NBER WORKING PAPER SERIES

THE ROLE OF NONEMPLOYERS IN BUSINESS DYNAMISM AND AGGREGATE PRODUCTIVITY

Pedro Bento
Diego Restuccia

Working Paper 25998
http://www.nber.org/papers/w25998

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2019, Revised November 2022

For helpful comments and suggestions we thank the editor, two anonymous referees, Benjamin Bridgman, John Haltiwanger, Pete Klenow, Benjamin Pugsley, and audiences at the 2021 Society for Economic Dynamics Meetings at the University of Minnesota, 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 or the National Bureau of Economic Research and are the authors' alone.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Pedro Bento and Diego Restuccia. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Role of Nonemployers in Business Dynamism and Aggregate Productivity
Pedro Bento and Diego Restuccia
NBER Working Paper No. 25998
June 2019, Revised November 2022
JEL No. E02,E1,O1,O4,O51

ABSTRACT

The well-documented decline in business dynamism, measured by the net entry rate of employer firms, has been proposed as an explanation for the productivity growth slowdown in the United States. We assess the role of nonemployers, firms without paid employees, in business dynamism and aggregate productivity. Including nonemployers, the total number of firms has instead increased since the early 1980s, which in the context of a standard model of firm dynamics implies an annualized growth of measured aggregate productivity of 0.22%, one-quarter of the productivity growth in the data. Further accounting for time changes in the share of nonemployer firms and in the distribution of employment across firms, we find that productivity growth is even higher (0.47% per year). The productivity growth slowdown is not due to changes in net firm entry.

Pedro Bento
Texas A&M University
1101 Hollow Creek Dr. Unit 2208
Austin, TX 78704
pbento@tamu.edu

Diego Restuccia
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON M5S 3G7
and NBER
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 growth slowdown in aggregate productivity (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 growth in aggregate total factor productivity (TFP) before the Great Recession. In this paper, we construct a broader measure of the total number of firms that contributes to a more comprehensive picture of business dynamism in the United States, and use this measure to assess the quantitative contribution of net firm entry on aggregate productivity growth.

Canonical theories of firm size and firm dynamics, such as Hopenhayn (1992), are used to draw implications for aggregate TFP from data on business dynamism. In these models, aggregate output depends on aggregate 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 this measure of the number of firms has increased substantially. Typical measures of business dynamism are based on employer firms, firms with at least one paid employee. 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 show using standard models of firm dynamics to draw TFP implications from firm counts or startup rates, that inferences are substantially biased if only a subset of all firms is counted.

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 1982 to 2014. We focus on the number of firms per worker, which in theory is the relevant measure when drawing implications for aggregate productivity (Hopenhayn, 1992; Karahan et al., 2019). Although the number of employers per worker decreased by −8.7% from 1982 to 2014, consistent with the findings in Karahan et al. (2019), the total number of firms per worker in our measure with nonemployers increased by 41% over the same period.

We consider a standard model of firm dynamics with endogenous entry in order to quantitatively assess the impact of the surge in the number of firms on aggregate TFP. In the baseline model, we assume that the distribution of firm-level productivities is constant over time but we allow for changes in aggregate employment as observed in the data. We infer changes over time in the cost of entry such that the model matches the observed evolution in the number of firms; and study the aggregate productivity implications. The calibrated model implies that aggregate measured TFP associated with the change in the number of firms per worker grew at an annualized rate of 0.22% from 1982 to 2014, 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 (−0.06% per year). Over 32 years, the compound effect in aggregate TFP is an increase of 7.1% when using the total number of firms per worker and a decrease of −1.8% when using the number of employers per worker. These findings illustrate the importance of a comprehensive measure of the total number of businesses for productivity implications. Moreover, the baseline model implies a substantial decline in entry costs which is consistent with some evidence (Koske et al., 2015; Bollard et al., 2016) and robust to changes in other parameter values.

Our baseline model allows us to compare aggregate productivity implications when nonemployers are included in firm counts and when they are not, assuming the distribution of firm-level productivity is constant over time. But while the baseline model matches the increase in the total number of firms, we show that it does not account for the full rise in the share of nonemployers, potentially affecting the productivity implications. It also implies a time invariant

employment size distribution, while the data and recent studies of market concentration suggest otherwise (Autor et al., 2020; Rossi-Hansberg et al., 2021). To accommodate these changes in the data, we extend the model to allow for changes in average productivity over time arising from differential and changing exit rates for employers and nonemployers, and from changes over time in the productivity of incumbents. While the baseline model allows us to draw aggregate productivity implications from the rise in the number of firms over time, keeping the distribution of firm-level productivity constant, the extended model allows us to separably infer the aggregate productivity implications of the observed shift in economic activity across firms over time.

We discipline the extended model using data on differential exit rates between employers and nonemployers, changing exit rates over time, and data on changes in the employment-size distribution of employer entrants and incumbents as well as changes in the revenue share of nonemployers. Under relatively weak structural assumptions commonly made in the firm-dynamics literature, we show how this additional data can be used 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 the implied average firm-level productivity has increased from 1982 to 2014, a TFP effect that is similar in magnitude to the effect of the increase in the number of firms, implying a combined cumulative increase in aggregate TFP of 16.4% compared to 7.1% in the baseline model. Because in our framework the TFP effect of changes in the number of firms is the same in the baseline and extended models, the additional productivity increase is due to a shift in the firm size distribution whereby incumbents have become more productive relative to entrants over time, due to a combination of lower exit rates and higher productivity growth over the life of the average firm. Importantly, our strategy to infer changes in average firm-level productivity in the extended model accounts for the productivity effect of any changes in firm dynamics driving the shift in economic activity from entrants to incumbents over time.

Our more comprehensive measure of business dynamism that includes nonemployers comple-

ments the important work of Decker et al. (2014) and Decker et al. (2016), who emphasize the persistent decline in net entry rates for employer firms. We show that the net entry rate of all firms declined until the mid-1990s, then began increasing. Karahan et al. (2019) 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 important driver of TFP trends. We show instead that when incorporating nonemployers in the total number of firms, net firm entry has contributed substantially to productivity growth. Karahan et al. (2019) and Hopenhayn et al. (2019) also document an increase in the average age of employer firms over time, driven by lower exit rates. We show that accounting for lower exit rates over time increases the implied contribution of net firm entry to aggregate TFP growth, as older firms tend to be much more productive than young firms. Pugsley and Şahin (2019) provide evidence that growth in the numbers of employer and nonemployer firms tend to move together over time between 1997 and 2012, and that these growth rates are correlated with aggregate employment. We extend their analysis by looking at nonemployers over a longer time period, and show that while annual growth rates are correlated, the number of nonemployers grew much faster than the number of employers.

The literature on business dynamism has also documented a drop over time in the job reallocation rate of employer firms, measured by the sum of job creation and job destruction normalized by aggregate employment, as another metric of the decline in business dynamism (Decker et al., 2014, 2016). Hyatt et al. (2020) construct a new measure of job reallocation that accounts for flows to and from nonemployers in addition to employers, and find the decline over time in the job reallocation rate is smaller compared to measures that abstract from nonemployers. Our analysis does not account for changes over time in job reallocation rates or labor flows that might affect the effective productivity of the workforce. We instead focus on the aggregate productivity impact of changes in business dynamism arising from changes in the total number of firms per worker, the share of nonemployers, and the employment size distribution. Nevertheless, to the extent that lower rates of job destruction are driven by lower exit and higher

growth rates for the most productive firms, our model suggests lower destruction rates may be accompanying the higher aggregate productivity driven by the shift in employment towards these firms.

A recent literature has identified policy and institutional distortions in developing countries that encourage more firm entry while distorting the allocation of labor across firms, thereby lowering aggregate TFP (Hsieh and Klenow, 2014; Bento and Restuccia, 2017, 2021; Bento, 2020). We assess whether changes in correlated distortions (distortions that increase with firm size) may be driving trends in business dynamism in the United States using establishment data on employment and revenue; finding that changes in correlated distortions are not substantial in accounting for the U.S. experience in recent decades. Guner et al. (2008) study a particular class of policies, size-dependent policies, which effectively distort the operation of firms after a certain firm-size threshold, thereby encouraging smaller firms and lowering aggregate productivity. Motivated by their work, we consider whether the rise in nonemployers relative to employers could be due to an increase over time in effective distortions faced by firms hiring employees. To do so, we extend the baseline model by imposing a proportional revenue tax to firms demanding more than one unit of labor (employer firms) and infer the evolution over time of this distortion to exactly match the increase in the fraction of firms that are nonemployers. We find that such effective taxes must be very small (less than 2% of firm revenue) and barely change over time. More importantly, the inferred contribution of the increase in the number of firms to aggregate productivity over time remains largely unchanged relative to the baseline model. We conclude that changes in the effective cost of being an employer firm, relative to a nonemployer, are not important in accounting for the U.S. experience.

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 describe 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 differences and changes in exit rates as well as growth in

firm-level productivity, 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 (non-farm) firms in the U.S. economy to assess the role of changes in net firm entry on aggregate productivity. We focus on a measure of the total number of firms that includes nonemployer businesses. Nonemployers are businesses with no paid employees, including self-employed entrepreneurs, that have annual business receipts of $1,000 or more ($1 or more in construction) and are subject to federal income taxes. In 2012, 86% of nonemployers were sole proprietors such as real estate agents and independent contractors, 7% were partnerships, and 7% were corporations. 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, 2021).

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 2012, they represent only about 3% 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 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 small 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. Nonemployers compare to small employers in similar proportions as small employers compare to medium or large employers. For instance, in terms of average revenue per firm in 2007, nonemployers are about 12% of that of small employers (less than 5 employees). Small employers in turn are about 17% of that of medium size employers (10 to 19 employees), which in turn are 16% of that of large employers (50 to 99 employees). Nonemployers certainly do not resemble large employers in terms of average revenue, but the difference between nonemployers and small employers resembles the difference between small and large employers. The main difference between employer and nonemployer firms other than size appears to be their rate of exit, consistent with a decreasing exit rate with size (Haltiwanger et al., 2013). In particular, while about 8 to 9% of all employers exit each year, the exit rate of all nonemployers is higher, about 15% (Davis et al., 2009). The smallest employers (less that 5 employees) exit at a very similar rate of 14%.

We also note that most owners of nonemployer firms also work in other jobs while running their nonemloyer firms. For example in 2017 only about one third of all entrepreneurs identified self-employment as their main occupation according to Current Population Survey data. See also Abraham et al. (2019) for a discussion of the distinction between measures of self-employment inferred from the Current Population Survey and Nonemployer Statistics.

From these facts we conclude that, as a first pass and absent more detailed data, it is reasonable

to assume nonemployer businesses are similar to small employers but operating at a lower scale, possibly because of lower productivity, with higher exit rates. This interpretation is similar to how small employers are treated in the literature, relative to larger employers. A key distinction is that if nonemployers are defined as firms demanding less than one unit of labor, then labor of nonemployers is supplied by owner-maganers who also may supply labor to other firms, while labor of larger firms (employers) is supplied by owners and hired labor. We use these facts and interpretation to develop a model of employer and nonemployer firms in Section 4.

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.

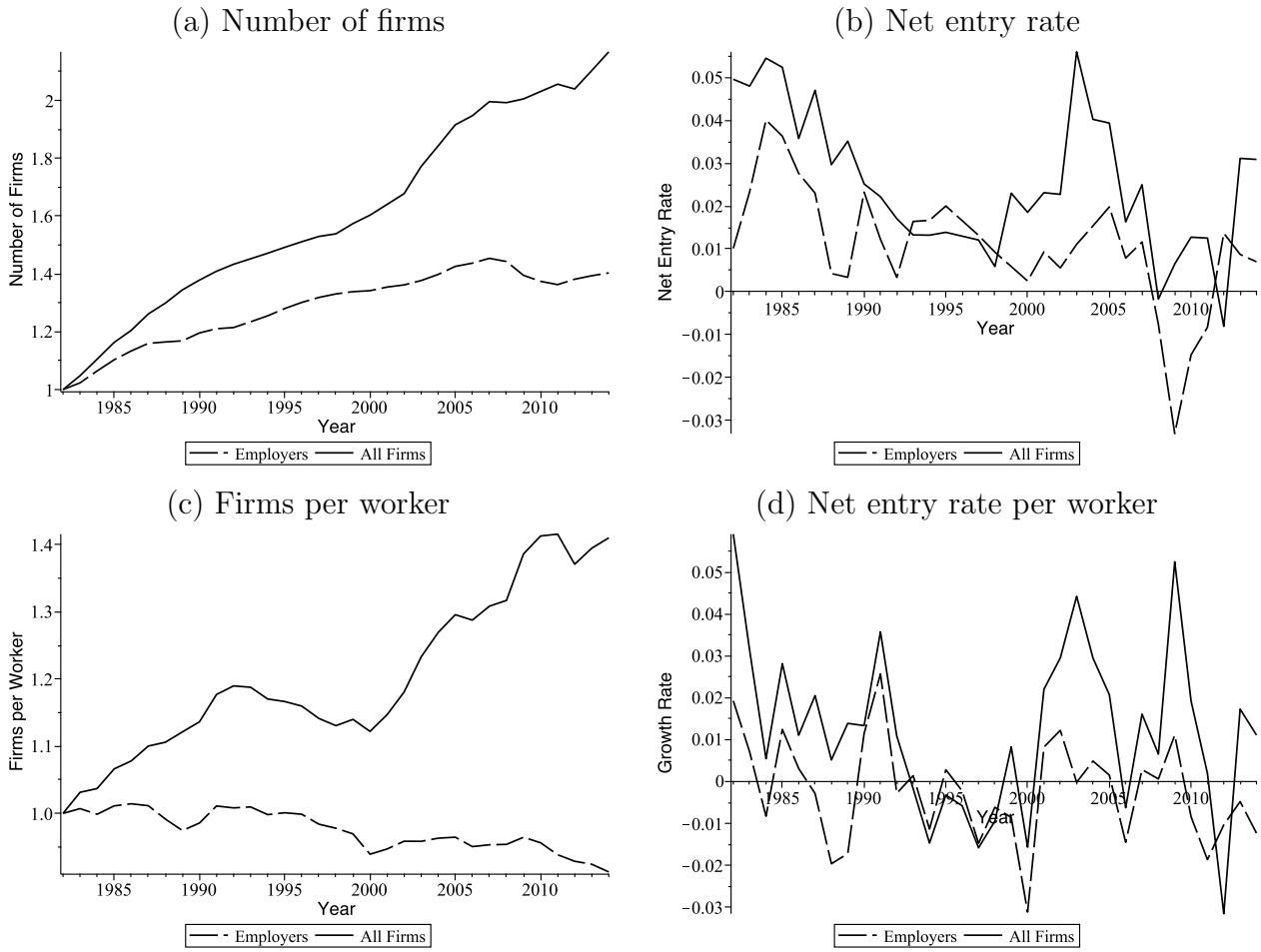
To construct our measure of the total number of firms, we simply add nonemployer businesses to employer firms. This is done for the years 1992 and 1997 to 2014, for which we have data for both nonemployers and employer firms. We impute nonemployer counts for the years 1981 to 1991 and 1993 to 1996. For 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 United States. 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 1982. 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 1982 to 2014, while the number of employers increased by 40%, the total number of firms increased by a striking 117%. Note that although we have data for 1981, our analysis requires changes in variables, as a result, for consistency we focus on the period from 1982 to 2014.

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 (minus government workers) 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. The CPS data includes farm employment, while our firm data excludes farms. We address this inconsistency in Appendix D, where we show our results are unaffected if we adjust aggregate employment to remove farm workers. 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 1982 to 2014, the number of firms per worker increases by 41%, whereas the number of employers per worker decreases by 8.7%. Our measure of CPS employment is consistent with the measures of labor force participation analyzed in Karahan et al. (2019) and Hopenhayn et al. (2019).

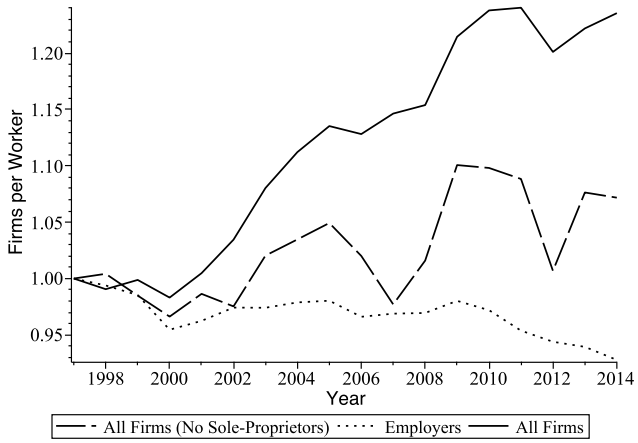
Figure 1: Evolution of Firms and Firms per Worker



Notes: Panel (a) reports the number of all firms and the number of employer firms with levels normalized to one in 1982. 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.

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. As a fraction of total nonemployers, the share of sole-proprietors is 87.2% in 1997 and 86.4% in 2014. Clearly, the number of partnerships and incorporated nonemployers increased even more than did all nonemployers. Figure 2 compares the number of all firms per worker, the number of firms without counting 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, a 7% increase rather than a 24% increase with all firms, it is still markedly higher than the −7% for employers per worker. We also note

Figure 2: Firms per Worker with and without Sole-Proprietor Nonemployers



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

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. We document in Appendix B 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, we show via a counterfactual that only 20% of the increase in the number of firms per worker can be accounted for by the change in economic structure during the period.

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 somewhat over the last three decades, the number of all firms per worker has risen substantially by 41%.

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 over 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. We assume that entry costs scale up with output per capita, consistent with the evidence in Bollard et al. (2016) and Bento and Restuccia (2021). We allow c_E to change over time in order to match the evolution of the net entry rate of firms as we explain below. We assume potential entrants draw their productivity z from some constant cumulative distribution function $G(z)$, and learn their productivity after entry. We assume firm-level productivity stays fixed over the lifetime of a firm and all firms face an exogenous probability of exit λ in each period after production. We also assume there is no fixed overhead cost for producers but in Appendix C we show that incorporating overhead costs does not change the main implications of the model.

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 elaborate in Appendix E that this assumption about beliefs does not affect our main results other than the implied value of entry costs. 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 $\pi(z)$.

(*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 $\rho = (1-\lambda)/(1+R)$ is the effective discount factor.

(*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 average firm-level idiosyncratic productivity, 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 (N/L);

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

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:

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

Along with equations (5) and (6), the free entry condition implies the following characterization of the number of firms per worker;

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

We emphasize that in our framework, the number of firms per worker in equation (8) does not depend on the common productivity term A, on average firm-level idiosyncratic productivity $\bar{z}$, or on the size of the workforce L.

Output per capita and the wage can be expressed as functions of exogenous variables;

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

$$w = \alpha \cdot \frac{Y}{L}. \tag{10}$$

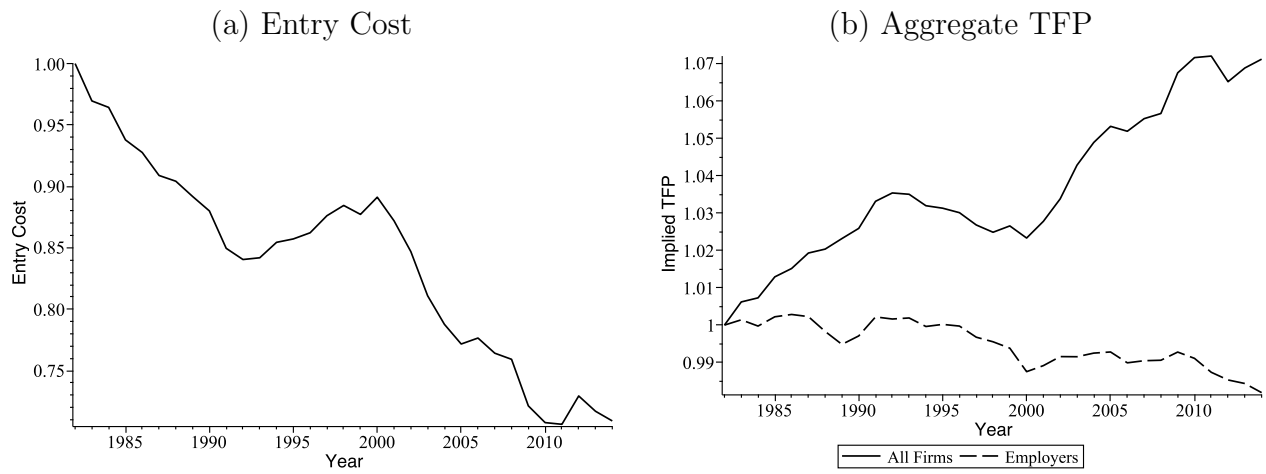
We define aggregate TFP in the model analogous to how it is measured in the data, aggregate output per unit of aggregate composite variable inputs, which is a constant-returns-to-scale function of aggregate variable inputs. As such, we refer to aggregate TFP in the model as measured TFP. We do this in order to make direct comparisons of how changes in firm net entry contribute to changes in measured TFP in the U.S. data. Given that labor is the only variable input in the model, aggregate TFP in the model is the same as aggregate output per worker Y/L.

We note that while TFP depends on the common productivity term A, the number of firms per worker does not. This implies that in our quantitative analysis, we can treat A as a free parameter that captures all changes in measured TFP in the data that are not accounted for by the endogenous forces in the model such as the change in the number of firms per worker. As a result, we report below the change in TFP associated with the change in the number of firms per worker.

4.3 Implications

Entry costs. In the model, the number of firms depends on employment L and the entry cost c_E (as well as α, R, and λ which we assume constant), whereas the change in the number of firms per worker depends only on the change in entry costs as indicated in equation (8). Hence, given the employment data, we infer changes in entry costs to match changes in the number of firms per worker in the data, normalizing the initial entry cost to one. Figure 3a documents the evolution of entry costs c_E which follows the inverse of the number of firms per worker. We note that the implied entry costs depend on the assumptions about beliefs held by firms with respect to future aggregate employment growth and future entry costs. The overall decline in entry costs is robust to alternative assumptions about firms' expectations, as we show in Appendix E when firms know the future paths of c_E and L.

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



Notes: Entry costs c_E calibrated to match the changes in total number of firms per worker in the data, normalized to 1 in 1982. Aggregate TFP in each year relative to 1982 is from equation (11) using the total number of firms per worker (solid line) and the number of employer firms per worker (dashed line).

The implied change in entry costs also depends on our assumption that the labor share α, the real interest rate R, and the exit rate λ are constant over time. Allowing any of these parameters to change could therefore lead to a different path for entry costs. We note, however, that given observed changes in firms per worker, any changes in R and λ would not affect the conclusions about the impact on aggregate productivity in our framework. We start by considering potential changes in the labor share α. Equation (8) shows that a decrease in α increases the number of firms, essentially by increasing firm profit margins. Note that such a decrease in α would lower each firm's labor expenditure relative to revenue. But Kehrig and Vincent (2021) report a small *increase* in this ratio over the time period we consider, almost 2% higher. This suggests that changes in α are not driving the dramatic rise in the number of firms per worker we document.

Next, we consider potential changes in the real interest rate R and in the exit rate λ. There is clear evidence that real interest rates and firm exit rates (at least for employers) have both declined over time. Equation (8) shows that these trends should have contributed to the rise in the number of firms per worker. To assess their quantitative effect on the implied changes in entry costs, we recalculate entry costs for 1982 and 2014 using different values for R and λ. For

R we use 0.058 for 1982 and 0.005 for 2014. These are averages over 1982-1986 and 2010-2014 using annual 10-year real interest rates from Federal Reserve Bank of Cleveland (2022). For λ, given our lack of data for nonemployer exit rates in years other than 1997, we assume a survival rate of 0.86 across all firms in 1997, and then assume this rate changes proportionately with employer survival rates from the BDS data. We then average these survival rates over 1982-1986 and 2010-2014. We obtain exit rates of 0.148 for 1982 and 0.128 for 2014. Using these changes for R and λ, we find a decline in entry costs of 28% from 1982 to 2014, almost identical to the 29% decline in entry costs illustrated in Figure 3a.

Is a decline of almost 30% in entry costs plausible as a source of the increase in the number of firms per worker? Keeping in mind the lack of direct measures for the costs of starting and operating a firm, we discuss two pieces of supporting evidence: a direct qualitative measure and an indirect quantitative measure. First, the OECD (Organisation for Economic Co-operation and Development) constructs an index of how restrictive regulatory barriers to entrepreneurship are for a number of countries, using objective data on *de jure* laws and regulations (Koske et al., 2015). They report a consistent decline in barriers from 1998 to 2013 in the United States, consistent with Figure 3a. Second, Bollard et al. (2016) use a framework similar to ours to infer how entry costs change over time. Acknowledging lack of data on nonemployers, they infer changes in entry costs using firm-level data on value-added over time. They note that the present discounted value of lifetime value-added for entrants should be proportional to entry costs, as in our framework. The analysis suggests that total entry costs have decreased by about 50% from 1982 to 2012 (from Figure 1 in Bollard et al., 2016), relative to aggregate output per worker. This is even larger than the 29% decline we find in our analysis.

Aggregate productivity. Aggregate labor productivity and aggregate TFP coincide in the model 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. As a result, we connect 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

assume a constant distribution of productivity:

$$\frac{TFP_t}{TFP_{1982}} = \left(\frac{A_t}{A_{1982}}\right)^{1-\alpha} \left(\frac{N_t/L_t}{N_{1982}/L_{1982}}\right)^{1-\alpha}. \tag{11}$$

To calculate the implied change in aggregate TFP (from 1982 to 2014) associated with the change in the number of firms per worker, we use a value for α equal to 0.8, consistent with much of the firm-dynamics literature. Note that $\alpha = 0.8$ implies a conservative estimate of TFP growth, as the lower values occasionally used in the literature imply larger effects on TFP growth from changes in the number of firms per worker. Our focus is on the implications for TFP from the change in firms per worker, so we abstract from changes in A over time that would force the variation in TFP in the model exactly match the data. Figure 3b illustrates the main result on the implied TFP in equation (11) using the total number of firms per worker (solid line), contrasting it with the number of employers per worker (dashed line) as an alternative measure of the number of firms. The increase in the total number of firms per worker from 1982 to 2014 implies a 7.1% cumulative increase in TFP over the 32 year period, or an annualized growth rate of 0.22%. This is substantial relative to the data. We consider a measure of TFP in the data from Fernald (2012), which adjusts for capacity utilization and is updated to 2014, to calculate an annual growth of 0.87% in measured TFP in the United States between 1982 and 2014. The implication is that about one quarter of the growth in TFP during the period can be attributed to the substantial change in the number of firms per worker, whereas using the number of employers per worker as is standard in the literature, the productivity effect would be a decrease of −0.06% per year during the period.

The number of nonemployer firms. To derive the implications of the model for the share of employer and nonemployer firms, we assume that nonemployers are firms with optimal labor demand less than one unit of labor, with labor supplied by an owner-manager. This implies that owners of nonemployer firms use any remaining labor endowment to supply labor to employer firms and earning a competitive wage, consistent with our discussion of the self-employment

data in Section 2. We derive the demand for labor by firms by combining the firm's optimal demand for labor (equation 2), which is increasing in productivity z and decreasing in the wage, with the market clearing wage rate (equation 5), which is increasing in the number of firms per worker and average firm-level productivity $\bar{z}$;

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

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. Using equation (12) we can solve for a productivity threshold z_1 below which a firm is counted as a nonemployer;

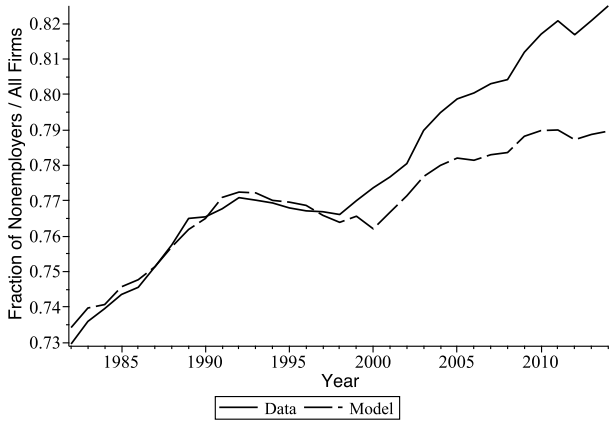
$$z_1 = \bar{z}\left(\frac{N}{L}\right). \tag{13}$$

The number of nonemployers per worker is the number of firms below the productivity threshold z_1, which is increasing in the number of firms per worker. If the number of firms per worker increases, the fraction of firms that are nonemployers $G(z_1)$ increases as z_1 rises. We calibrate $G(z)$ to U.S. data, targeting BDS data on the firm size distribution for employers and data on the revenue share of nonemployers in 1987. We provide more details in Appendix F.

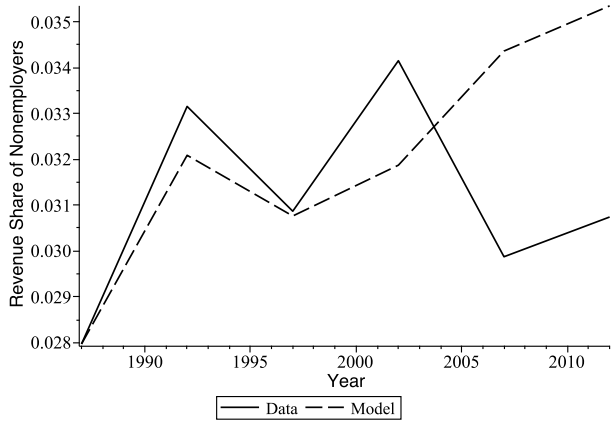
Figure 4a compares the fraction of firms classified as nonemployers in the model to the data. The fraction of nonemployers in the model tracks the data closely until the late 1990's as the fraction of nonemployers grows from 74% in 1982 to 77% in the 1990s. After 1998 the fraction of nonemployers increases more in the data than in the model—the data shows an increase in the share of nonemployers to 82% by 2014, while the model implies a smaller increase to 79%. Figure 4b illustrates the same comparison between the model and the data for the revenue share of nonemployers. Note that revenue data for nonemployers and employers is only available during census years starting in 1987. The model implies a relatively steady increase in the revenue share of nonemployers from 2.8% to 3.6%. Although the model-implied share follows the data relatively closely up to 2002, it then diverges from the data through 2012.

Figure 4: Nonemployer Firms in the Model and Data

(a) Fraction of Nonemployers to All Firms



(b) Revenue Share of Nonemployers



To summarize, the increase in firms per worker in the data can account for much of the change over time in the share of nonemployer firms, but starting in the late 1990s other forces may be contributing to changes in the firm-size distribution, for instance changes in exit rates or in the productivity distribution which are assumed constant in the model. We assess the importance of changes in the firm-size distribution for productivity implications in the next section.

5 Extended Model

The baseline model does not capture the entire shift in the share of nonemployer firms observed over time, from 74% in 1982 to 82% in 2014, potentially affecting the productivity implications. If the rise in the share of nonemployer firms implies a shift of employment to nonemployers, then average firm-level idiosyncratic productivity could fall as a result. This potential effect is abstracted from in the baseline model by assuming constant productivity and employment distributions.

Empirically, recent studies highlight a shift in employment shares from small to large firms, consistent with a recent literature documenting broader trends in market concentration of economic activity at the top of the firm distribution (Autor et al., 2020; Rossi-Hansberg et al., 2021). Our data also suggests that the average employment size of all incumbent firms relative

to entrants rose from a factor of 5 in the early 1980s to 7.5 by 2014. The shift in employment shares is relevant in our context since it may reflect an increase in the average productivity of incumbents relative to entrants. For instance Hopenhayn et al. (2019) note that average exit rates for incumbents have fallen over time and Aghion et al. (2020) point to a falling cost of expansion for incumbents. Both of these factors could result in a higher concentration of employment at incumbent firms, relative to entrants. While the baseline model captures the aggregate productivity implications from the rise in the number of firms, in this section we extend the model to fully account for the observed shift in the share of nonemployer firms and to account for the shift in economic activity within the distribution of firms.

To derive implications for aggregate productivity over time, we use equation (6) with 1982 as the base year:

$$\frac{TFP_t}{TFP_{1982}} = \left(\frac{A_t}{A_{1982}}\right)^{1-\alpha} \left(\frac{\bar{z}_t}{\bar{z}_{1982}}\right)^{1-\alpha} \left(\frac{N_t/L_t}{N_{1982}/L_{1982}}\right)^{1-\alpha}. \tag{14}$$

Then, changes in TFP can then be decomposed into: (a) changes in a common productivity term for all firms A; (b) changes in average idiosyncratic productivity across firms $\bar{z}$; and (c) changes in the number of firms per worker, which as discussed previously, is independent in our framework from A and $\bar{z}$.

We continue to use our measure of the total number of firms per worker over time for N/L, but unlike in the baseline model where average firm-level productivity $\bar{z}$ is assumed constant, in the extended model we endeavour to infer the evolution of $\bar{z}$ that could result from the change in the share of nonemployers. We emphasize that the contributions of changes in A, N/L, and $\bar{z}$ to TFP enter multiplicatively and since N/L is independent of A and $\bar{z}$, the contribution of changes in the number of firms N/L to productivity growth remains the same as in Section 4. As in the baseline model, we treat the common productivity term A as a residual, equal to any growth in TFP not accounted for by the model. We now describe the specifics of our approach, starting with how we infer productivity from firm size.

The model provides a mapping from relative size to relative productivity, given by equation (12). Denoting the average employment size of all firms and entrants by $\bar{\ell}$ and $\bar{\ell}_{ent}$, average

productivity of all firms relative to entrants in each year is;

$$\frac{\bar{z}_t}{\bar{z}_{ent,t}} = \frac{\bar{\ell}_t}{\bar{\ell}_{ent,t}} \qquad \Rightarrow \qquad \bar{z}_t = \bar{z}_{ent,t}\frac{\bar{\ell}_t}{\bar{\ell}_{ent,t}}.$$

Note that, from equations (2) and (3), the same mapping applies between productivity z and output y or revenue, hence, the above expressions hold when substituting revenue for labor. Since entrants draw z from a constant distribution $G(z)$ (their average and expected z is constant), we infer average idiosyncratic productivity of all firms in each year relative to 1982 as;

$$\frac{\bar{z}_t}{\bar{z}_{1982}} = \frac{\bar{z}_{ent,t}}{\bar{z}_{ent,1982}} \cdot \left(\frac{\bar{\ell}_t/\bar{\ell}_{ent,t}}{\bar{\ell}_{1982}/\bar{\ell}_{ent,1982}}\right) = \frac{\bar{\ell}_t/\bar{\ell}_{ent,t}}{\bar{\ell}_{1982}/\bar{\ell}_{ent,1982}}, \tag{15}$$

noting that relative employment variables could be replaced with relative revenue.

To illustrate what kind of additional data is needed to infer changes in average idiosyncratic productivity, it is useful to decompose average productivity relative to entrants $\bar{z}/\bar{z}_{ent}$ as follows;

$$\frac{\bar{z}}{\bar{z}_{ent}} = \frac{N_{ent}}{N}\left(\frac{\bar{z}^{non}_{ent}N^{non}_{ent} + \bar{z}^{non}_{inc}N^{non}_{inc} + \bar{z}^{emp}_{ent}N^{emp}_{ent} + \bar{z}^{emp}_{inc}N^{emp}_{inc}}{\bar{z}^{non}_{ent}N^{non}_{ent} + \bar{z}^{emp}_{ent}N^{emp}_{ent}}\right), \tag{16}$$

where $\bar{z}$ on the left-hand side is a weighted average of the productivities of nonemployer and employer entrants, and nonemployer and employer incumbents, weighted by the number of firms in each category. Average productivity of entrants $\bar{z}_{ent}$ is defined analogously. This decomposition illustrates that $\bar{z}/\bar{z}_{ent}$ can be calculated given values for each of the eight variables on the right-hand side: four variables on the number of entrant and incumbent firms for employers and nonemployers; and four variables related to the average idiosyncratic productivity for each type of firm. As a result, with separate measures of labor ℓ (or revenue) for entrant and incumbent nonemployers and employers, in addition to measures of the number of firms of each type, equations (15) and (16) can be used to infer changes in average productivity relevant in equation (14).

While the BDS data provide information for entrant employer firms, the NES data do not

distinguish between entrant and incumbent nonemployers. Similarly, while the revenue share of nonemployers (for some years) is reported, the data do not distinguish between entrants and incumbents. We specify below how we separate nonemployer entrants and incumbents using additional data and assumptions. We proceed in two steps.

First, to infer the number of nonemployer entrants N_{ent}^{non}, we use information on exit rates. In particular, we use a survival rate of 85% for nonemployers in 1997 from Davis et al. (2009) and given the lack of exit data for nonemployers over time, we assume that the survival rate of nonemployers changes over time in proportion to the survival rate for employers, which we take from the BDS data. Hence, we impute the survival rate of nonemployers as follows:

$$(1 - \lambda_t^{non}) = 0.85 \cdot \frac{(1 - \lambda_t^{emp})}{(1 - \lambda_{1997}^{emp})}.$$

Given that the exit rate for employers from BDS data decreases from 10% in 1982 to 7.5% in 2014, the implied exit rate for nonemployers decreases from 15.8% to 13.5%. Using these exit rates and data on the total number of nonemployers each year N_t^{non}, we calculate the number of nonemployer entrants as:

$$N_{ent,t}^{non} = N_t^{non} - (1 - \lambda_{t-1}^{non}) \cdot N_{t-1}^{non}. \tag{17}$$

With the total number of nonemployers in each year, the number of incumbent nonemployers is given by $N_{inc,t}^{non} = N_t^{non} - N_{ent,t}^{non}$. Note that we abstract from nonemployer to employer transitions since we cannot differentiate between new employers and transitioning nonemployers. As a result, we potentially overstate the number of employer entrants. However, we argue this bias should not be quantitatively substantial given the small number of transitioning nonemployers in a given year (Davis et al., 2009). The number of incumbent employer firms N_{inc}^{emp} is simply calculated from data on the total number of employer firms N^{emp} and the number of employer entrants N_{ent}^{emp}.

This analysis implies that as a fraction of the total number of firms, the share of nonemployer

incumbents increases from 57% in 1982 to 68.9% in 2014, whereas the share of employer incumbents decreases from 23.7% to 16.1%, and the share of entrants nonemployer and employer decrease from 15.9% to 13.6% and 3.4% to 1.4%.

Second, to infer average idiosyncratic productivity of entrants and incumbents for employer and nonemployer firms given the lack of separate data for nonemployer firms, we make the following two assumptions: (a) average productivity of nonemployer entrants $\bar{z}_{ent}^{non}$ is constant over time and without loss of generality we normalize it to one, i.e., $\bar{z}_{ent}^{non,t} = 1$ for all t, and (b) average productivity of all entrants relative to nonemployer entrants $\bar{z}_{ent}/\bar{z}_{ent}^{non}$ is constant. Given the normalization in (a), this implies that $\bar{z}_{ent}$ is constant to some value. We explain below how we discipline this value.

Since average productivity of entrants is the average productivity of employer and nonemployer entrants weighted by their share of entrant firms,

$$\bar{z}_{ent} N_{ent} = 1 \cdot N_{ent}^{non} + \bar{z}_{ent}^{emp} N_{ent}^{emp},$$

we compute average employer entrant productivity for each year as;

$$\bar{z}_{ent}^{emp} = \frac{(\bar{z}_{ent} N_{ent} - 1 \cdot N_{ent}^{non})}{N_{ent}^{emp}}, \tag{18}$$

where $\bar{z}_{ent}$ is given from assumption (b) above. Using BDS data on employment at employer firms for entrants and incumbents, we calculate employer incumbent productivity as;

$$\bar{z}_{inc}^{emp} = \bar{z}_{ent}^{emp} \cdot \left(\frac{\bar{\ell}_{inc}^{emp}}{\bar{\ell}_{ent}^{emp}} \right). \tag{19}$$

We do not have data on employment at nonemployer firms (unpaid or owner-provided). To infer average idiosyncratic productivity of incumbent nonemployers $\bar{z}_{inc}^{non}$, we therefore use data on the revenue share of nonemployers, which the model assumes is equal to their employment share. Denote the revenue share of nonemployers as R^{non}. The model implies the following

relationship between the revenue share of nonemployers and average productivity $\bar{z}$;

$$R^{non} = \frac{\bar{z}^{non}_{ent} N^{non}_{ent} + \bar{z}^{non}_{inc} N^{non}_{inc}}{\bar{z}^{non}_{ent} N^{non}_{ent} + \bar{z}^{non}_{inc} N^{non}_{inc} + \bar{z}^{emp}_{ent} N^{emp}_{ent} + \bar{z}^{emp}_{inc} N^{emp}_{inc}},$$

which we use to infer the average idiosyncratic productivity of nonemployer incumbents;

$$\bar{z}^{non}_{inc} = \frac{R^{non} \cdot (\bar{z}^{emp}_{ent} N^{emp}_{ent} + \bar{z}^{emp}_{inc} N^{emp}_{inc}) - (1 - R^{non}) \cdot 1 \cdot N^{non}_{ent}}{(1 - R^{non}) \cdot N^{non}_{inc}}. \tag{20}$$

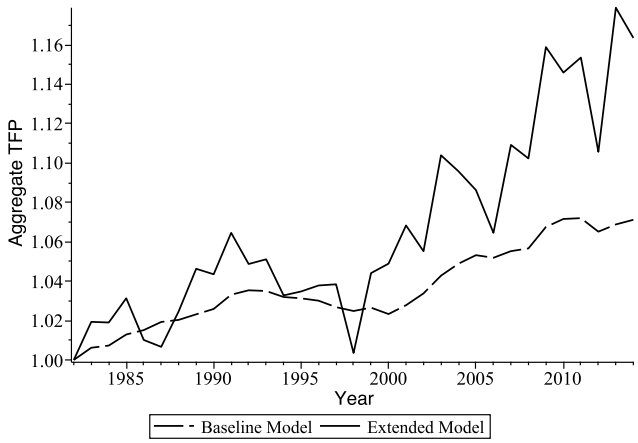
Note that we have data for the revenue share of nonemployers in Census years starting in 1987. For the years between census years, we assume that the average productivity of nonemployers relative to all firms changes smoothly from one census year to the next, and calculate the corresponding revenue share. We then calculate the average annual growth rate of this measure between 1987 and 2012, and use this growth rate to impute values before 1987 and after 2012. Our imputed revenue share of nonemployers increases from 2.71% in 1982 to 3.11% in 2014, and (as in the data) changes from 2.80% in 1987 to 3.07% in 2012.

We discipline the value of average productivity of entrants $\bar{z}_{ent}$ by ensuring that it implies a reasonable value for the productivity of nonemployer incumbents z^{non}_{inc}. We find that values of $\bar{z}_{ent} \geq 7$ ensure that $z^{non}_{inc} > z^{non}_{ent}$ for all years, with $\bar{z}_{ent} = 7$ generating an average $z^{non}_{inc}/z^{non}_{ent}$ equal to about 1.2 across the first five years (the implied z^{non}_{inc} is generally increasing over time). The implied value of z^{non}_{inc} increases with higher values of $\bar{z}_{ent}$, as a result, we use $\bar{z}_{ent} = 7$ as a conservative choice since the implied growth in aggregate TFP from 1982 to 2014 is also (slightly) increasing in $\bar{z}_{ent}$.

To summarize, equations (18) to (20) identify the values for $\bar{z}^{emp}_{ent}$, $\bar{z}^{emp}_{inc}$, and $\bar{z}^{non}_{inc}$, which together with the assumption that $\bar{z}^{non}_{ent} = 1$ provide all the additional information required to calculate average productivity relative to entrants $\bar{z}/\bar{z}_{ent}$ using equation (16). With $\bar{z}^{non}_{ent}$ normalized to one, this analysis implies that between 1982 to 2014, $\bar{z}^{non}_{inc}$ increases from 1.15 to 1.86, $\bar{z}^{emp}_{ent}$ from 35.6 to 65.3, and $\bar{z}^{emp}_{inc}$ from 118.9 to 268.7.

Using the implied values for $\bar{z}_t/\bar{z}_{ent,t}$ and data for N_t/L_t, we calculate the implied aggregate

Figure 5: Aggregate TFP in Extended Model



Notes: The solid line reports aggregate TFP in the extended model due to changes in the number of firms and average idiosyncratic productivity $\bar{z}$ relative to 1982. The dashed line is aggregate TFP in the baseline model. The ratio between the solid and dashed lines is the contribution to aggregate TFP of changes in average idiosyncratic productivity.

TFP from the last two terms in equation (14). Recall that A is treated as a residual to isolate implied changes in TFP due to changes in firm dynamics. Figure 5a reports aggregate TFP in the extended model compared with the baseline model. The implied measure of average firm-level productivity (15), equal to the ratio of the two measures of TFP in Figure 5a (normalized to 1 in 1982), is generally increasing from 1982 to 2014. This growth averages 0.07% per year between 1982 and 1997, then a much a higher 0.42% per year from 1997 to 2014. As a result, the cumulative increase in aggregate TFP is higher in the extended model than in the baseline model. From 1982 to 2014, the cumulative increase in TFP is 16.4% compared to 7.1% in the baseline model. The cumulative growth in TFP observed in the data between 1982 and 2014 amounts to 32%, and so the implied measure of TFP in the extended model accounts for 51% of the actual increase.

Since the model implies that the shares of employment and revenue of nonemployers are the same, we use revenue shares of nonemployers to bypass the lack of data on employment. This implicitly assumes that revenue per unit of labor for nonemployers is the same as revenue per labor of all firms. Is this a reasonable assumption? Presumably, revenue per labor of nonemployers is closer to small employers than to all employers. We look at disaggregated

employer data from the U.S. Census' Statistics of U.S. Businesses for 2012 and find that revenue per employee in the smallest employers (less than 5 employees) is equal to or greater than revenue per employee for all employers in 70 percent of the 2-digit industries (14 industries out of 20). It is also not clear there are systematic patterns of change over time. For instance, across all industries, revenue per employee across small employers is 13% lower than that across all employers in 2012 but about 8% higher in 1997. Nevertheless, we assess the potential relevance of our assumption using the revenue per employee of small employers. Taking the value for 2012, we assume a constant revenue per unit of labor for nonemployers that is 13% lower than for all employers. We find that the productivity effect of this adjustment is very small, an implied TFP increase between 1982 and 2014 of 16.38% compared with 16.35% using our simpler assumption in the extended model.

Aggregate TFP growth implied by the model can be decomposed into two parts. First, the increase in the number of firms per person by itself implies cumulative growth in TFP of 7.1% from 1982 to 2014, as captured by the baseline model.[1] This growth in productivity is the result of smaller firms on average and hence a higher marginal (and average) product of labor. Second, the shift in the firm size distribution suggests that incumbents have become more productive than entrants over time, due to a combination of lower exit rates and higher productivity growth over the life of the average firm. This shift suggests average productivity across all firms has increased by an additional 8.6% since 1982. We emphasize that our analysis allows for any underlying exit and growth dynamics by size and age among incumbent employers. What ultimately matters for our calculation is the average $\bar{z}$ among incumbents, relative to entrants (all relative to 1982), regardless of which incumbent employers are growing more or less, or surviving with higher or lower probabilities.

We also note that a 10-year moving average of our measures of TFP growth features a much higher correlation with productivity growth in the data (22.3%), suggesting that net firm entry

[1] When calculating firms per worker in Section 3 we use CPS data for our measure of aggregate employment. In Appendix D we recalculate firms per worker using a measure of aggregate employment inferred from the BDS data for robustness.

when including nonemployers in fact plays an important quantitative role in driving trends in aggregate productivity, especially relative to a measure of implied TFP growth in the baseline model that includes only employers, which features a correlation of -35% with productivity growth in the data. While our analysis is not designed to account for the short- and medium-run movements in TFP that surely depend on many factors not included in the model, we emphasize that the extended model significantly improves on the strong negative correlation between observed TFP growth and that implied by the model with employer firms as highlighted by Decker et al. (2016) and Li (2017).

6 Alternative Drivers of Firm Dynamics

There is a growing literature in macroeconomic development linking low average firm size (the ratio of workers to the number of firms) and low aggregate productivity to policy distortions in developing countries. It is therefore relevant to ask whether the increase in the number of firms, and hence the observed decline in average firm size when including nonemployers, may be due to changes in distortions rather than changes in exit and productivity growth patterns considered in our model. If evidence of this exists, then the increase in the number of firms may in fact imply a *negative* contribution to aggregate TFP, in contrast to what we report in Section 4. We consider two broad mechanisms as potential alternative drivers of the change in net firm entry: increased factor misallocation and increased barriers for firms to enter new markets.

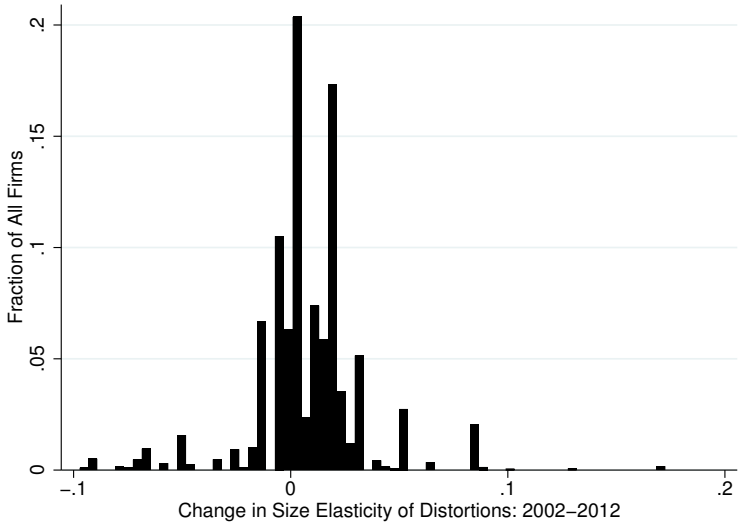
Increasing misallocation among employer firms. The literature investigating the extent and impact of the misallocation of productive factors across firms shows that 'random' misallocation need not affect the equilibrium number of firms. But if larger and more productive firms are effectively taxed at higher rates (or face tighter constraints) than smaller 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 firms that are less productive on average. 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. We therefore consider whether an increase in the prevalence of idiosyncratic distortions may be driving the increase in the number of firms in the U.S. economy over time.

To assess this mechanism, we use publicly-available 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. For manufacturing industries, we use establishments rather than firms, and value added rather than revenue. Although we only have data from 2002 to 2012, this period is still characterized by a substantial increase in the number firms of 22%. For each size bin within an 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 that in the absence of distortions, firms within an industry choose labor inputs to equalize the average product of labor across firms within the industry. 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 larger implicit taxes within a particular industry.

For each industry, we regress (logged) average products on (logged) employment size, using the number of firms per bin/observation as weights, to obtain industry-specific estimates of the elasticity of distortions with respect to size in each year as our measure of misallocation. 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 -0.075, with a variance across industries of 0.02. Taken at face value, this suggests that a typical industry features larger firms behaving as if they face slightly *lower* effective taxes (or constraints) compared with smaller firms. For 2012, we find an average elasticity closer to zero, with a lower variance.

Figure 6: Distribution of Changes in the Size Elasticity of Distortions



These results suggest that there is essentially no change in the extent of misallocation over time as measured by the size elasticity of distortions. To address the potential issue of unpaid workers not being counted in small firms, we have reproduced these estimates using only firms with at least 5 employees, with the resulting estimated elasticities even closer to zero and even lower variances, reinforcing our findings.

While the average industry has low and roughly constant elasticity, it is still possible that a significant number of firms are in industries that experienced larger increases in the size elasticity of distortions. Figure 6 addresses this possibility, reporting the fraction of firms in industries that saw a given change in the size elasticity of distortions from 2002 to 2012. The histogram indicates that the vast majority of firms are in industries that saw little change in this elasticity over time. This finding is consistent with Bils et al. (2020) who use more disaggregated data to account for measurement error and show that there is no significant change in misallocation over time in the U.S. economy. Although we do not measure total misallocation, the focus of Bils et al., our results suggest that correlated distortions (a particularly harmful driver of misallocation) are not becoming worse over time. We conclude that increased misallocation across employer firms is not driving the increase in the number of firms in the U.S. economy.

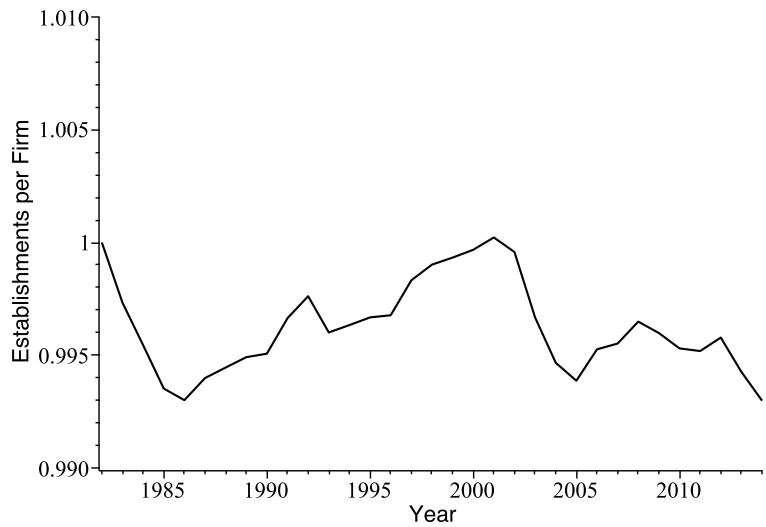
Increasing misallocation between employer and nonemployers. The previous evidence relies on measures of distortions inferred from employer data, and as a result it abstracts from potential distortions affecting firms' decisions whether to have employees or not. We now turn to assess this possibility. Consider that all potential entrants must pay an entry cost (c_E) to become a producer as before, but now firms that demand more than one unit of labor, that is demand labor beyond the labor supplied from the owner-manager, must incur a proportional tax on revenue equal to τ_e. This tax, a size-dependent distortion, discourages firms that would otherwise hire more than one unit of labor, and results in an inaction zone between two distinct thresholds of productivity, z_1 and z_e. This mechanism is similar to that in Guner et al. (2008), who study the impact of size-dependent distortions across countries. They focus on effective taxes incurred by employer firms that apply after some legal size threshold. Firms with productivity $z < z_1$ demand labor given the wage below one unit of labor (interpreted as the labor supplied by the owner-manager). Firms with productivity $z > z_e$ hire labor optimally given the wage and the tax τ_e. And firms with productivity $z \in (z_1, z_e)$, in the inaction zone, choose to operate with only one unit of labor to avoid the tax. In Appendix F we show that a decrease in c_E raises z_1 and z_e in the same proportion, whereas an increase in τ_e lowers z_1 and increases z_e/z_1. In this setting, it is possible that the observed increase in the fraction of nonemployers over time could be associated with an increase in labor misallocation between employer and nonemployer firms and lower aggregate TFP, relative to that inferred by the baseline model. If an increase in τ_e were responsible for the increase in nonemployers, a growing fraction of firms would be firms that choose to stay small in order to avoid τ_e. In Appendix F, we assess this possibility by calibrating the tax on employers to match the observed increase in the fraction of nonemployers over time. The τ_e we infer from the data is very low and increases only slightly over time, from 1.44% in 1982 to 1.87% in 2014. When combined with the inferred drop in c_E (virtually identical to that implied by the baseline model), we find that the large drop in c_E *decreases* the fraction of firms that are constrained nonemployers (those that would be employers if $\tau_e = 0$), even while the total fraction of nonemployers increases as in the data. The net result is an implied cumulative increase in aggregate TFP from 1982 to 2014 equal

to 7.1%, which is the same increase as in the baseline model. Although the slightly increasing implied tax by itself lowers TFP by increasing misallocation, the declining implied entry cost increases TFP by both increasing the number of firms (as in our baseline model) and reducing misallocation. This reinforces our previous conclusion that increased misallocation is not likely driving the increase in the number of firms or the share of nonemployers in the U.S. economy over time.

Increasing entry barriers to new markets. Another possible alternative explanation for the increase in the number of firms is higher barriers to entering a market. Bento (2020) 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 7 reports that the number of establishments per firm essentially remained constant from 1982 to 2014, even as the number of firms per worker grew by 41%. Note that Aghion et al. (2020) and Cao et al. (2020) 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.

Increasing constraints for entrants. Our framework assumes that entering firms operate at their optimal scale, while in practice there may be frictions that affect their operational scale

Figure 7: The Number of Establishments per Firm (Relative to 1982)



and changes in these frictions could affect our inferences. In particular, if financial constraints have gotten worse for entrants over time relative to other firms, then this could account for some of the increase in the relative employment or revenue of incumbents to entrants. This feature would imply lower idiosyncratic productivity for entrants and, as a result, our analysis would be over-stating the increase in average idiosyncratic productivity. To assess the potential impact of this, we note that we don't have data on capital inputs by firm age, nevertheless, we do have revenue and payroll data by age for employer firms in 2007 and 2012 that can be used to investigate the possibility of changing constraints over time. From Hsieh and Klenow (2009), the ratio of revenue to wage bill can be interpreted as a broad measure of the labor wedge $(1 + \tau_L)$, that is implicit 'taxes' on labor inputs. The idea is that firms that are financially constrained should have problems financing employees through 'working capital.' As a result, financial constraints should show up as contributing to the distortion measure τ_L. If τ_L is increasing over time more for entrants than for incumbent firms, this would suggest that financial constraints are becoming more binding for entrants (relative to incumbents) over time. Based on these assumptions we find: (a) the ratio of revenue to wage bill is lower for entrants in both 2007 and 2012; and (b) the ratio of revenue to wage bill for entrants relative to all firms decreases from 2007 to 2012. We take this admittedly crude evidence to suggest that our analysis is not

over-stating the increase in average idiosyncratic productivity $\bar{z}$ due to changes in constraints over time.

In summary, the evidence presented in this section suggests that various mechanisms encouraging firm entry while lowering TFP, which may be prevalent in developing countries, may not be driving the increase in firm net entry or the increase in the share of nonemployer firms in the United States; and we find no evidence of increasing constraints to entering firms. Nevertheless, the evidence is tentative due to the lack of more granular data at the firm level and future work may be needed to reassess these conclusions.

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 employer firms 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 small relative to employers, the increase in the number of firms and in firm-level productivity together imply that the increase in net firm entry has been responsible for between one quarter to one half of observed aggregate productivity growth from 1982 to 2014. This is in striking contrast to a decrease in productivity implied by a model considering only employer firms.

Decker et al. (2016) and Li (2017) show that standard measures of business dynamism, which focus 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 firm net entry, which accounts for nonemployers and the evolution of the size distribution over time, follows TFP growth in the data more closely since the 1980s. Nevertheless, an important question remains

on what accounts for the productivity slowdown in the U.S. economy in recent decades. Our results indicate it is not a decline in firm net entry. One promising recent study instead focuses on the decline in the quality of innovative activity resulting from a misallocation of R&D investments across firms (Ayerst, 2020), driven by firm-level heterogeneity in the wedge between the private and social return to innovation.

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 and price-cost markups, as documented in De Loecker et al. (2020), and Rossi-Hansberg et al. (2021). 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 net entry. We have abstracted from the underlying causes of changes in exit rates and productivity growth across firms and over time. Understanding these patterns remains an important area of research (Aghion et al., 2020; Cao et al., 2020). Similarly, given the growing importance of nonemployer firms in the U.S. data, it is essential to document and better understand the nature of nonemployer business activity. We leave these important explorations for future research.

References

Abraham, Katharine G., John Haltiwanger, Kristin Sandusky, and James Spletzer, "The Rise of the Gig Economy: Fact or Fiction?," *AEA Papers and Proceedings*, 2019, *109*, 357–361.

Acs, Zoltan J., Brian Headd, and Hezekiah Agwara, "Nonemployer Start-up Puzzle," Technical Report, Small Business Administration, Office of Advocacy 2009.

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li, "A Theory of Falling Growth and Rising Rents," 2020. Unpublished manuscript.

Akcigit, Ufuk and Sina T. Ates, "What Happened to U.S. Business Dynamism?," *NBER Working Paper No. 25756*, 2019.

Asker, John, Allan Collard-Wexler, and Jan De Loecker, "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 2014, *122* (5), 1013–1063.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," *Quarterly Journal of Economics*, 2020, *135* (2), 645–709.

Ayerst, Stephen, "Innovator Heterogeneity, R&D Misallocation, and the Productivity Growth Slowdown," 2020. Unpublished manuscript, International Monetary Fund.

Bento, Pedro, "Competition, Innovation, and the Number of Firms," *Review of Economic Dynamics*, 2020, *37*, 275–298.

_ **and Diego Restuccia**, "Misallocation, Establishment Size, and Productivity," *American Economic Journal: Macroeconomics*, 2017, *9* (3), 267–303.

_ **and** _ , "On average establishment size across sectors and countries," *Journal of Monetary Economics*, 2021, *117*, 220–242.

Bils, Mark, Peter J. Klenow, and Cian Ruane, "Misallocation or Mismeasurement?," *NBER Working Paper No. 26711*, 2020.

Bollard, Albert, Peter J. Klenow, and Huiyu Li, "Entry Costs Rise with Development," 2016. Unpublished manuscript, Stanford University.

Cao, Dan, Henry R. Hyatt, Toshihiko Mukoyama, and Erick Sager, "Firm Growth Through New Establishments," 2020. Unpublished manuscript.

Davis, Steven J., John Haltiwanger, Ronald S. Jarmin, C. J. Krizan, Javier Miranda, Alfred Nucci, and Kristin Sandusky, "Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Nonemployer Universes," *Producer Dynamics: New Evidence from Micro Data*, 2009, pp. 329–366.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger, "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda, "Declining Business Dynamism: Implications for Productivity," *Hutchins Center Working Paper No. 23*, 2016.

Decker, Ryan, John Haltiwanger, Ron S. Jarmin, and Javier Miranda, "The Secular Decline of Business Dynamism in the United States," 2014. Unpublished manuscript, University of Maryland.

Fairlie, Robert W., Javier Miranda, and Nikolas Zolas, "Job Creation and Survival among Entrepreneurs: Evidence from the Universe of U.S. Startups," 2018. Unpublished manuscript, University of California, Santa Cruz.

Federal Reserve Bank of Cleveland, "10-Year Real Interest Rate [REAINTRATREARAT10Y]," Technical Report, retrieved from FRED, Federal Reserve Bank of St. Louis 2022. https://fred.stlouisfed.org/series/REAINTRATREARAT10Y (accessed October 21, 2022).

Fernald, John G., "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," *FRBSF Working Paper No. 2012-9*, 2012.

Furman, Jason and Peter Orszag, "Slower Productivity and Higher Inequality: Are They Related?," *PIIE Working Paper No. 18-4*, 2018.

Guner, N., G. Ventura, and Y. Xu, "Macroeconomic Implications of Size-Dependent Policies," *Review of Economic Dynamics*, 2008, *11* (4), 721–744.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda, "Who Creates Jobs? Small versus Large versus Young," *Review of Economics and Statistics*, 2013, *95*, 347–361.

Hopenhayn, Hugo A., "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 1992, *60*, 1127–50.

Hopenhayn, Hugo, Julian Neira, and Rish Singhania, "From Population Growth to Firm Demographics: Implications for Concentration, Entrepreneurship and the Labor Share," *NBER Working Paper No. 25382*, 2019.

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

_ and _ , "The Life Cycle of Plants in India and Mexico," *Quarterly Journal of Economics*, August 2014, *129* (3), 1035–1084.

Hyatt, Henry, Seth Murray, and Kristin Sandusky, "Business Income Dynamics and Labor Market Fluidity," 2020. Unpublished manuscript.

Karahan, Fatih, Benjamin Pugsley, and Ayşegül Şahin, "Demographic Origins of the Startup Deficit," 2019. Unpublished manuscript.

Kehrig, Matthias and Nicolas Vincent, "The Micro-Level Anatomy of the Labor Share Decline," *Quarterly Journal of Economics*, 2021, *136* (2), 1031–1087.

Koske, Isabell, Isabelle Wanner, Rosamaria Bitetti, and Omar Barbiero, "The 2013 update of the OECD's database on product market regulation: Policy insights for OECD and non-OECD countries," *OECD Economics Department Working Papers*, 2015, *1200.*

Li, Huiyu, "How Does Business Dynamism Link to Productivity Growth?," *FRBSF Economic Letter No. 2017-01*, 2017.

Pugsley, Benjamin and Ayşegül Şahin, "Grown-Up Business Cycles," *Review of Financial Studies*, 2019, *32*, 1102–1047.

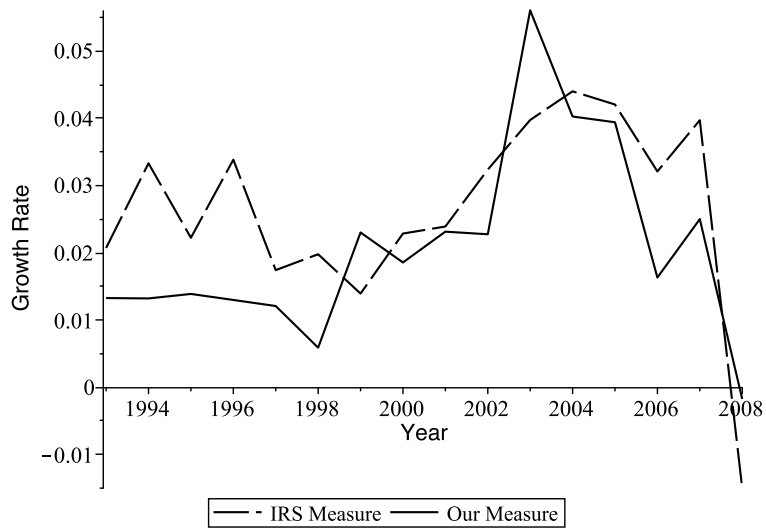
Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter, "Diverging Trends in National and Local Concentration," *NBER Macroeconomics Annual 2020*, 2021, *35*, 115–150.

Appendix

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 A.1 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 A.1: The Growth Rate in the Number of Firms, IRS and Our Measure



B Sectoral Composition in the 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.[2] 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%).

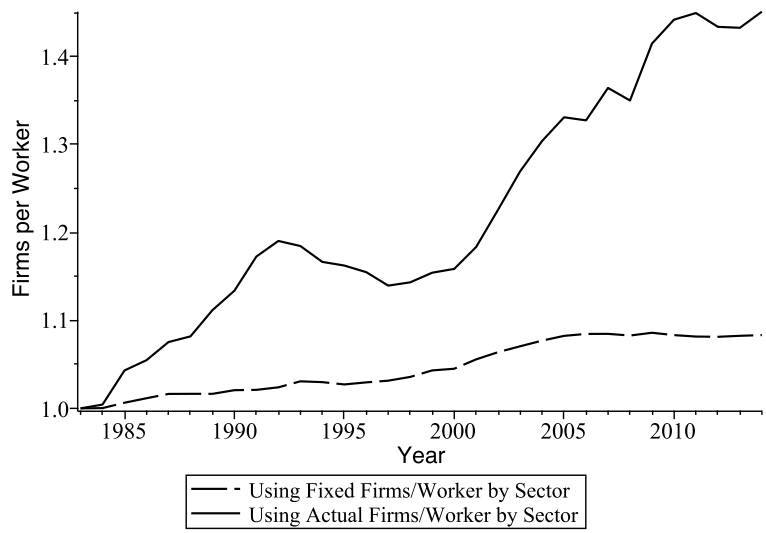
Table B.1: The Role of Structural Transformation in Total Firms per Worker

	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

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 B.2 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.

[2]We do not have data identifying government workers by industry, so sectoral measures of employment include these workers.

Figure B.2: 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 (sectoral) 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.

C Extension with Overhead Costs

In the baseline model in Section 4, we assume entry costs are the only non-production costs incurred by firms. We extend the baseline model to show that allowing for realistically-specified overhead costs (and allowing these costs to change over time) does not affect the main implications of the model. We start by noting that specifying overhead costs as fixed in nature (unrelated to the productivity of a firm) is inconsistent with the presence of very small firms in the data, a point made by Hsieh and Klenow (2014) among others. As a result of this insight, the literature specifies overhead costs as increasing in firm productivity. We therefore follow Asker et al. (2014) in assuming that a producer with productivity z must incur an output cost equal to $z \cdot c_p \cdot Y/L$ each period in order to operate. Instead of equation (4), operating profits for a producer with productivity z can now be expressed as;

$$\pi(z) = Az\left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}}(1-\alpha) - z \cdot c_p \cdot Y/L = \frac{w}{\alpha} \cdot \left(\frac{z}{\bar{z}}\right)\left[(1-\alpha)\left(\frac{L}{N}\right) - \bar{z}c_p\right], \tag{C.1}$$

where the right-hand-side expression takes into account equilibrium values of w and Y/L. Note that the square-bracketed terms imply that a firm's decision about whether to operate is independent of productivity. If a firm chooses to operate, all firms do (conditional on the number of firms). Free entry ensures that the discounted value of expected lifetime profits (net of operating costs) at entry is exactly equal to the cost of entry, resulting in the following

characterization for the number of firms in equilibrium;

$$\frac{N}{L} = \frac{1-\alpha}{c_E(1-\rho) + \bar{z}c_p}. \tag{C.2}$$

For a given average productivity $\bar{z}$, the number of firms per person is increasing in the sum of $c_E(1-\rho) + \bar{z}c_p$. The only difference with the baseline model, then, is that an increase in the number of firms per person implies a decrease in the discounted expected value of all non-production costs. Because the addition of c_p does not change the relationship between TFP and N/L, and does not affect $\bar{z}$, the model-implied contribution of a higher number of firms per person to TFP is left unchanged.

D Alternative Measures of Aggregate Employment

We assess our main results using alternative measures of aggregate employment.

Accounting for farm workers in the CPS. Farm workers are included in the CPS data we use for aggregate employment, while farms are not included in our firm counts. We do not have data specifically for farm workers, which is why we use the aggregate CPS data in our calculations. But the CPS does report the total number of workers in farming, fishing, and forestry occupations starting in 1983. Although this includes workers in agricultural firms accounted for in our firm data, we can nevertheless check whether removing these employees results in any difference in aggregate labor trends over time. We find it does not. After removing these workers, the overall increase in aggregate employment is exactly the same as for the aggregate CPS data we use in the main body of our paper. As a consequence, using this alternative measure of aggregate employment does not affect any of our results.

Aggregate employment from BDS data. In Section 5 we combine BDS employment data for employers with revenue share data for both employers and nonemployers to infer the share of aggregate employment in nonemployer firms. But we continue to use CPS data as our measure of aggregate employment. It is important to note that these two data sources capture two different concepts of of employment – the CPS is measuring workers, while the BDS is measuring formal jobs. A worker can be employed in multiple jobs (formal and informal), while a job may employ either a part-time or full-time (formal) worker. Clearly a better measure of employment across firms for our purposes would be number of hours, but we use BDS data as the best available alternative. For aggregate employment, we believe the CPS measure matches the model most closely. Although it does not account for hours, neither does the BDS data.

And the CPS has two important advantages. First, it does not double-count workers with multiple jobs. Second, it accounts for workers not being paid formal wages—owners, family workers, under-the-table workers, among others. Nevertheless, we can recalculate our measure of firms per worker each year and the contribution of its increase to aggregate TFP, using the same strategy to infer employment at nonemployers based on BDS employment and revenue share data, and simply adding this to total BDS employment (ignoring the CPS). When we do this, we find aggregate employment increases by a larger 61% from 1982-2014, compared to 54% in the CPS data. As a result, the number of firms per worker grows by a lower 35%, rather than 41%. This translates into a cumulative increase in aggregate TFP of 6.2%, compared to the 7.1% we report in Section 4.

E Alternative Beliefs

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 1982, such that L and c_E were previously constant at 1982 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.

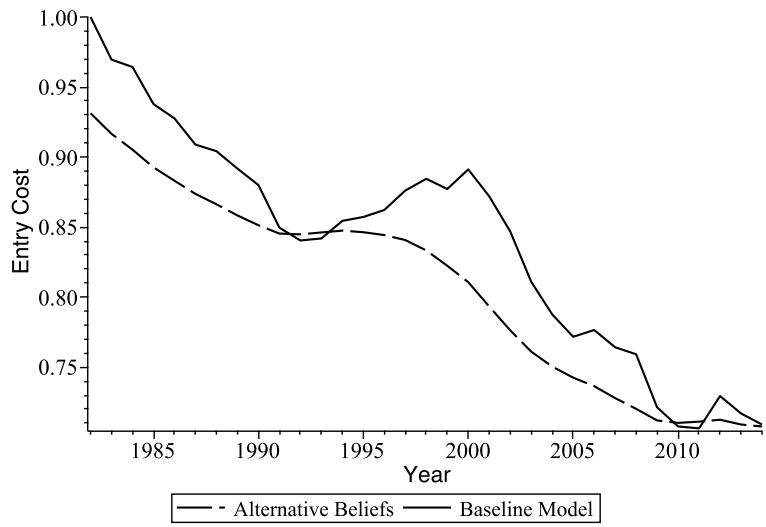
Free entry requires that the expected discounted profits of entrants are exactly equal to the entry cost in the period they enter. This can be expressed recursively as;

$$c_{e,t} = \frac{(1-\alpha)}{(N_t/L_t)} + \rho c_{e,t+1}, \tag{E.3}$$

$$c_{e,2014} = \frac{(1-\alpha)}{(N_{2014}/L_{2014})(1-\rho)}. \tag{E.4}$$

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

Figure E.3: Implied Entry Cost with Alternative Beliefs



current number of firms per worker. In Figure E.3 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 1982, 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.

F Share of Nonemployers

To assess the ability of our baseline model to explain the change in the number of nonemployers over time, given observed changes in the total number of firms, we calibrate the distribution of productivity across firms $G(z)$ to match U.S. data for 1987. The BDS data includes the number of employer firms falling within 12 different employment-size bins, along with average employment per employer firm within each bin. Two issues arise here. First, we do not have data for labor used in nonemployer firms (ie., owners, informal workers). Second, the BDS data does not report informal workers and owners in employer firms. To address these issues we use data from the Survey of Business Owners and the Statistical Abstract of the U.S., from which we can calculate the share of aggregate revenue generated by nonemployer firms. This data is available in census years starting in 1987.

We start by noting that in the model the share of aggregate revenue generated by a group

of firms is equal to its share of aggregate labor, from (2) and (3). We therefore assume that the share of labor used across all nonemployers is equal to the share of revenue generated by nonemployers. For our purposes, we need a distribution of labor across nonemployers and across small employers (those with less than 5 formal employees). We make a further assumption about the employment data, that the difference between total persons engaged (from the CPS) and formal employees (from the BDS) is distributed across all firms in proportion to their reported number of employees. In other words, if firm A is reported to have 5 times the employees of firm B, then we assume firm A also has 5 times the number of total workers as firm B. In 1987 nonemployers are therefore assumed to employ 2.8% of aggregate labor, while small employers employ 6.2%. Taking into account the numbers of each type of firm, these translate to an average employment size of 0.20 for nonemployers and 2.41 for small employers.

To map the above results to relative firm-level productivity, we note that equation (2) implies the following productivity ratio between two arbitrary firms i and j with different levels of employment;

$$\frac{z_i}{z_j} = \frac{\ell_i}{\ell_j}. \tag{F.5}$$

To infer the distribution of productivity across nonemployers and small employers, we assume nonemployers are firms with optimal (reported) labor less than 1, which translates to optimal actual labor less than 1.05. For comparison, small employers are firms with actual labor between 1.05 and 5.24 (1-5 multiplied by 1.049). We assume the distribution of productivity within each group is described by the Beta distribution, $\text{Beta}(a, 1|z_L, z_U)$, where z_L and z_U are the bounds of each distribution, $z_L = 0$ for nonemployers, and a is specific to the group.[3] The flexibility of this distribution allows us to match the average z (average labor) for each group from the data. Doing this, we obtain $a_{non} = 0.24$ and $a_{small} = 0.48$.[4] Given $G(z)$, we then calculate the fraction of firms that are nonemployers generated by the model as the number of firms changes each year (again, we assume the entry cost changes over time to generate this outcome).

G Tax on Employers

We extend the baseline model by assuming that any firm demanding more than one unit of labor (an 'employer') must pay a proportional tax rate τ_e on revenue. This size-dependent tax discourages firms that would otherwise demand more than one unit of labor from doing so, and results in two distinct thresholds of productivity, z_1 and z_e. Firms with productivity

[3] The pdf of this distribution is $f(z \mid z_L < z < z_U) = \frac{a(z-z_L)^{a-1}}{(z_U-z_L)^a}$.

[4] We only need within-group distributions of productivity for nonemployers and small employers, as the share of these firms in 1987 is 88%, while the share of nonemployers never rises above 82.5%.

$z < z_1$ demand less than one unit of labor given the wage, which is assumed to be supplied by the owner-manager. Firms with productivity $z > z_e$ demand more than one unit of labor optimally given the wage and the tax τ_e. And firms with productivity $z \in (z_1, z_e)$, the inaction zone, choose to operate as nonemployers with only with one unit of labor supplied by the owner-manager. We now have three expressions characterizing optimal labor conditional on productivity z;

$$\ell(z \mid z < z_1) = Az\left(\frac{\alpha}{w}\right)^{\frac{1}{1-\alpha}}, \tag{G.6}$$

$$\ell(z \mid z_1 < z < z_e) = 1, \tag{G.7}$$

$$\ell(z \mid z > z_e) = Az\left(\frac{\alpha}{w}\right)^{\frac{1}{1-\alpha}}(1-\tau_e)^{\frac{1}{1-\alpha}}, \tag{G.8}$$

with the following corresponding expressions for output and operating profits;

$$y(z \mid z < z_1) = Az\left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}}, \tag{G.9}$$

$$y(z \mid z_1 < z < z_e) = (Az)^{1-\alpha}, \tag{G.10}$$

$$y(z \mid z > z_e) = Az\left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}}(1-\tau_e)^{\frac{\alpha}{1-\alpha}}, \tag{G.11}$$

$$\pi(z \mid z < z_1) = (1-\alpha)Az\left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}}, \tag{G.12}$$

$$\pi(z \mid z_1 < z < z_e) = (Az)^{1-\alpha} - w, \tag{G.13}$$

$$\pi(z \mid z > z_e) = (1-\alpha)Az\left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}}(1-\tau_e)^{\frac{1}{1-\alpha}}. \tag{G.14}$$

We can pin down z_1 using $\ell(z_1) = 1$, resulting in the following;

$$z_1 = A^{-1}\left(\frac{\alpha}{w}\right)^{\frac{-1}{1-\alpha}}. \tag{G.15}$$

The threshold z_e can be characterized by noting that a firm with productivity z_e is indifferent between hiring more than one unit of labor while paying the tax τ_e and using only one unit of labor without paying the tax;

$$(1-\alpha)Az_e\left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}}(1-\tau_e)^{\frac{1}{1-\alpha}} = (Az_e)^{1-\alpha} - w.$$

Combining the above expression with equation (G.15), we get z_e/z_1 as an implicit function of τ_e;

$$(1-\alpha)(1-\tau_e)^{\frac{1}{1-\alpha}} = \left(\frac{z_1}{z_e}\right)^{\alpha} - \alpha\left(\frac{z_1}{z_e}\right). \tag{G.16}$$

Given that the right-hand side of the above equation is decreasing in z_e/z_1, the ratio z_e/z_1 must increase when τ_e increases.

For tractability we now assume $G(z)$ can be described by a Pareto distribution with lower bound equal to 1, scale parameter $\xi > 1$, and cdf $G(z) = 1 - z^{-\xi}$.[5] Given that potential entrants only learn their z after incurring the entry cost, $G(z)$ also describes the distribution of z across producers. The fraction of firms that are employers, i.e., use labor other than that of the firm owner, is therefore;

$$\frac{N_e}{N} = z_e^{-\xi}. \tag{G.17}$$

Further, note the number of firms that are employers as a fraction of firms that would be employers if $\tau_e = 0$ (given w) is;

$$\frac{1 - G(z_e)}{1 - G(z_1)} = \left(\frac{z_1}{z_e}\right)^{\xi},$$

which is decreasing as τ_e increases.

Labor market clearing now implies the following;

$$\begin{aligned} \frac{L}{N} &= G(z_1) A \left(\frac{\alpha}{w}\right)^{\frac{1}{1-\alpha}} \int_{z<z_1} z \; dG(z|z < z_1) + G(z_e) - G(z_1) \\ &\quad + [1 - G(z_e)] \, A \left(\frac{\alpha}{w}\right)^{\frac{1}{1-\alpha}} (1 - \tau_e)^{\frac{1}{1-\alpha}} \int_{z>z_e} z \; dG(z|z > z_e), \end{aligned}$$

or

$$\frac{L}{N} = \left(\frac{\xi}{\xi - 1}\right) \frac{z_1^{\xi-1} - 1}{z_1^{\xi}} + \frac{1}{z_1^{\xi}} \left(1 - \left(\frac{z_1}{z_e}\right)^{\xi}\right) + \frac{(1 - \tau_e)^{\frac{1}{1-\alpha}}}{z_1^{\xi}} \left(\frac{\xi}{\xi - 1}\right) \left(\frac{z_1}{z_e}\right)^{\xi-1}. \tag{G.18}$$

Aggregate output per worker can be calculated in a similar way, resulting in the following expression;

$$\frac{Y}{L} = \frac{A^{1-\alpha}\xi}{z_1^{\xi+\alpha-1}} \left(\frac{N}{L}\right) \left[\frac{z_1^{\xi-1} - 1}{\xi - 1} + \frac{1}{\xi + \alpha - 1} \left(1 - \left(\frac{z_1}{z_e}\right)^{\xi+\alpha-1}\right) + \frac{(1 - \tau_e)^{\frac{\alpha}{1-\alpha}}}{\xi - 1} \left(\frac{z_1}{z_e}\right)^{\xi-1}\right]. \tag{G.19}$$

We change slightly how we specify the cost of entry, relative to the baseline model. The cost is still in terms of the consumption good, but we now assume it scales up with the wage w rather

[5] Note that the lower bound of the productivity distribution is a free parameter in this framework. By normalizing it to 1, we assign all exogenous productivity changes over time to A.

than Y/L. This serves to simplify the free entry condition, which can be expressed as;

$$\begin{aligned} w \cdot c_E &= \frac{w}{1-\rho} \cdot \Bigg[G(z_1) A \left(\frac{1-\alpha}{\alpha}\right) \left(\frac{\alpha}{w}\right)^{\frac{1}{1-\alpha}} \int_{z<z_1} z dG(z|z<z_1) \\ &\quad + (G(z_e) - G(z_1)) \left(\frac{A^{1-\alpha}}{w} \int_{z_1<z<z_e} z^{1-\alpha} dG(z|z_1<z<z_e) - 1 \right) \\ &\quad + [1 - G(z_e)] A \left(\frac{1-\alpha}{\alpha}\right) \left(\frac{\alpha}{w}\right)^{\frac{1}{1-\alpha}} (1-\tau_e)^{\frac{1}{1-\alpha}} \int_{z>z_e} z \; dG(z|z>z_e) \Bigg], \end{aligned}$$

or

$$\begin{aligned} \frac{c_E \alpha z_1^{\xi}(1-\rho)}{\xi} &= \frac{1-\alpha}{\xi-1}\left(z_1^{\xi-1} - 1\right) + \frac{1}{\xi+\alpha-1}\left(1 - \left(\frac{z_1}{z_e}\right)^{\xi+\alpha-1}\right) \\ &\quad - \frac{\alpha}{\xi}\left(1 - \left(\frac{z_1}{z_e}\right)^{\xi}\right) + \frac{(1-\alpha)(1-\tau_e)^{\frac{1}{1-\alpha}}}{\xi-1}\left(\frac{z_1}{z_e}\right)^{\xi-1}. \end{aligned} \tag{G.20}$$

The revenue share of nonemployers can be derived and characterized as follows;

$$rev.\ share_{non} = \frac{\frac{z_1^{\xi-1}-1}{\xi-1} + \frac{1}{\xi+\alpha-1}\left(1 - \left(\frac{z_1}{z_e}\right)^{\xi+\alpha-1}\right)}{\frac{z_1^{\xi-1}-1}{\xi-1} + \frac{1}{\xi+\alpha-1}\left(1 - \left(\frac{z_1}{z_e}\right)^{\xi+\alpha-1}\right) + \frac{(1-\tau_e)^{\frac{\alpha}{1-\alpha}}}{\xi-1}\left(\frac{z_1}{z_e}\right)^{\xi-1}}. \tag{G.21}$$

From the above expressions, we derive the following observations from comparative statics. First, a higher τ_e increases z_e/z_1 and therefore increases the number of firms that are nonemployers because of τ_e, relative to the number of actual employers. Second, a higher c_E does not affect z_e/z_1. Third, if increases in c_E and τ_e lower the equilibrium wage, then they must lower z_1.

To assess the quantitative implications of the model, we need values for c_E, τ_e, and ξ. For ξ, we exploit the mapping from relative firm-level employment to relative firm-level productivity across employer firms, choosing $\xi = 1.08$ to match the ratio of average to median employment for employer firms (BDS data for manufacturing firms in 2007). We then choose parameter values for c_E and τ_e in each year to match the number of firms per person engaged and the fraction of firms that are employers. We obtain these values, along with values for z_1 and z_e, using equations (G.16), (G.17), (G.18), and (G.20). We cannot target the revenue share of nonemployers, as this is not possible when the productivity distribution is assumed to be Pareto. We nevertheless report the change over time in model-implied revenue share.

Table G.2 reports outcomes of interest for 1982 and 2014, comparing model-generated outcomes to the data. Our inferred cost of entry c_E decreases by 29%, as in the baseline model. This

Table G.2: Tax on Employers, 2014 vs. 1982

	Model		Data	
	1982	2014	1982	2014
c_E	7.99	5.68	–	–
τ_e	1.44%	1.87%	–	–
Fraction nonemployers to all firms	0.73	0.82	0.73	0.82
Fraction firms to persons	1.00	1.41	1.00	1.41
Fraction constrained nonemployers to all firms	0.18	0.15	–	–
Revenue share nonemployers (1987, 2012)	0.096	0.119	0.028	0.031
Aggregate TFP (1982=1)	1.000	1.071	1.000	1.320

by itself leads to a higher z_1 and an identical increase (in percentage terms) in z_e, implying no increase in the number of constrained nonemployers (those who would choose to be employers if $\tau_e = 0$) relative to the total number of would-be employers (all firms who would choose to be employers if $\tau_e = 0$). In Section 4 we note that such a drop in c_E can not increase the fraction of nonemployers enough relative to what we observe in the data. This rationalizes why our inferred tax τ_e must increase in the model in order to push z_e higher to match the increase in the share of nonemployers.

We emphasize that the inferred tax is a very low 1.44% in 1982 and only increases to 1.87% by 2012. This leads to both a drop in z_1 and an increase in z_e/z_1, resulting in a slight increase in the number of constrained nonemployers relative to would-be employers. But as a fraction of all firms, the number of constrained nonemployers drops from 18% to 15% as the total number of nonemployers increases. While the increase in τ_e leads to a shift in employment from employers to nonemployers, the drop in the number of constrained nonemployers offsets this, resulting in the same net increase in measured TFP as in the baseline model. To gain a better understanding of how the increase in τ_e inferred in this experiment contributes to the outcomes in Table G.2, we perform an experiment where we increase τ_e by the same amount but keep c_E fixed at its 1982 level. We find that that the fraction of nonemployers increases only slightly from 73% to 75%, and the number of constrained nonemployers, relative to all firms, increases slightly from 18% to 21%. This implies the decline in entry costs more than offsets any increase in misallocation due to a higher τ_e.

The assumption of a Pareto distribution of productivity does not allow the model to match the revenue share of nonemployers observed in the data. Nevertheless, we emphasize that the nonemployer revenue share generated by this model increases by 23% from 1987 to 2012 (the period for which we have revenue share data for comparison), much larger than observed in the data and only slightly lower than that implied by the baseline model.

In summary, while considering an effective tax on employers allows the model to match exactly the increase over time in the fraction of nonemployers, the implications for aggregate productivity are identical to those from the baseline model. We conclude that an increase in the effective cost of being an employer, relative to a nonemployer, is not a significant driver of trends in the U.S. economy. We also conclude that relaxing the assumption of an invariant incumbent firm dynamics over time, as we have done in the extended model of Section 5, is important for capturing the quantitative importance of the relatively constant revenue share of nonemployers observed in the data.