University of Toronto
and NBER**

January 2020

Abstract

A decline in the net entry rate of employer firms in the United States in the last decades, a decline in business dynamism, may explain the observed productivity slowdown. We consider the role of nonemployers, businesses without paid employees, in business dynamism and aggregate productivity. Despite the decline in the growth of employer firms, the total number of firms has increased since the early 1980s, which in the context of a standard model of firm dynamics implies an average annual growth of aggregate productivity of 0.26-0.39%, over one quarter of the productivity growth in the data.

Keywords: nonemployers, employer firms, business dynamism, productivity, TFP.
JEL codes: O4, O51, E1.

†For helpful comments and suggestions we thank Benjamin Bridgman, John Haltiwanger, Benjamin Pugsley, and audiences at Texas A&M University, the University of Texas at Austin, the 2019 Midwest Macroeconomic Meetings at Michigan State University, the 2019 RCEA Growth, Innovation and Entrepreneurship Conference at Wilfrid Laurier University, and the 2020 Econometric Society North American Winter Meetings in San Diego. All errors are our own. Restuccia gratefully acknowledges the support from the Canada Research Chairs program and the Bank of Canada Fellowship. The views expressed herein are not necessarily those of the Bank of Canada and are the authors alone.

*Department of Economics, Texas A&M University, 254 Liberal Arts Social Sciences Building, 4228 TAMU, College Station, TX 77843, USA. E-mail: pbento@tamu.edu.

**Department of Economics, University of Toronto, 150 St. George Street, Toronto, ON M5S 3G7, Canada. E-mail: diego.restuccia@utoronto.ca.

1 Introduction

A number of studies have documented a slowdown in business startups and entrepreneurship in the United States over the last several decades. The decline since the Great Recession, in particular, has been proposed as a potential source of the slowdown in aggregate productivity growth (Decker et al., 2016; Furman and Orszag, 2018). However, Decker et al. (2016) and Li (2017) have noted that standard measures of business dynamism appear unrelated to estimates of aggregate total factor productivity (TFP) growth before the Great Recession. In this paper, we construct a broader measure of the total number of firms we believe paints a more comprehensive picture of business dynamism in the United States over time.

Canonical theories of firm size and firm dynamics, such as Hopenhayn (1992), have been used to draw implications for TFP from data on business dynamism. In these models, TFP depends on factor inputs and a term that aggregates the productivity of all firms, which also depends on the total number of firms. In this context, measured TFP—the aggregate amount of output per unit of composite aggregate inputs—depends on the total number of firms. We construct a comprehensive measure of U.S. businesses that includes nonemployers, that is businesses that are subject to federal income tax but have no paid employees, composed solely of owner-managers and unpaid workers such as family members. We show that the startup rate for this measure of firms has not declined, instead it has increased substantially. The literature on business dynamism has thus far focused on employer firms, firms with at least one paid employee, abstracting from nonemployer firms. But nonemployers account for 82% of all firms in 2014, suggesting their evolution over time is an important determinant of changes in the total number of firms. We combine employers data from the Business Dynamics Statistics (BDS) with nonemployers data from the Nonemployer Statistics (NES) and other data sources, to construct a measure of the total number of U.S. businesses from 1981 to 2014. We focus on the number of firms per worker, which according to standard theory, is the relevant measure when drawing implications for aggregate productivity (Hopenhayn, 1992; Karahan et al., 2018). Although the number of employers per worker decreased by 4.5% from 1981 to 2014—consistent

with the findings in Karahan et al. (2018)—the total number of firms per worker in our measure increased by 53% over the same period.

We consider a standard model of firm entry in order to quantitatively assess the impact of the surge in the number of firms on aggregate TFP. In the baseline model the distribution of firm-level productivities is constant over time but we allow for changes in aggregate employment as observed in the data. We assume that the cost of entry changes over time in order to match the observed evolution in the number of firms. The calibrated model suggests aggregate TFP grew at an annualized rate of 0.26% from 1981 to 2014 due to the increase in firms, one quarter of the actual growth in measured TFP. In contrast, using the number of employers as is standard in the literature, the model implies slightly negative growth in TFP. Over 33 years, these implied rates compound into a cumulative increase in aggregate TFP of 9% when using the total number of firms per worker and -1% when using the number of employers per worker.

Our baseline model implies an invariant employment size distribution across producers over time, while the data and recent studies of market concentration suggest otherwise (Autor et al., 2019; Rossi-Hansberg et al., 2019). We extend the baseline model to allow for changes in average productivity over time, which might arise from differential exit rates of employers and nonemployers and from changes over time in the dynamics of incumbents. We discipline these features using data on exit rates and changes in the employment-size distribution of entrants and incumbents over time. Under relatively weak structural assumptions commonly made in the firm-dynamics literature, we show how this additional data can be combined with data on the total number of firms to derive the implied change in the average productivity of firms over time up to a constant, that is, relative to any change in productivity common to all firms. We find that implied average firm-level productivity has increased from the 1980s to 2014, making the implied cumulative increase in aggregate TFP in the extended model 13.3% compared to 9% in the baseline model. TFP growth implied by the extended model correlates well with observed TFP growth over the medium and longer run relative to a model that includes only employer firms.

We provide a comprehensive measure of business dynamism that includes nonemployers, complementing the important work of Decker et al. (2014), who emphasize the decline in net entry rates for employer firms. We show that the net entry rate of all firms has not declined. Karahan et al. (2018) and Hopenhayn et al. (2019) document a marginally declining trend in employers per worker and conclude that changing business dynamism has not been a quantitatively significant driver of TFP trends. We show instead that when considering nonemployers in the total number of firms, business dynamism has contributed substantially to positive productivity growth.

In the next section, we discuss the evidence on nonemployers and why they may matter for business dynamism. Section 3 describes the data for employers and nonemployers and documents trends in the variables of interest. In Section 4 we present our baseline model of firm entry to assess the quantitative impact of firms per worker on aggregate productivity. Section 5 extends the analysis to include differential firm exit rates and firm-level productivity growth showing that implied TFP growth from firm dynamics is even larger than in the baseline model. Section 6 provides a discussion of potential alternative drivers of business dynamism. We conclude in section 7.

2 Nonemployer U.S. Businesses

We construct a comprehensive measure of the total number of firms in the U.S. economy to assess the role of changes in net entry on aggregate productivity. We focus on a measure of the total number of firms that includes nonemployer businesses. Nonemployer businesses are firms with no paid employees, including self-employed entrepreneurs. A comprehensive measure of firms may be relevant in understanding changes in net entry rates over time, as is the case when considering very small firms in the context of cross-country differences in establishment size (Bento and Restuccia, 2017, 2020).

For many economic questions it is reasonable to abstract from nonemployers, as they contribute

little to aggregate output in the U.S. economy. Although nonemployers constitute 82% of all U.S. businesses in 2014, they represent only about 4% of total revenues. However, theories of firm size and firm dynamics suggest patterns of firm entry and exit are essential for aggregate productivity implications. In this context, it is important for the analysis to account for all firms. This is the case even if nonemployers are less productive than employer firms and account for a small proportion of output and employment, although these characteristics need to be taken into account.

Including nonemployers in the total measure of firms raises important questions. Are nonemployer firms using different technologies than employer firms or operating in different product markets? Or are nonemployers the same as employer firms albeit with lower productivity? Our data, together with recent papers by Acs et al. (2009), Davis et al. (2009), and Fairlie et al. (2018), provide us with a characterization of U.S. nonemployers that can be compared with employer firms. In the data, nonemployers coexist and compete with employers within narrow industries. In each of the nine industries we consider, nonemployers represent more than 55% of all firms in 2007 (more than 80% in five industries) and are more prevalent in industries with smaller average employment firm size. The survival rate of nonemployer startups is close to that of employer startups. Data on employment in nonemployers (i.e., owner-managers and unpaid workers) are not available, but average growth rates of revenue are similar to that of small employer firms. A small percentage of nonemployers transition into employer status each year, roughly consistent with employment growth rates among small employers. Considering differences in average revenue across different firm-size classes, nonemployers do not appear distinct from small employer firms in terms of relative size. For instance, comparing employer firms with 10 to 20 employees to employer firms with 1 to 5 employees—a 5-fold difference in average employment between these two groups—we find that there is a 6-fold difference in average revenues. Comparing employer firms with 1 to 5 employees to nonemployers, we find a close 5-fold factor difference in average revenues. The main difference between employer and nonemployer firms other than size appears to be their probability of exit, which is consistent

with a declining exit rate with firm size. While about 8 to 9% of employers exit each year, nonemployers exit at a higher rate of 15% (Davis et al., 2009).

From these facts we conclude it is reasonable to treat nonemployer businesses as employers but operating at a lower scale—possibly because of lower productivity—with higher exit rates.

3 Data

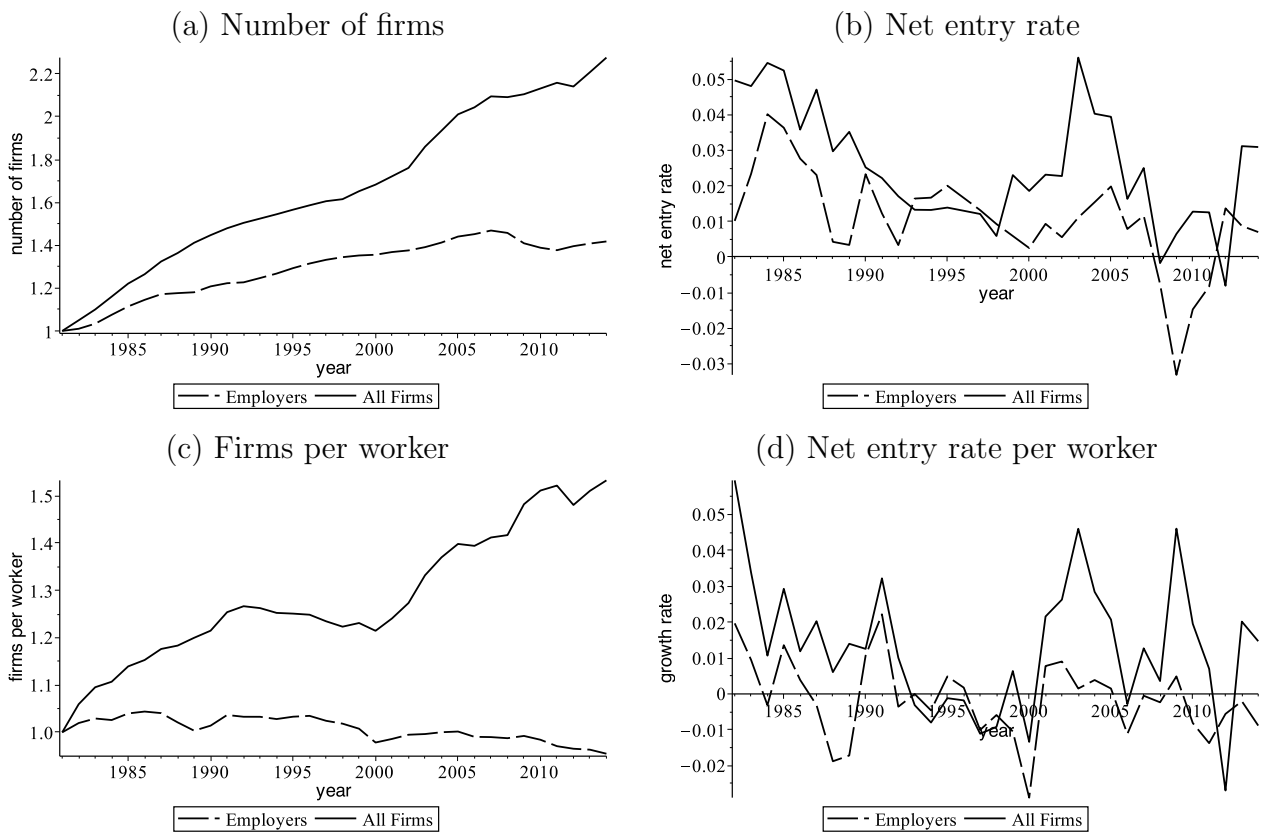
We describe the data and procedure used to construct our measure of the total number of firms over time in the U.S. economy. Data for employer firms is from the U.S. Census Bureau's Business Dynamics Statistics (BDS), the standard data source in the business dynamism literature (Decker et al., 2014). The employer data contains employer-firm counts by industry, employment size, and age from 1977 to 2014. All non-farm firms with at least one formal employee are included.

Data for nonemployers is from the U.S. Census Bureau's Nonemployer Statistics (NES). NES contains economic data for businesses that have no paid employees and are subject to federal income tax, providing nonemployer business counts by industry for 1992 and from 1997 onward. The U.S. Internal Revenue Service (IRS) tax return data is used by the Census Bureau to identify the universe of potential nonemployers. IRS counts up to 2008 are reported in U.S. Statistical Abstracts. Care is then taken to identify duplicates (multiple tax numbers belonging to one firm), and reclassify nonemployers when they are properly part of an employer firm.

To construct our measure of the total number of firms, we simply add nonemployer businesses to employer firms. This is done for the years 1992 and 1997 to 2014 for which we have data for both nonemployers and employer firms. We impute nonemployer counts for the years 1981 to 1991 and 1993 to 1996. For the years 1993 to 1996, we simply assume that the number of nonemployers increased smoothly from 1992 to 1997, and add the implied number of nonemployers to the observed number of employers. For the years 1981 to 1991, we impute the number of nonemployers using IRS data. We describe this imputation in Appendix A.

Figure 1 documents the evolution of the number of firms and firms per worker in the U.S. Panel (a) reports our measure of the number of firms and the more common measure of the number of employers over time, normalized to one in 1981. Panel (b) reports the net entry rate (growth in the number of firms) of all firms and employers. Two features of the data stand out. First, the net entry rate of all firms has been consistently higher than that of employer firms. Second, the net entry rate of all firms declined along with that of employer firms from the early 1980s, but then diverged sharply starting in the late 1990s. From 1981 to 2014, while the number of employers increased by 42%, the total number of firms increased by a striking 128%.

Figure 1: Evolution of Firms and Firms per Worker



Notes: Panel (a) reports the number of all firms and the number of employer firms with levels normalized to one in 1981. Panel (b) reports the net entry rate of all firms and of employer firms. Panel (c) reports the number of firms per worker and panel (d) the net entry rate per worker.

Theories of firm dynamics suggest the more relevant measure of business dynamism when drawing implications for TFP is the number of firms per worker. Using data on the total employed civilian non-institutional population from the U.S. Bureau of Labor Statistics' Current

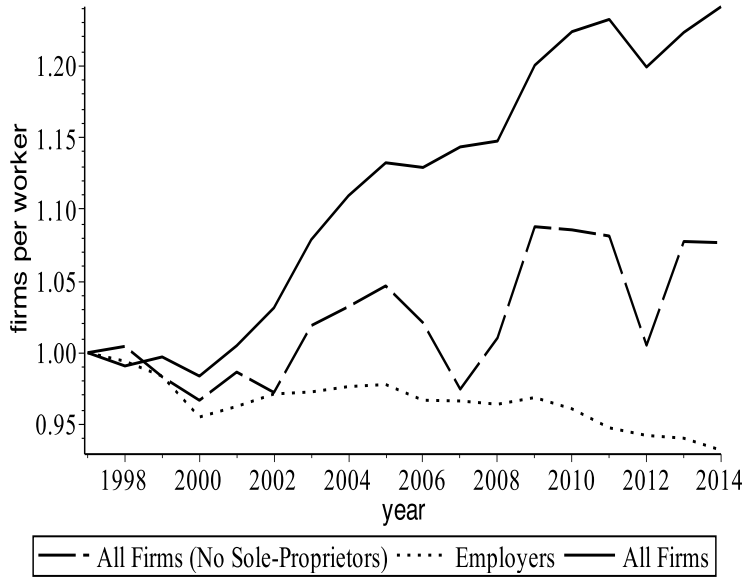
Population Survey (CPS), Figure 1 panel (c) documents the number of firms per worker for all firms and for employers, and panel (d) the net entry rate per worker. Both the total number of firms per worker and the number of employers per worker drop during the 1990s. But whereas the growth rate of employer firms per worker stays negative (on average) after 2000, the total number of firms per worker recovers and grows at a positive rate. From 1981 to 2014, the number of firms per worker increases by 53%, whereas the number of employers per worker decreases by 4.5%.[1]

The striking difference in the number of firms over time between all firms and employers is robust to removing sole-proprietors from nonemployer counts. Data on the legal form of nonemployers is available from 1997 onwards. Figure 2 compares the number of all firms per worker, the number of firms without nonemployer sole-proprietors, and the number of employers. Although the cumulative increase in the number of firms per worker since 1997 is lower when removing sole-proprietors, an 8% increase rather than a 24% increase with all firms, it is still markedly higher than the −7% for employers per worker. We also note the increase in the number of firms per worker over time occurs within sectors and is not the result of reallocation across sectors with different firms per worker. Appendix B documents that firms per worker have increased in seven out of nine sectors and that most employment reallocation has occurred between manufacturing and other services with similar increases in firms per worker. More formally, Appendix B shows via a counterfactual that only 20% of the increase in the number of firms per worker can be accounted for by structural change.

In summary, accounting for nonemployer businesses in firm counts dramatically changes the pattern of net entry over time in the U.S. economy. While the number of employers per worker has fallen slightly over the last three decades, the number of all firms per worker has risen substantially by more than 50%.

[1]The measure of CPS employment we use is fairly consistent with the measures of labor force participation analyzed in Karahan et al. (2018) and Hopenhayn et al. (2019).

Figure 2: Number of Firms per Worker with and without Sole-Proprietors



Notes: The solid line represents the total number of firms per worker in the data, the doted line is employer firms per worker, and the dashed line is the total number of firms per worker excluding sole-proprietors. In each case, levels are normalized to 1 in 1997.

4 Baseline Model

We consider a version of the firm dynamics model in Hopenhayn (1992) in order to provide a mapping from changes in net entry of firms to aggregate TFP. We also use the model to assess the factors leading to the divergence overt time between the number of employer and nonemployer firms.

4.1 Environment

At each date, a single homogeneous good (the numéraire) is produced by firms. Firms have access to a decreasing returns to scale technology in variable inputs and are heterogeneous with respect to their productivity z:

$$y = (Az)^{1-\alpha}\ell^{\alpha}, \tag{1}$$

where y is output, ℓ is the labor input, and A an exogenous productivity term common to all firms that can change over time. Decreasing returns to scale in variable inputs implies $\alpha \in (0, 1)$, hence the optimal scale of a firm depends non-trivially on productivity. More productive firms operate at a larger scale by hiring more inputs, producing more output, and generating higher profits. Firms take the current real wage w as given, and the only cost incurred by incumbents is their wage bill.

There are a large number of potential entrants that can become producers by incurring an entry cost equal to $c_e \cdot Y/L$, where Y/L is aggregate output per worker.[2] We allow c_e to change over time. We assume potential entrants draw their productivity z from some constant cumulative distribution function $G(z)$, and learn their productivity after entry. There is no fixed operating cost for producers, and firm-level productivity is assumed to stay fixed over the lifetime of a firm. We assume all firms face an exogenous probability of exit λ in each period after production.

We denote employment by L which can change over time. We assume that firms always believe current levels of L, c_e, and A persist forever. At the beginning of each period they learn otherwise, but then again believe current levels of these variables persist. We discuss below and in Appendix C that this assumption about beliefs does not affect our main results, it mainly affects the implied entry cost c_e in the model. We abstract from household choices by assuming a constant exogenous real interest rate R.

4.2 Equilibrium

A *competitive equilibrium* is defined by a wage rate w, firm-level functions labor demand $\ell(z)$ and per-period profits $\pi(z)$, and number of firms N, given exogenous entry cost $c_e \cdot Y/L$, labor supply L, real interest rate R, and firm-level productivity distribution $G(z)$, such that:

(*i*) Given w, firms choose $\ell(z)$ to maximize $\pi(z)$.

[2] Having entry costs scale up with output per capita is consistent with the evidence in Bollard et al. (2016) and Bento and Restuccia (2020).

(*ii*) Free entry ensures the expected present value of lifetime profits for an entrant is equal to the entry cost, $\int_z \frac{\pi(z)}{1-\rho} dG(z) = c_e \cdot Y/L$, where $\rho = (1-\lambda)/(1+R)$.

(*iii*) The labor market clears: aggregate labor is equal to the quantity of labor demanded by firms, $L = N \int_z \ell(z) dG(z)$.

The equilibrium in each period can be easily solved. Producers choose labor to maximize operating profits, resulting in the following optimal demand for labor, output, and operating profits, expressed as functions of z;

$$\ell = Az \left(\frac{\alpha}{w}\right)^{\frac{1}{1-\alpha}}, \tag{2}$$

$$y = Az \left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}}, \tag{3}$$

$$\pi = Az \left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}} (1-\alpha). \tag{4}$$

Labor market clearing implies total labor is equal to aggregate labor demand;

$$L = N \cdot A \left(\frac{\alpha}{w}\right)^{\frac{1}{1-\alpha}} \bar{z},$$

where $\bar{z}$ is equal to average firm-level productivity. This is in turn equal to the expected value of each draw z from $G(z)$, or $\int_z z \, dG(z)$. The wage can therefore be expressed as a function of the number of firms per worker;

$$w = \alpha (A\bar{z})^{1-\alpha} \left(\frac{N}{L}\right)^{1-\alpha}. \tag{5}$$

Using equations (3) and (5), aggregate output per worker as a function of firms per worker is;

$$\frac{Y}{L} = \frac{N}{L} \cdot A \left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}} \bar{z} = (A\bar{z})^{1-\alpha} \left(\frac{N}{L}\right)^{1-\alpha}. \tag{6}$$

Free entry ensures N in each period is such that the discounted expected profits of an entrant

are equal to the cost of entry:

$$c_e \cdot \frac{Y}{L} = \frac{A(1-\alpha)\bar{z}}{1-\rho}\left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}}. \tag{7}$$

This free entry condition holds as long as c_e is not too high relative to previous periods. If c_e were to increase too much from period $t-1$ to period t, then the number of firms in period t would be $N_t = (1-\lambda)N_{t-1}$. This constraint never binds when we take this model to the data. Along with equations (5) and (6), the free entry condition implies the following characterization of the number of firms per worker;

$$\frac{N}{L} = \frac{1-\alpha}{c_e(1-\rho)}. \tag{8}$$

Note that firms per worker does not depend on the common productivity term A, on average firm-level productivity $\bar{z}$, or on the size of the workforce L.

Lastly, output per capita, aggregate TFP, and the wage can be expressed as functions of exogenous variables;

$$\frac{Y}{L} = TFP = (A\bar{z})^{1-\alpha}\left(\frac{1-\alpha}{c_e(1-\rho)}\right)^{1-\alpha}, \tag{9}$$

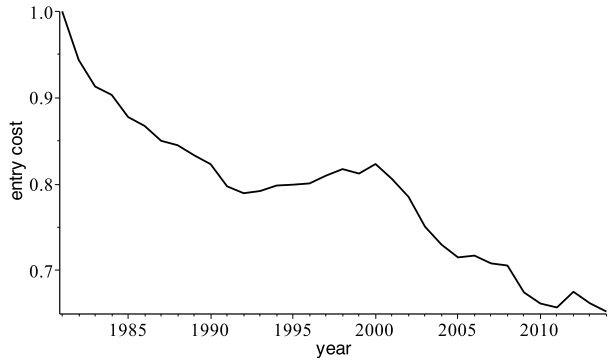
$$w = \alpha \cdot TFP. \tag{10}$$

4.3 Implications

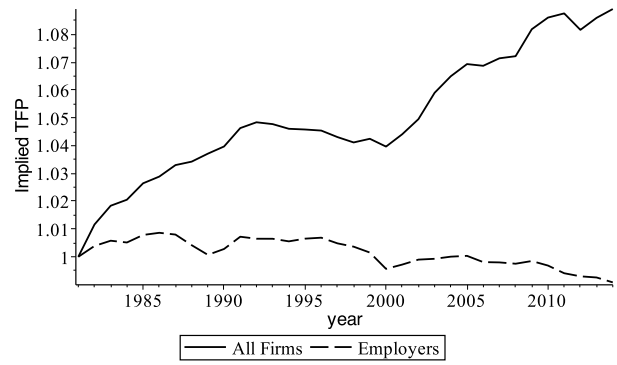
Entry costs. In the model, the number of firms responds solely to changes in employment L and the entry cost c_e, and the number of firms per worker depends only on the entry cost (8). Therefore, we calculate the implied evolution of c_e after 1981 using equation (8). Figure 3a documents the implied evolution of c_e which follows the inverse of the number of firms per worker. We note that the implied entry cost depends on assumptions about the beliefs held by firms with respect to future aggregate employment growth and future entry costs. But regardless of what we assume about firms' expectations, the overall decline in the implied entry cost would still hold. Appendix C shows how entry costs change when firms know the future

Figure 3: Implied Entry Costs and Aggregate TFP in Baseline Model

(a) Entry Cost



(b) Aggregate TFP



Notes: Entry costs c_e is from equation (8) in order to match the total number of firms per worker in the data, normalized to 1 in 1981. Aggregate TFP in each year relative to 1981 is from equation (11) using the total number of firms per worker (solid line) and the number of employer firms per worker (dashed line).

paths of c_e and L.

Aggregate Productivity. In the model aggregate labor productivity and aggregate TFP coincide and are characterized by equation (9) as a function of parameters and exogenous variables such as the entry cost, or by equation (6) as a function of exogenous variables and the number of firms per worker. Using equation (6), we link the effect of changes in the number of firms per worker in the data to changes in aggregate TFP, noting that average productivity $\bar{z}$ drops out since we have assumed that the distribution of productivity is constant:

$$\frac{TFP_t}{TFP_{1981}} = \left(\frac{A_t}{A_{1981}}\right)^{1-\alpha} \left(\frac{N_t/L_t}{N_{1981}/L_{1981}}\right)^{1-\alpha}. \tag{11}$$

To calculate implied TFP over time (relative to 1981), we use a value for α equal to 0.8, consistent with much of the firm-dynamics literature. This relatively high value can be seen as giving a conservative estimate of implied TFP, as lower values (used occasionally in the literature) imply larger effects on TFP from changes in the number of firms per worker. Figure 3b illustrates our main results on the implied growth in aggregate TFP in equation (11) using the total number of firms per worker (solid line), contrasting it with the implied change in TFP in the model if instead we use employers per worker (dashed line) as the measure of the number of firms. Through the lens of the model, the increase in the total number of firms per worker

from 1981 to 2014 implies a 9% cumulative increase in TFP, or an annualized growth rate of 0.26%. This is substantial relative to the observed 0.85% annual growth rate of TFP in the United States over the same time period.[3] In other words, over one quarter of the growth in TFP during the period can be attributed to the change in the number of firms per worker. In contrast, using the number of employers per worker as is standard in the literature, implied TFP decreases by 1% between 1981 and 2014.

Evolution of employer and nonemployer firms. Even with a constant distribution of productivities, the number of employers grows slower, and could even drop, as the total number of firms increases. For intuition, note that a firm's optimal demand for labor in equation (2) is increasing in productivity z and decreasing in the wage. From labor market clearing, the wage is therefore increasing in both the number of firms per worker and average firm-level productivity $\bar{z}$. Combining optimal labor demand in equation (2) with the equilibrium wage (5), we can express a firm's choice of labor as a function of productivity and the number of firms per worker;

$$\ell = \alpha \left(\frac{z}{\bar{z}}\right) \left(\frac{N}{L}\right)^{-1}. \tag{12}$$

If the number of firms per worker increases, demand for labor increases and pushes up the equilibrium wage (equation 5). This results in lower employment for any given level of productivity, as shown in equation (12).

Assume nonemployers are firms with optimal labor demand less than one, with labor provided by an owner-manager. Using equation (12) we can solve for a productivity threshold z_1 above which a firm is counted as an employer;

$$z_1 = \alpha^{-1} \bar{z} \left(\frac{N}{L}\right). \tag{13}$$

The number of employers per worker is the number of firms above the productivity threshold

[3] Our measure of observed TFP growth is utilization-adjusted TFP growth from Fernald (2012), updated to 2014.

z_1, which is increasing in the number of firms. This number can be decomposed as follows:

$$\frac{N^{emp}}{L} = \left(\frac{N}{L}\right) \cdot [1 - G(z_1)]. \tag{14}$$

If the number of firms increases, the number of employers increases less as z_1 rises and hence $G(z_1)$ rises. Depending on the shape of the distribution $G(\cdot)$, the number of employers in principle could even drop.

In Appendix D we quantitatively assess the importance of this mechanism for explaining the decrease over time in the number of employers. After calibrating $G(z)$ to U.S. data, we conclude that this mechanism cannot account for a significant part of the drop in employers over time. This suggests that changes in the productivity distribution across producers, assumed constant in the baseline model, could be quantitatively important. We address this shortcoming of the baseline model in the next section by accounting for changes in the productivity distribution over time.

5 Extended Model

We have derived implications of changes in the number of firms per worker on TFP growth assuming that the distribution of productivity across firms remains constant over time. One implication of this assumption is that the model does not capture the shift over time in the share of firms classified as nonemployers, which could affect our implied TFP growth calculations. Another implication is the distribution of employment across firms remains unchanged. This is inconsistent with the observed shift in employment from small to large firms. For instance, we find the average employment size of all incumbent firms relative to entrants rose from about 4 in the early 1980s to more than 5 after 2010. This shift in employment from small to large firms is consistent with a recent literature documenting broader trends in market concentration of economic activity at the top of the firm distribution (Autor et al., 2019; Rossi-Hansberg et al., 2019).

Through the lens of the model, the shift in employment shares may reflect an increase in the average productivity of incumbents relative to entrants. There are several potential reasons for this trend. For instance, Hopenhayn et al. (2019) note that average exit rates for incumbents have fallen over time. If lower exit rates are due to higher productivity growth among incumbents (rather than lower operating costs), then we expect higher average productivity of incumbents relative to entrants. Aghion et al. (2019) point to a falling cost of firm expansion which benefits high-productivity firms more than low-productivity firms, which leads to higher average employment in incumbent firms relative to entrants.

We address this issue by extending the baseline framework to account for observed changes in the employment size distribution and the share of firms that are employers. To do so, we allow for changes in the exit rate of employers and changes in the productivity distribution of incumbents over time. As before, entrants draw from a constant distribution of productivities $G(z)$ and aggregate labor supply L varies according to data.

In each period, optimal firm-level labor, output, and profits are functions of productivity z and the wage rate as described in equations (2) through (4). From labor market clearing, the equilibrium wage rate is a function of average productivity and the number of firms per worker in equation (5). As a result, aggregate TFP which is equal to aggregate output per worker in our model, is still described by equation (6).

We use equation (6) to derive implications for aggregate TFP over time using 1982 as the benchmark year since we now use data on firm exit:

$$\frac{TFP_t}{TFP_{1982}} = \left(\frac{A_t}{A_{1982}}\right)^{1-\alpha} \left(\frac{\bar{z}_t}{\bar{z}_{1982}}\right)^{1-\alpha} \left(\frac{N_t/L_t}{N_{1982}/L_{1982}}\right)^{1-\alpha}. \tag{15}$$

Changes in TFP can be decomposed into terms representing the evolution of: (a) the common productivity term A, (b) average productivity across firms, and (c) the number of firms per worker. As before, we use data on the total number of firms per worker over time as our measure of N/L. We depart from the baseline model in that we now use additional data to infer the

evolution of average firm-level productivity $\bar{z}$ previously assumed constant. We use data from the BDS and NES on the exit rate of employers over time and the evolution of the average size of all firms relative to the average size of entrants. We now describe the specifics of our approach.

The model provides a mapping from relative productivity to relative size across firms. Denote the average employment size of entrants and all firms by $\bar{\ell}_{ent}$ and $\bar{\ell}_{all}$. Given the mapping from size to productivity (up to a constant) in equation (12), average productivity of all firms relative to entrants in each year is;

$$\frac{\bar{z}_{all,t}}{\bar{z}_{ent,t}} = \frac{\bar{\ell}_{all,t}}{\bar{\ell}_{ent,t}}.$$

Given our assumption that entrants draw z from a constant distribution $G(z)$, we infer average firm-level productivity in each year relative to 1982;

$$\frac{\bar{z}_{all,t}}{\bar{z}_{all,1982}} = \frac{\bar{z}_{ent,t}}{\bar{z}_{ent,1982}} \cdot \left(\frac{\bar{\ell}_{all,t}/\bar{\ell}_{ent,t}}{\bar{\ell}_{all,1982}/\bar{\ell}_{ent,1982}} \right) = \frac{\bar{\ell}_{all,t}/\bar{\ell}_{ent,t}}{\bar{\ell}_{all,1982}/\bar{\ell}_{ent,1982}}. \tag{16}$$

From BDS we know the number of employees for both entrant and incumbent employer firms. To impute the labor input of nonemployers, we assume a uniform distribution of employment between 0 and 1 for all nonemployers (both entrants and incumbents) in each year. Hence, the average size of all firms and entrants in any year is:

$$\bar{\ell}_{all} = \frac{L_{all}}{N_{all}} = \frac{L_{all}^{emp} + 0.5 \cdot N_{all}^{non}}{N_{all}}, \qquad \bar{\ell}_{ent} = \frac{L_{ent}}{N_{ent}} = \frac{L_{ent}^{emp} + 0.5 \cdot N_{ent}^{non}}{N_{ent}}, \tag{17}$$

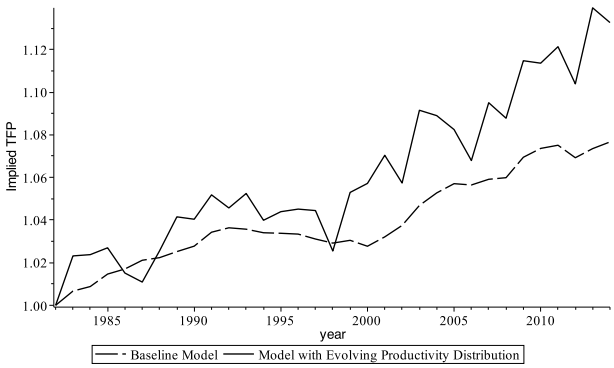
where superscripts distinguish between employers and nonemployers, and subscripts between entrants and all firms.

We have data on the number of employer entrants but not the corresponding measure for nonemployers. We therefore assume a constant exit rate $\lambda = 0.15$ for nonemployers as reported by Davis et al. (2009) and use it to infer the number of nonemployer entrants in each year;

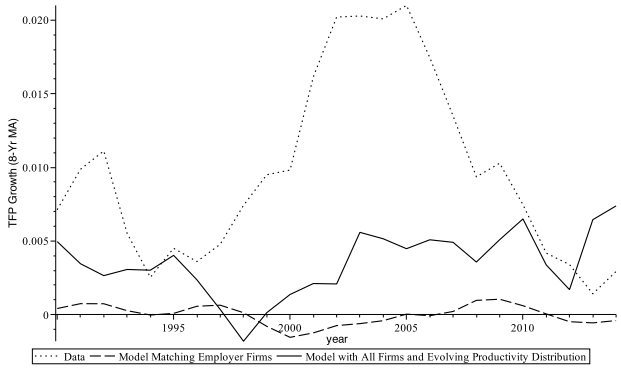
$$N_{ent,t}^{non} = N_{all,t}^{non} - (1 - \lambda) N_{all,t-1}^{non}. \tag{18}$$

Figure 4: Aggregate TFP in Extended Model

(a) Aggregate TFP



(b) Growth in TFP



Notes: TFP growth in the data and implied by the models are reported as 10-year moving averages. TFP data from Fernald (2012), updated to 2014.

Note that we abstract from any transitions from nonemployer to employer status since we cannot differentiate between truly new employers and transitioning nonemployers. As a result, we overstate the number of employer entrants, however, given the small number of transitioning nonemployers in a given year documented by Davis et al. (2009), it should not be quantitatively important.

We calculate average firm-level productivity in each year, relative to 1982, using equations (16)-(18), and use it along with the number of firms per worker to calculate implied TFP from equation (15). Figure 4a reports the results compared with those of the baseline model. Our inferred measure of average firm-level productivity is generally increasing from 1982 to 2014. As a result, the implied cumulative increase in aggregate TFP is even higher than in the baseline model. From 1982 to 2014, the cumulative increase in TFP is 13.3%, compared to 7.7% in the baseline model over these years.

Figure 4b reports the implied growth rate of aggregate TFP over time in the extended model, reported as a 10-year moving average. From 1982 to 2014, implied annual growth averaged 0.39%. The doted line is the 10-year moving average of TFP growth in the U.S. data. The evolution of TFP growth over time implied by the model follows the data well over the medium and long term (correlation 21%), suggesting that business dynamism in fact plays an important quantitative role in driving trends in aggregate productivity, as suggested by canonical theories

of firm dynamics such as Hopenhayn (1992). The cumulative growth in TFP observed in the data between 1982 and 2014 amounts to 32%, meaning our implied measure of TFP can account for more than 40% of the actual increase. The dashed line in Figure 4b shows a corresponding measure of implied TFP growth when employers are used as the number of firms in the baseline model. Not only does this measure imply a decline in aggregate TFP over time, it is also strongly negatively correlated with observed productivity growth (correlation -32%).

6 Alternative Drivers of Firm Dynamics

An important literature links low average firm size (the ratio of workers to the number of firms) to policy distortions in developing countries. We consider two mechanisms highlighted in this literature as potential alternative drivers of the trends in business dynamism: (a) an increase in resource misallocation driven by size-dependent distortions (Hsieh and Klenow, 2014; Bento and Restuccia, 2017, 2020) and (b) increasing barriers for firms entering new markets (Bento, 2019).

If large productive firms are effectively taxed at higher rates than small unproductive firms, then all firms reduce investment in productivity. This effectively reduces non-production costs for all firms, thereby *increasing* profitability and encouraging entry. In equilibrium, these size-dependent distortions result in more but less-productive firms. Appendix E provides evidence that the relationship between firm size and effective tax rates in the U.S. is insignificant and has not grown over time.

If firms operate in multiple markets, Bento (2019) shows that increasing barriers to market entry (distinct from barriers to starting a firm) can result in more firms, each competing in fewer markets. As a result each market is characterized by fewer competing firms, and aggregate TFP drops even as the aggregate number of firms increases. To the extent that firms create multiple establishments to access multiple markets, higher market entry barriers should result in fewer establishments per firm as the number of firms increases. Appendix E shows this has

not been the case for the U.S. over time.

In summary, the evidence suggests that various mechanisms encouraging firm entry while lowering TFP—which are prevalent in developing countries—are not driving the increase in business dynamism in the United States.

7 Conclusions

An important literature documenting a decline in business dynamism in the U.S. over the last several decades has focused solely on employer firms. We consider a broader measure of firms that includes nonemployers, and find that the total number of firms has diverged dramatically from the number of employers over time. We interpret this fact, along with the evolution of the employment distribution across firms, through the lens of a model of firm dynamics based on Hopenhayn (1992). We show that accounting for nonemployers drastically changes the implications for aggregate productivity. Although nonemployers are (by definition) small relative to employers, the increase in the number of firms and in firm-level productivity together imply that business dynamism has been responsible for over one third of observed aggregate productivity growth from 1982 to 2014. This is in striking contrast to the decrease in TFP implied by a model considering only employer firms.

Decker et al. (2016) and Li (2017) note that previous measures of business dynamism (focusing on employer firms) do not correlate well with TFP growth, casting doubt on the quantitative importance of theories of firm dynamics. Our broader measure of business dynamism, which accounts for nonemployers and the evolution of the size distribution over time, follows TFP growth in the data more closely since the 1980s.

Our results suggest several avenues for future research. It would be useful to relate our comprehensive measure of the number of firms with recently documented trends in market concentration, price-cost markups, and job reallocation rates as documented in Decker et al. (2014), De Loecker et al. (2018), and Rossi-Hansberg et al. (2019). Relatedly, theories developed to

explain increasing markups and market concentration, as well as the declining labor share of aggregate income, have taken as given a decline in the number of firms. For instance, Akcigit and Ates (2019) relate these trends to declining business dynamism. As a result, an important direction for future research may be exploring mechanisms that can account for these trends in the context of higher firm entry. Finally, we have abstracted from the underlying causes of changes in exit rates and productivity growth across firms and over time. Understanding these patterns, as explored in Aghion et al. (2019) and Cao et al. (2019), remains an important area for further work.

References

Acs, Z. J., Headd, B., and Agwara, H. (2009). Nonemployer start-up puzzle. Technical report, Small Business Administration, Office of Advocacy.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., and Li, H. (2019). A theory of falling growth and rising rents. Unpublished manuscript.

Akcigit, U. and Ates, S. T. (2019). What happened to u.s. business dynamism? Unpublished manuscript.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2019). The fall of the labor share and the rise of superstar firms. *Quarterly Journal of Economics*, forthcoming.

Bento, P. (2019). Competition, innovation, and the number of firms. Technical report, Texas A&M University.

Bento, P. and Restuccia, D. (2017). Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics*, 9(3):267–303.

Bento, P. and Restuccia, D. (2020). On average establishment size across sectors and countries. *Journal of Monetary Economics*.

Bollard, A., Klenow, P. J., and Li, H. (2016). Entry costs rise with development. Unpublished manuscript, Stanford University.

Cao, D., Hyatt, H. R., Mukoyama, T., and Sager, E. (2019). Firm growth through new establishments. Unpublished manuscript.

Davis, S. J., Haltiwanger, J., Jarmin, R. S., Krizan, C. J., Miranda, J., Nucci, A., and Sandusky, K. (2009). Measuring the dynamics of young and small businesses: Integrating the employer and nonemployer universes. *Producer Dynamics: New Evidence from Micro Data*, pages 329–366.

De Loecker, J., Eeckhout, J., and Unger, G. (2018). The rise of market power and the macroeconomic implications.

Decker, R., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2014). The secular decline of business dynamism in the united states. Unpublished manuscript, University of Maryland.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., Krizan, C. J., and Miranda, J. (2016). Declining business dynamism: Implications for productivity. Technical report.

Fairlie, R. W., Miranda, J., and Zolas, N. (2018). Job creation and survival among entrepreneurs: Evidence from the universe of u.s. startups. Technical report, University of California, Santa Cruz.

Fernald, J. G. (2012). A quarterly, utilization-adjusted series on total factor productivity. *FRBSF Working Paper*, 2012-9.

Furman, J. and Orszag, P. (2018). Slower productivity and higher inequality: Are they related?

Hopenhayn, H., Neira, J., and Singhania, R. (2019). From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share. *National Bureau of Economic Research Working Paper No. 25382*.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60:1127–50.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics*, 124(4):1403–1448.

Hsieh, C.-T. and Klenow, P. J. (2014). The life cycle of plants in india and mexico. *Quarterly Journal of Economics*, 129(3):1035–1084.

Karahan, F., Pugsley, B., and Şahin, A. (2018). Demographic origins of the startup deficit. Unpublished manuscript.

Li, H. (2017). How does business dynamism link to productivity growth? *FRBSF Economic Letter*.

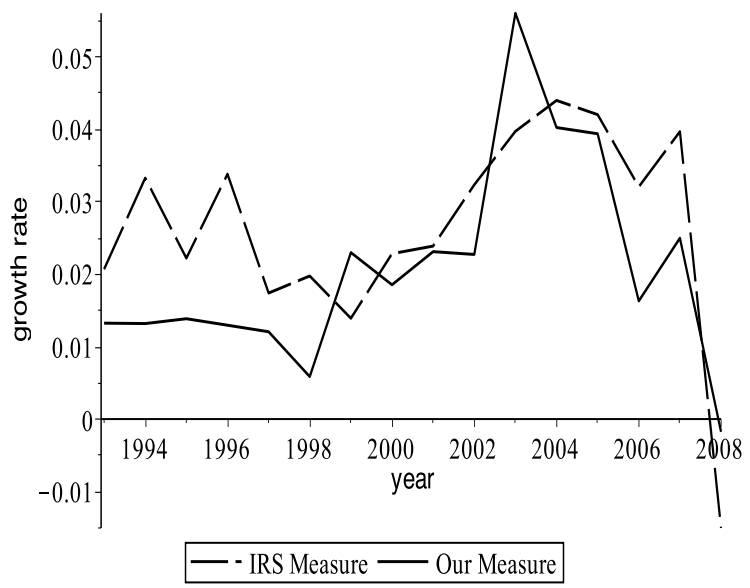
Rossi-Hansberg, E., Sarte, P.-D., and Trachter, N. (2019). Diverging trends in national and local concentration.

On-line Appendix (Not for publication)

A Data Imputation

For the years between 1981 to 1991, we impute the number of nonemployers by using the growth rate in the total number of firms reported by the IRS (constructed using tax returns). We work backwards from 1992, imputing the total number of firms using the growth rate in each year from the IRS data, then subtracting the number of employers (from BDS) to obtain the number of nonemployers. Figure 5 documents the fact that the growth rate in the total number of firms reported by the IRS tracks very closely our measure of the growth rate of the total number of firms over the years for which we have data for nonemployers and employers. Hence, we argue this imputation of the number of nonemployer firms is reasonable.

Figure 5: The Growth Rate in the Number of Firms, IRS and Our Measure



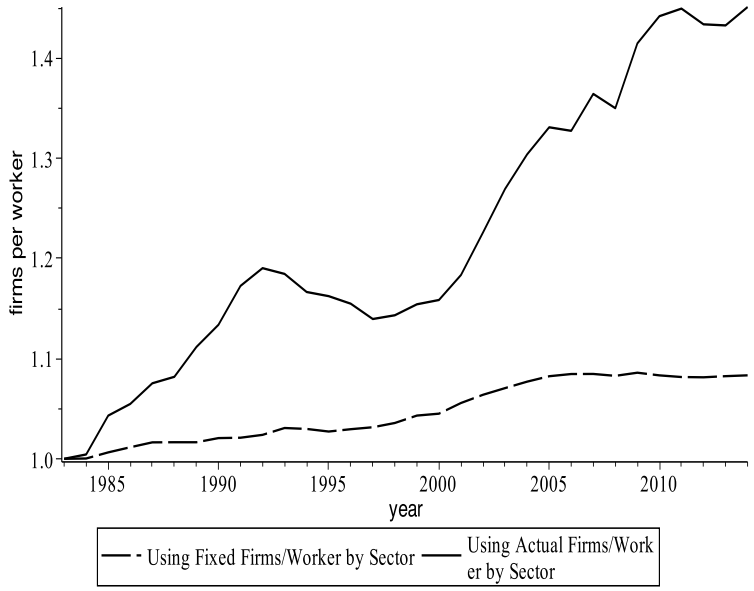
B Sectoral Composition in Number of Firms

Given the structural transformation in the U.S. economy over the last several decades, it is important to assess whether the large increase in the total number of firms per worker is driven by within-sector changes in net entry or by changes in sectoral employment shares, that is changes from sectors with a low number of firms per worker to sectors with a high number of firms per worker. We analyze how sectoral employment shares have evolved over time between 1983 to 2014 for 9 sectors of the economy: agriculture, forestry, and fishing; mining; construction; manufacturing; wholesale trade; retail trade; transportation, communication, and utilities; finance, insurance, and real estate; and other services. We find that the most significant change is the reallocation of employment away from manufacturing to other services. Within these two sectors, firms per worker in manufacturing rose by 47%, while firms per worker in other services rose by a close 46%, which suggests that the process of structural transformation is not driving the increase in the total number of firms per worker. Indeed, firms per worker rose in seven out of nine sectors. The only sectors that experienced a drop in the number of firms per worker are Mining (−52%) and Retail Trade (−16%).

Table 1: The Role of Structural Transformation in Total Firms per Worker

	Employment share (%)		Firms per worker (×100)	
	1983	2014	1983	2014
Sectors				
Agriculture, forestry, and fishing	4	2	9	16
Mining	1	1	25	12
Construction	7	7	21	29
Manufacturing	20	11	3	4
Wholesale trade	4	3	17	20
Retail trade	12	12	21	17
Transportation, communication, and utilities	6	6	11	24
Finance, insurance, and real state	7	7	29	38
Other services	40	52	15	22
Aggregate	100	100	14	20

Figure 6: Total Number of Firms per Worker, Actual vs. Counterfactual



Notes: The solid line represents the evolution of the total number of firms per worker in the data, whereas the dashed line is the counterfactual evolution of the total number of firms per worker when firms per worker in each sector is kept fixed at 1983 levels.

Nevertheless, to get a more concrete quantitative assessment of the importance of structural transformation to the increase in the number of firms per worker, we compute a counterfactual aggregate number of firms per worker assuming that the number of firms per worker in each sector is fixed at 1983 levels. Changes in this counterfactual measure over time are therefore solely driven by changes in sectoral employment shares. Figure 6 reports this counterfactual measure of the aggregate number of firms per worker, along with the actual number of all firms per worker for comparison. The counterfactual shows that only 20% of the increase in the total number of firms per worker can be accounted for by structural change.

C Alternative Beliefs in the Baseline Model

We show that our assumptions about the beliefs of firms in the baseline model affect the implied evolution of the entry cost over time, but do not affect our calculation of the implied evolution

of aggregate productivity. In the baseline model we assume firms always believe the current supply of labor L and entry cost c_e will persist indefinitely, and are subsequently shocked each period. We now assume that firms know the future paths of both L and c_e with certainty. For this exercise we assume that the economy is in a steady state in 1981, such that L and c_e were previously constant at 1981 levels. We further assume these variables stop changing after 2014.

Per-period optimal output, labor demand, and profits, are still described by equations (2) through (4), as functions of firm productivity. The wage is still described by equation (5), as a function of the number of firms per worker. That equation (5) still holds in each period implies that our calculation of implied TFP (11) also holds, given the observed number of firms per worker in each year in the data.

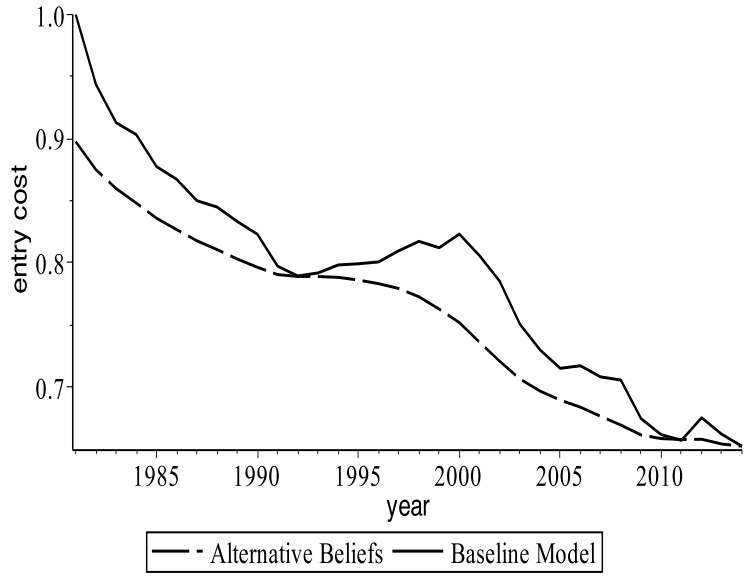
Free entry requires that the expected discounted profits of entrants are exactly equal to the entry cost in the period they enter. This can be expressed recursively as;

$$c_{e,t} = \frac{(1-\alpha)}{(N_t/L_t)} + \rho c_{e,t+1}, \tag{19}$$

$$c_{e,2014} = \frac{(1-\alpha)}{(N_{2014}/L_{2014})(1-\rho)}. \tag{20}$$

Potential entrants now take into account future entry and labor supply growth when making their entry decision. Taking as given the number of firms per worker in the data, the more firms per worker in the future, the lower the current entry cost must be to rationalize a given current number of firms per worker. In Figure 7 we show how the implied entry cost c_e must evolve over time in order to match the observed evolution of the number of firms per worker in the data. Compared to the baseline, this alternative implied entry cost must be lower in 1981, given that firms are now taking into account the future observed increase in the number of firms per worker. By 2014, the two measures converge, as must be the case since the two sets of beliefs also converge. With these alternative beliefs, the implication for the evolution of entry costs is much the same, except here the implied entry cost is less volatile over time.

Figure 7: Implied Entry Cost with Alternative Beliefs



D Share of Employers in Baseline Model

To assess the ability of our baseline model to explain the change in the number of employers over time, given observed changes in the total number of firms, we calibrate the distribution of productivity across firms $G(z)$ to match the employment distribution of firms in 1981. To do this we need data on employment in nonemployers, which is not available. We therefore assume labor is distributed uniformly between 0 and 1 across nonemployers, with owner-managers spending the rest of their time working at other firms. The BDS data reports the number of employer firms falling within 12 different size bins, and the top bin is open-ended. We assume that size is uniformly distributed between the lower and upper bound of each bin, and choose an upper bound for the top bin such that the average size of firms in the top bin implied by a uniform distribution is equal to the average size in the data. Note that equation (2) implies the following productivity ratio between two arbitrary firms i and j with different levels of employment;

$$\frac{z_i}{z_j} = \frac{\ell_i}{\ell_j}. \tag{21}$$

Figure 8: The Number of Employer Firms in the Model and Data

Given the observed employment-size distribution, equation (21) gives the distribution of productivity across firms, up to a constant. Given $G(z)$, we calculate the number of employers generated by the model as the number of firms changes each year (again, we assume the entry cost changes over time generate this outcome). Figure 8a reports the number of employer firms in the model and data and Figure 8b the growth rate in these two time series. The model clearly overstates the number of employer firms. For instance by the end of the sample, the number of employer firms in the model is a factor of 2-fold that in the data. This result implies that the mechanism we focus on, namely the effect of the number of firms on the nonemployer threshold, cannot by itself explain the majority of the divergence between the total number of firms and the number of employers over time. What is striking, however, is that while the growth rate of employers implied by the model is consistently higher than in the data, the growth rates move together quite closely.

E Evidence on Alternative Drivers of Firm Dynamics

Here we assess whether the mechanisms highlighted in a recently literature linking low average firm size to low aggregate TFP in poor countries are relevant for the U.S. experience since the early 1980s. We start by considering whether an increase in the extent of size-dependent distortions may be driving the increase in the number of firms in the U.S. economy over time.

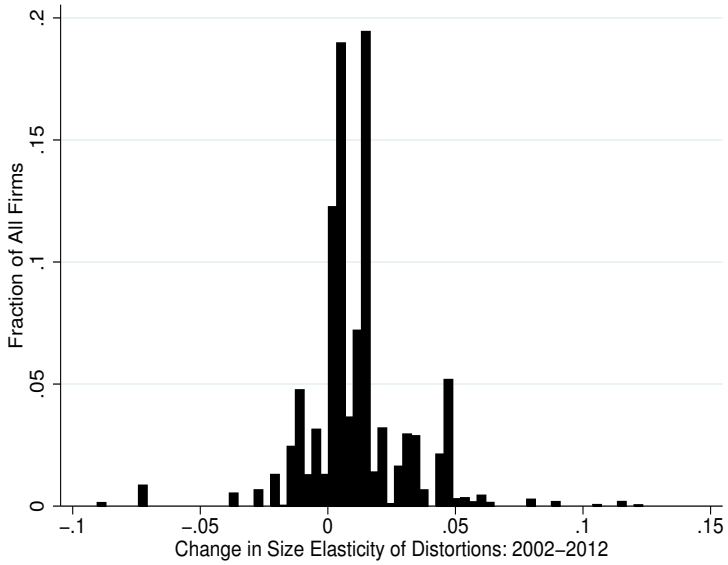
Hsieh and Klenow (2014) and Bento and Restuccia (2017) show that cross-country differences in the extent to which firm-level distortions are positively related to firm size can go a long way to rationalizing the large differences in average firm size across countries at differing levels of development. To assess this basic mechanism in the U.S. data over time, we use data from the Economic Census for 74 3-digit NAICS industries for the years 2002 and 2012. Within each industry, we have data on the total number of firms, the number of firms in each size bin, the number of employees per firm within each bin, total payroll, and total revenue.[4] Although we only have data from 2002 to 2012, this period is still characterized by a substantial increase in the number firms by 22%. For each size bin within each industry, we use the ratio of revenue to payroll as our measure of the average distortion faced by each firm within the bin. Hsieh and Klenow (2009) show that under certain structural assumptions, profit maximization implies each firm within an industry should be choosing its labor input such that all firms are left with the same average product of labor. To the extent that the average product is higher for firms in a large size category relative to a small category, we interpret this as evidence that larger firms face a larger effective tax.

As in Hsieh and Klenow (2014) and Bento and Restuccia (2017), we regress (logged) average products on (logged) employment size within each industry to obtain an estimate of the elasticity of distortions with respect to size.[5] The higher the elasticity, the larger the effective tax rate faced by large firms relative to small firms. For 2002, we find an average elasticity across industries equal to -0.01, with a variance across industries of 0.01. For 2012, we find an average elasticity even closer to zero, with a similar variance. These numbers suggest there is very little systematic misallocation on average, as well as very little change in the extent of systematic misallocation over time. From 2002 to 2012, Figure 9 presents a histogram of the fraction of firms in industries that saw a given change in the size elasticity of distortions, indicating that the vast majority of firms are in industries that saw very little change in this elasticity over

[4]For manufacturing industries, we use establishments rather than firms, and value added rather than revenue.

[5]For these regressions we exclude all firms with less than 5 employees, to address the problem of unpaid workers not being counted in small firms. For the purposes of this discussion, the 'number of firms' is defined as the number of firms with at least 5 employees.

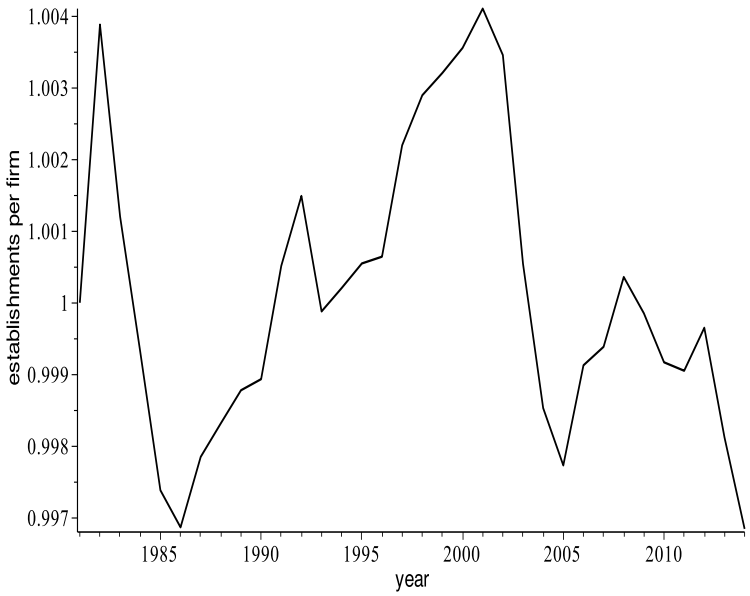
Figure 9: Distribution of Changes in the Size Elasticity of Distortions



time. This suggests that the type of misallocation responsible for cross-country differences in firm dynamics is not driving the increase in the number of firms in the U.S. economy over time.

Another possible explanation for the increase in the number of firms is higher barriers to entering a market. Bento (2019) shows that when firms choose how many markets to enter, and the cost of entering is increasing in the number of markets entered, barriers to market entry (distinct from barriers to starting a firm) encourage more firm startups, with each firm competing in fewer markets in equilibrium. As a result, each market is characterized by fewer competing firms, and aggregate productivity drops even as the aggregate number of firms increases. If barriers to market entry have been increasing in the U.S. economy, we should observe fewer firms competing in each local market, even as the aggregate number of firms increases. We do not have data on the number of firms present in each market, and defining a market is difficult. But we can consider how the aggregate number of establishments changes over time. To the extent that firms create multiple establishments to access multiple markets, the number of establishments per firm can serve as a proxy for the number of markets per firm. Figure 10 reports that the number of establishments per firm essentially remained constant from 1981 to

Figure 10: The Number of Establishments per Firm (1981=1)



2014, even as the number of firms per worker grew by 53%. Note that Aghion et al. (2019) and Cao et al. (2019) make a similar observation about establishments per firm over time in the context of employer firms. This suggests that increasing barriers to market entry are not likely driving the increase in the number of firms per worker in the U.S. economy.