American Economic Review 2021, 111(2): 547–579
https://doi.org/10.1257/aer.20190748

The Nature of Firm Growth[†]

By Vincent Sterk ⓡ Petr Sedláček ⓡ Benjamin Pugsley*

About one-half of all startups fail within five years, and those that survive grow at vastly different speeds. Using Census microdata, we estimate that most of these differences are determined by ex ante heterogeneity rather than persistent ex post shocks. Embedding such heterogeneity in a firm dynamics model shows that the presence of ex ante heterogeneity (i) is a key determinant of the firm size distribution and firm dynamics, (ii) can strongly affect the macroeconomic effects of firm-level frictions, and (iii) helps understand the recently documented decline in business dynamism by showing a disappearance of high-growth startups ("gazelles") since the mid-1980s. (*JEL* D22, D24, E24, J23, L11, M13)

There are enormous differences across firms. On the one hand, many startups fail within the first year and most of those that survive do not grow. On the other hand, a small fraction of high-growth startups, so-called "gazelles," makes lasting contributions to aggregate job creation and productivity growth (see, e.g., Haltiwanger et al. 2016). While firm dynamics have long been recognized as a key determinant of macroeconomic outcomes, little is known about why firm performance is so different or how the nature of firm growth affects macroeconomic outcomes.

One view in the literature is that, following entry, firms are hit by ex post shocks to productivity or demand: some startups are lucky and grow into large firms. An alternative view is that there are ex ante differences across firms: some types of startups are poised for growth, for example due to a highly scalable technology or business idea.[1]

*Sterk: University College London and CEPR (email: v.sterk@ucl.ac.uk); Sedláček: University of Oxford and CEPR (email: petr.sedlacek@economics.ox.ac.uk); Pugsley: University of Notre Dame (email: bpugsley@nd.edu). Emi Nakamura was the coeditor for this article. First version December 2017. We thank Costas Arkolakis, Eric Bartelsman, Christian Bayer, Mark Bils, Richard Blundell, Vasco Carvalho, Steven Davis, Pablo D'Erasmo, Chris Edmond, Doireann Fitzgerald, Xavier Gabaix, Urban Jermann, Greg Kaplan, Pawel Krowlikowski, Erzo Luttmer, Ezra Oberfield, Vasia Panousi, Fabien Postel-Vinay, Michael Siemer, Kjetil Storesletten, Emily and Robert Swift as well as three anonymous referees, and participants of numerous seminar and conference presentations for their helpful comments. We also are grateful for expert research assistance by Harry Wheeler and Lan Dinh. Pugsley gratefully acknowledges financial support from the Ewing Marion Kauffman Foundation. Sedláček gratefully acknowledges financial support from the European Commission: this project has received funding from the European Research Council [grant number 802145]. Any opinions and conclusions expressed herein are solely those of the authors and do not necessarily represent the views of the US Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

[†]Go to https://doi.org/10.1257/aer.20190748 to visit the article page for additional materials and author disclosure statements.

[1]Another important dimension of heterogeneity, on which we do not focus in this paper, relates to the role of supply versus demand factors. For evidence on this, see, e.g., Hottman, Redding, and Weinstein (2016) and Foster, Haltiwanger, and Syverson (2016).

In this paper, we provide empirical evidence on the relative importance of ex ante and ex post heterogeneity in shaping firms' growth paths. We then bring this evidence to a structural firm dynamics model and show that the precise nature of firm growth has strong implications for the macroeconomy and the way in which it is affected by firm-level frictions. Since Hopenhayn and Rogerson (1993), a growing literature uses quantitative heterogeneous-firms models to evaluate the micro- and macroeconomic effects of policies and/or frictions. Our results demonstrate the importance of accounting carefully not only for the amount of heterogeneity across firms, but also for its transience and for the moment of its inception, i.e., before or after startup.

To establish these results, we make use of the Longitudinal Business Database (LBD), an administrative panel covering nearly all private employers in the United States from 1976 to 2012. Our central piece of empirical evidence is the cross-sectional autocovariance function of business-level employment by age. We thereby take inspiration from the earnings dynamics literature, which has long recognized that autocovariances help to distinguish shocks from deterministic profiles (see, e.g., MaCurdy 1982; Abowd and Card 1989; Guvenen 2009; Guvenen and Smith 2014). To the best of our knowledge, even the basic autocovariance structure of employment by age has not been systematically documented in the firm dynamics literature which, instead, has focused on the age profiles of average size and exit.[2]

We begin the analysis using a reduced-form statistical model of firm-level employment, which allows for the possibility that differences across businesses are a result of both ex ante heterogeneous growth profiles and ex post shocks. A major benefit of the statistical model is its simplicity, yielding analytical formulas which help to understand the identification of the key parameters. In particular, it makes clear that crucial information about the extent of ex ante heterogeneity across firms is contained in the long-horizon autocovariances of firm-level employment.

Estimation of the statistical model on the autocovariance matrix reveals a key finding of our study: ex ante heterogeneity accounts for a large share of the cross-sectional dispersion in employment. In the first year after entry, ex ante heterogeneity accounts for more than 90 percent of the cross-sectional dispersion in employment. More importantly, even after *twenty* years, ex ante factors still explain about 40 percent of the cohort's employment dispersion. This finding is consistent with empirical evidence that certain observable characteristics at the time of startup can partly predict firm growth, see Guzman and Stern (2015, 2019, 2020), and Belenzon, Chatterji, and Daley (2017). Beyond its value summarizing the importance of ex ante heterogeneity for observed employment dynamics, our statistical model is easily adapted to the driving process of a structural model.

Next, we take the data to a full-blown structural macroeconomic model with firm dynamics in order to answer other important questions which the statistical model cannot address. The structural model follows the tradition of Hopenhayn (1992), Melitz (2003), and Luttmer (2007), and features endogenous entry, exit, and general equilibrium forces. Following the statistical model, we introduce a multidimensional

[2] See, e.g., Haltiwanger, Jarmin, and Miranda (2013); Hsieh and Klenow (2014); and Akcigit, Alp, and Peters (2017). Cabral and Mata (2003) also document the evolution of the skewness of the size distribution with age.

idiosyncratic process into this framework, which allows not only for persistent and transitory ex post shocks, but also for heterogeneity in ex ante growth and survival profiles. We demonstrate that a combination of ex ante heterogeneity and ex post shocks is in fact necessary to obtain a good fit with the empirical autocovariance structure.

While our baseline model contains no explicit frictions, we also consider a version with imperfect information, in the spirit of Jovanovic (1982), in which ex ante heterogeneity is disentangled from ex post shocks only gradually. In addition, we consider a version in which firms endogenously invest into demand accumulation, subject to adjustment costs. Although these extensions could in principle offer a different perspective on the empirical patterns, this turns out not to be the case. In particular, ex ante differences still emerge as the key source of heterogeneity.

We estimate the model by matching not only the autocovariance function of employment at the firm level, but also the average size and exit profiles, conditional on age. We then use the structural model for three purposes: (i) to revisit the results from the reduced-form model while accounting for endogenous selection, and to extend these results to other outcomes such as exit rates and the average firm growth profile; (ii) to understand how the presence of rich ex ante heterogeneity can change the macroeconomic effects of micro-level frictions; and (iii) to understand how the nature of firm growth has changed during recent decades and what have been the macroeconomic implications.

First, the model suggests that ex ante heterogeneity is not only an important determinant of the dispersion in firm-level employment, but also of firm exit and growth. That is, the fact that many young firms shut down while surviving businesses grow quickly is in large part driven by ex ante heterogeneity. Moreover, we find that "gazelles," a small fraction of startups with exceptional ex ante growth potential, account for a large share of average firm growth.

Second, we present model exercises to explore how the presence of ex ante heterogeneity can affect the impact of micro-level distortions on the aggregate economy. We do so by introducing micro-level frictions into the baseline economy and contrasting it with the same exercise conducted in a model with a restricted shock process. In particular, this restricted model features no permanent ex ante heterogeneity, but is conventional in the literature, see for instance Hopenhayn and Rogerson (1993).

In the main text, we consider two examples of micro-level distortions: nonconvex adjustment costs on demand accumulation and financial frictions forcing firms to shut down if they fail to meet a borrowing limit. We find that these frictions have quantitatively very different effects in the two versions of the model. This is primarily due to the fact that, while the two economies have almost identical firm size distributions, the baseline economy has a wider dispersion of firm values owing to the presence of permanent ex ante differences across firms.

As a result, the adjustment costs have a much smaller effect in the baseline model than in the restricted version. Intuitively, the higher dispersion of firm values implies that there are fewer "marginal" firms which are indifferent between adjusting or not. By contrast, the effects of financial frictions, which indiscriminately force firms to exit whenever they cannot meet the borrowing limit, are larger in the baseline model. The friction is particularly damaging in the baseline model as it eliminates young

firms with low current profitability but high long-run potential. In the restricted model, on the other hand, all firms have the same long-run potential and therefore indiscriminate exit is less harmful. These examples highlight that while the presence of ex ante heterogeneity does not always change the impact of micro-level distortions on aggregate outcomes in the same direction, the precise nature of firm growth may be crucially important for quantitative analysis.[3]

Third and finally, we use the model to understand how the nature of firm growth has changed over time and whether any such changes can shed light on the observed decline in business dynamism in the US economy. Specifically, we re-estimate the model on two subsamples, splitting our data in half. The results suggest that the prevalence of ex ante high-growth firms, i.e., gazelles, has substantially declined among the population of startups in the late sample compared to earlier years. In addition, we find that gazelles that do start up in the late sample do not grow as rapidly as their counterparts in the early sample. These findings provide a new angle to discussions on declining "dynamism" of US businesses, see, e.g., Decker et al. (2016). They also relate to Sedláček and Sterk (2017), who document strong cohort effects in firm-level employment. The latter focus on cyclical variations in entry conditions, whereas the change considered here appears permanent.

A major advantage of the Census data used in this paper is that they span the population of employers and therefore speak simultaneously to the micro and the macro level. An important next step is to investigate empirically what determines the ex ante and ex post differences documented in this paper and use this information to further endogenize firm dynamics in structural models. This, however, will require very different data sources with richer micro information relating to, e.g., entrepreneurial skills, business plans, financial characteristics, or the organizational structures of firms. Existing studies along these lines include, in addition to references above, Abbring and Campbell (2005), who study bars in Texas; and Campbell and De Nardi (2009) and Hurst and Pugsley (2011), who present survey evidence that many nascent entrepreneurs do not expect their business to grow large.[4] Our results show that the heterogeneity documented in these studies has important implications at the macro level.

The remainder of this paper is organized as follows. Section I presents the data, the reduced-form statistical model, and initial estimates of the importance of ex ante heterogeneity for size dispersion. Section II describes the structural firm dynamics model and revisits the importance of ex ante heterogeneity. Sections III and IV present results on, respectively, macroeconomic implications and changes in the nature of firm growth over time. Finally, Section V concludes.

I. Evidence from a Statistical Model

This section takes the first step in analyzing the importance of ex ante heterogeneity in driving observed differences in employment across firms.[5] Using a statistical

[3] Similar results for several other firm-level distortions are presented in the online Appendix.

[4] Schoar (2010) makes a distinction between "subsistence" and "transformational" entrepreneurship in this regard.

[5] See DeBacker, Panousi, and Ramnath (2018) for an analysis of household income risk from owning non-corporate private businesses using tax data and Gourio (2008) for an analysis of income risk for publicly held firms using investment data in Compustat.

model, we estimate the extent to which cross-sectional variation in employment is driven by ex ante heterogeneity and to what extent it results from ex post shocks. We begin by describing our dataset and the central piece of empirical evidence used in the estimation: the autocovariance function of log employment at the firm level. The simplicity of the statistical model allows us to show analytically how all the relevant model parameters map into the autocovariance function, shedding light on the identification of ex ante versus ex post heterogeneity. Moreover, the statistical model is a special case of the structural model, which we discuss further in Section II.

A. *Data*

The analysis is based on administrative microdata on employment in the United States, taken from the Census Longitudinal Business Database (LBD). These annual data cover almost the entire population of employers over the period between 1979 and 2012. We construct a panel of log employment at the firm level in the year of startup (age 0) up to age 19.[6] Prior to the analysis, we take out a fixed effect for the birth year of the establishment (or firm) and for its industry classification at the 6-digit level.[7] Throughout, unless stated otherwise, all results use data from the LBD.

The main text reports results only for firms and using data from the LBD, unless explicitly stated otherwise. Online Appendix Section G shows that our results also hold for establishments, for which ex ante heterogeneity is even slightly more important than for firms.[8,9]

B. *The Autocovariance Structure of Employment*

Figure 1 presents our main piece of empirical evidence: the cross-sectional autocovariance structure of log employment, conditional on age a. In order to understand this structure more easily, we present the autocovariances in terms of standard deviations (left panel) and autocorrelations (right panel). Since firms may exit at any age, we display patterns for a balanced panel (solid line) that includes only firms that survive for at least 20 years and for an unbalanced panel (dashed line) that includes all firms in our dataset. Interestingly, firm exit affects essentially only the cross-sectional employment dispersion by age; the autocorrelations are remarkably similar across the balanced and unbalanced panels.

[6] Employment is measured annually at the establishment level for the pay period including March 12. Establishments are physical locations, and a firm can consist of one or more establishments. The age of an establishment is measured from the year it first reports employment. Firm age is initialized as the age of its oldest establishment and the firm ages naturally thereafter.

[7] See online Appendix Section A for details on the panel construction and autocovariance estimation.

[8] Bento and Restuccia (2019) have recently drawn attention to non-employer firms, in the context of the discussion on business dynamism. While our data do not include non-employers, one might expect that the importance of ex ante heterogeneity would be even larger if such firms were included in our sample, to the extent that they have zero employees throughout their lives.

[9] In unreported results, we have also examined firm sales, which are available for a subset of firms starting in 1996. For these overlapping years and a shorter horizon, we find results for sales to be similar to those for employment. We also find that our main results are not sensitive to excluding large entrants, nor using an industry $\times$ cohort interacted fixed effect when residualizing log employment in place of separate fixed effects.

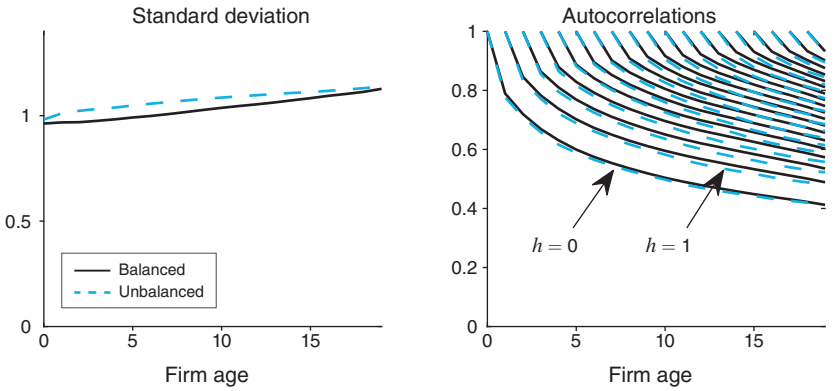

FIGURE 1. STANDARD DEVIATIONS AND AUTOCORRELATIONS OF LOG EMPLOYMENT BY AGE

Notes: The left panel shows cross-sectional standard deviations of log employment by age (a). The right panel shows cross-sectional correlations of log employment between ages a and age $h \leq a$. *Balanced* refers to a panel of firms which survived at least up to age 19, while *Unbalanced* refers to a panel of all firms.

Let us first focus on the cross-sectional standard deviations by age, shown in the left panel. Standard deviations are between 0.9 and 1.3, indicating large size differences even at young ages. Also, the cross-sectional dispersion generally increases with age and this is true for both the balanced and unbalanced panel. The latter indicates that the observed increase in size dispersion with age is not purely driven by selective exit of certain firms.

The right panel of Figure 1 depicts the cross-sectional correlations of logged employment between age a and an earlier age $h \leq a$. Keeping h fixed, the autocorrelations decline with age a. For instance, while the autocorrelation between logged employment at ages 0 and 10 is 0.50, the autocorrelation between ages 0 and 19 is 0.41. Importantly, the long-horizon autocorrelations appear to stabilize at relatively high levels.

On the other hand, for a fixed lag length $a - h$, the autocorrelations are increasing in age. For instance, the correlation of log employment between age 0 and age 9 is 0.52, whereas the corresponding correlation between age 10 and 19 is 0.73. These empirical patterns contain key information on the relative importance of ex ante heterogeneity and ex post shocks, as we will discuss below in detail.

C. *Employment Process*

To understand what we can learn from the autocovariances about the importance of ex ante versus ex post heterogeneity, we now consider a statistical model of employment which allows for both sources of heterogeneity. The statistical model abstracts from endogenous exit, but the model in Section II will explicitly incorporate these aspects. The process nests as special cases the shock processes considered in several prominent structural firm dynamics models, such as those of Hopenhayn and Rogerson (1993) and Melitz (2003), while at the same time it is flexible enough to fit the observed autocovariance structure well. Online Appendix Section B estimates several alternative model specifications (including conventional panel data

models) showing that our specification strikes a balance between model fit and parsimony.

Our baseline employment process features deterministic "ex ante" profile heterogeneity and "ex post" shocks. Let $n_{i,a}$ be the employment level of an individual firm i at age a and consider the following process for this variable:

$$\ln n_{i,a} = \underbrace{u_{i,a} + v_{i,a}}_{\text{ex ante component}} + \underbrace{w_{i,a} + z_{i,a}}_{\text{ex post component}}, \tag{1}$$

where

$$\begin{aligned}
&u_{i,a} = \rho_u u_{i,a-1} + \theta_i, \quad && u_{i,-1} \sim iid\left(\mu_{\tilde{u}}, \sigma^2_{\tilde{u}}\right), \quad && \theta_i \sim iid\left(\mu_\theta, \sigma^2_\theta\right), \quad && |\rho_u| \leq 1,\\
&v_{i,a} = \rho_v v_{i,a-1}, && v_{i,-1} \sim iid\left(\mu_{\tilde{v}}, \sigma^2_{\tilde{v}}\right), && && |\rho_v| \leq 1,\\
&w_{i,a} = \rho_w w_{i,a-1} + \varepsilon_{i,a}, && w_{i,-1} = 0, && \varepsilon_{i,a} \sim iid\left(0, \sigma^2_\varepsilon\right), && |\rho_w| \leq 1,\\
&z_{i,a} \sim iid\left(0, \sigma^2_z\right). &&&&&&
\end{aligned}$$

Here, all shocks are drawn from distributions which are i.i.d. across time and across firms, and we let μ denote a mean and σ^2 a variance.

In the process above, $\ln n^{EXA}_{i,a} = u_{i,a} + v_{i,a}$ captures the *ex ante* component of employment, where $u_{i,a}$ is a permanent part which converges to a certain level as the firm ages, and $v_{i,a}$ is a transitory part which converges to zero. Note that both parts come with their own persistence parameter, ρ_u and ρ_v, respectively, which are common across firms.

The ex ante component is governed by three firm-specific constants, which are random and drawn independently just prior to startup, i.e., at age $a = -1$. The constant θ_i determines the firm-specific long-run level of the ex ante component. The second and third constant, $u_{i,-1}$ and $v_{i,-1}$, represent two firm-specific initial conditions, corresponding to, respectively, the permanent and the transitory part of the ex ante component.

Note that this relatively parsimonious specification allows for rich heterogeneity in ex ante profiles. In particular, if $|\rho_u| < 1$ then the ex ante component converges to a long-run "steady state" level of $\ln n^{EXA}_{i,\infty} = \theta_i/(1 - \rho_u)$. Since this level differs across firms, the process admits heterogeneity in long-run steady-state employment. Moreover, since initial conditions differ across firms, we allow for heterogeneity in the paths from initial employment toward the steady states. Finally, since the process includes two separate initial conditions, each with its own degree of persistence, the process allows firms to gravitate toward their steady-state levels at different speeds.[10]

The ex post shocks enter the model via a second component, $\ln n^{EXP}_{i,a} = w_{i,a} + z_{i,a}$. The process for the ex post component is constructed such that its expected profile is

[10] By not restricting ρ_u and ρ_v to lay strictly inside the unit circle, we allow in principle for unit roots in the u and v terms. In this case, rather than an ex ante profile toward some expected long-run size, the ex ante terms would instead characterize heterogeneous growth rates from some initial size.

flat and zero so that it does not capture any of the heterogeneity in ex ante profiles. Specifically, $w_{i,a}$ captures persistent ex post shocks, and is modeled as an autoregressive process of order one, with i.i.d. innovations given by $\varepsilon_{i,a}$ and a persistence parameter denoted by $|\rho_w| \leq 1$. Notice that this formulation allows $w_{i,a}$ to follow a random walk, in which case each $\varepsilon_{i,a}$ may be interpreted as a growth rate shock. Because the u and v terms are meant to capture the entire ex ante profile, we normalize the initial condition of the persistent ex post shocks to $w_{i,-1} = 0$.

As described earlier, the process above nests various specifications commonly used in the firm dynamics literature to model firm-level shocks. For example, Hopenhayn and Rogerson (1993) assume an AR(1) for firm-level productivity, with a common constant across firms and heterogeneous initial draws. In their baseline model without distortions, the firm-level shocks map one-for-one into employment. We obtain their specification by setting $\rho_u = \rho_v = \rho_w$ and fixing $\theta_i = \mu_\theta$ and $u_{i,-1} = z_{i,a} = 0$, so $\sigma_\theta = \sigma_{\tilde{u}} = \sigma_z = 0$. By contrast, Melitz (2003) allows, like us, for heterogeneity in steady-state levels, but abstracts from ex post shocks and assumes that steady states are immediately reached. We obtain his process by setting $\rho_u = 0$ and $u_{i,-1} = v_{i,-1} = z_{i,a} = \varepsilon_{i,a} = 0$, which implies that $\ln n_{i,a} = \theta_i$ at any age. Similarly, we obtain the dynamics in Bartelsman, Haltiwanger, and Scarpetta (2013) under the same restrictions, but allowing for $z_{i,a} \neq 0$ with $\sigma_z > 0$.[11] Our baseline process also aligns with models with richer heterogeneity in ex ante profiles and/or ex post shocks, as proposed by for example Luttmer (2011) and Arkolakis (2016) and Arkolakis, Papageorgiou, and Timoshenko (2018).[12]

D. *Estimation Strategy and Results*

In what follows, we first discuss several key properties of the model-implied autocovariance function. Next, we present the estimation results and show how our baseline model fits the data. Finally, we provide intuition about the identification of the model parameters and how each of the model components maps into the empirical patterns.

Properties of the Autocovariance Function.—To explain our empirical strategy, we first demonstrate the usefulness of the autocovariance matrix in quantifying the role of ex ante versus ex post heterogeneity. All key parameters of the statistical model can be identified from the autocovariance matrix. For any pair of ages, the model-implied cross-sectional covariance of employment can be written as closed-form expression of the model parameters. The covariance of employment

[11] Our process also nests specifications commonly assumed in the econometrics literature on dynamic panel data models, see for example Arellano and Bond (1991). This literature typically assumes an autoregressive process, like Hopenhayn and Rogerson (1993), but allows for heterogeneity in the constant θ_i and thus in steady-state levels. Commonly, however, θ_i is differenced out and hence no estimate is provided for σ_θ, a key parameter in our analysis. Moreover, the panel data econometrics literature commonly assumes that $\rho_u = \rho_v = \rho_w$. In our application, it turns out that this assumption is too restrictive to provide a good fit of the observed autocovariance matrix (see online Appendix Section B). Our results thus caution against the use of standard panel data estimators when applied to employment dynamics of young firms.

[12] For further discussion, please refer to online Appendix Section B where we consider a number of alternative statistical models both as special cases and further generalizations of equation (1).

of a firm at age a and at age $h = a - j$, where $0 \leq j \leq a$ is the lag length, can be expressed as

$$(2)\quad \operatorname{cov}\left[\ln n_{i,a}, \ln n_{i,a-j}\right] = \underbrace{\left(\sum_{k=0}^{a} \rho_u^k\right)\left(\sum_{k=0}^{a-j} \rho_u^k\right)\sigma_\theta^2 + \rho_u^{2(a+1)-j}\sigma_{\tilde{u}}^2 + \rho_v^{2(a+1)-j}\sigma_{\tilde{v}}^2}_{\text{ex ante components}}$$
$$+ \underbrace{\sigma_\varepsilon^2 \rho^j \sum_{k=0}^{a-j} \rho_w^{2k} + \sigma_z^2 \mathbf{1}_{j=0}}_{\text{ex post components}}.$$

This result is derived in online Appendix Section B. The autocovariance function is a nonlinear function of the persistence and variance parameters of the components of the underlying process.[13] We can estimate the parameters of this process by matching the model's autocovariance structure to its empirical counterpart.

To understand the identification, it is useful to consider the case where ρ_u, ρ_v, and ρ_w are strictly inside the unit circle so that the process is covariance stationary in the long run. Then, at an infinite lag length, i.e., letting the age a approach infinity keeping the initial age $h = a - j$ fixed, the autocovariance is $\lim_{a\to\infty} \operatorname{cov}[\ln n_{i,a}, \ln n_{i,h}] = \left((1 - \rho_u^{h+1})/(1 - \rho_u)^2\right)\sigma_\theta^2$. When σ_θ equals zero, i.e., when there is no heterogeneity in steady-state levels, the autocovariance is zero. Thus, long-horizon autocovariances contain valuable information on the presence of ex ante heterogeneity in steady-state levels. In Figure 1, autocorrelations appear to stabilize at long lag lengths, i.e., at high levels of a given $h = a - j$, suggesting that such heterogeneity is indeed a feature of the data. More intuition on the identification of model parameters is presented below.

Parameter Estimates and Model Fit.—We formally estimate the eight parameters of the process using a minimum distance procedure, as proposed by Chamberlain (1984). Specifically, we minimize the sum of squared deviations of the 210 covariance moments from the upper triangular parts of the autocovariance matrix implied by the process, from its counterpart in the data. Because the size of the LBD ensures that each element of the empirical autocovariance matrix is precisely estimated, we use an identity weighting matrix in the estimation procedure. Throughout, our results apply to the balanced panel dataset, although they are similar using the unbalanced panel.[14]

Figure 2 shows the autocovariance structure in the data and in the estimated model. The figure has the same structure as the right panel of Figure 1, but plots autocovariances rather than autocorrelations (thus combining the information of the two panels in Figure 1). For instance, the bottom solid line shows the autocovariance of employment at a certain age a with employment at age 0. Figure 2 shows that model fit is very good, correctly capturing the convexly declining pattern of the

[13] Note that the mean parameters μ_θ, $\mu_{\tilde{u}}$, and $\mu_{\tilde{v}}$ are not identified by the autocovariance function. These parameters, however, are also not needed to quantify the relative importance of ex ante versus ex post heterogeneity.

[14] For brevity, we defer a detailed discussion of the estimation procedure to online Appendix Section B. There, we also include the estimated parameters using the unbalanced panel.

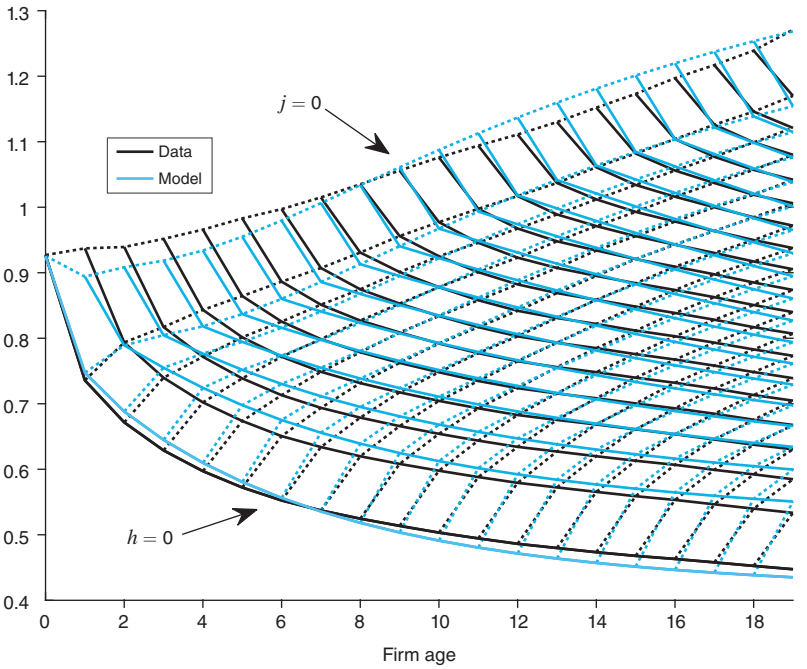

FIGURE 2. AUTOCOVARIANCE MATRICES: STATISTICAL MODELS VERSUS DATA

Notes: Cross-sectional covariance of log employment between age $a = h + j$ and age $h \leq a$ in the data, and in the baseline model. Results are shown for firms, using the balanced panel.

autocovariances in the lag length, given the initial age h, and the concavely increasing pattern in age given the lag length $j > 0$.

The corresponding parameter estimates are shown in Table 1. A key feature of our baseline process is the presence of dispersion in long-run steady states, governed by σ_θ and ρ_u. The point estimates imply a standard deviation of long-run steady-state employment levels of 0.71 log points. This value is substantial when considering that the overall cross-sectional dispersion of 20-year-old firms is about 1.3 log points (see Figure 1). Note also that the data reject the presence of a unit root process, in our sample. Such violations of Gibrat's law have been documented in the literature, in particular among younger firms, see, e.g., Haltiwanger, Jarmin, and Miranda (2013).

Mapping Model Components to the Data.—We now discuss in more detail the role of each of the model's components in generating the shape of the autocovariance function necessary to match the data. This will also provide further intuition about how the model parameters are identified by the information contained in the autocovariance matrix. We do so by estimating four restricted versions of our baseline model and considering their empirical fit, depicted in Figure 3 (see the figure's note for the specific restrictions imposed).

Restricted models I and II (top row) illustrate, respectively, why a combination of permanent ex ante heterogeneity and ex post shocks is needed to match the data. Model I is a popular specification in the firm dynamics literature, which essentially

TABLE 1—PARAMETER ESTIMATES FROM REDUCED-FORM MODEL

ρ_u	ρ_v	ρ_w	σ_θ	$\sigma_{\tilde{u}}$	$\sigma_{\tilde{v}}$	σ_ϵ	σ_z
0.218	0.832	0.963	0.555	1.743	0.695	0.255	0.272
(0.002)	(0.001)	(0.001)	(0.002)	(0.015)	(0.002)	(0.001)	(0.001)

Notes: Equally weighted minimum distance estimates of equation (2) for firms, using the balanced panel. See online Appendix Section B for estimates using the unbalanced panel and online Appendix Section G for results on establishments.

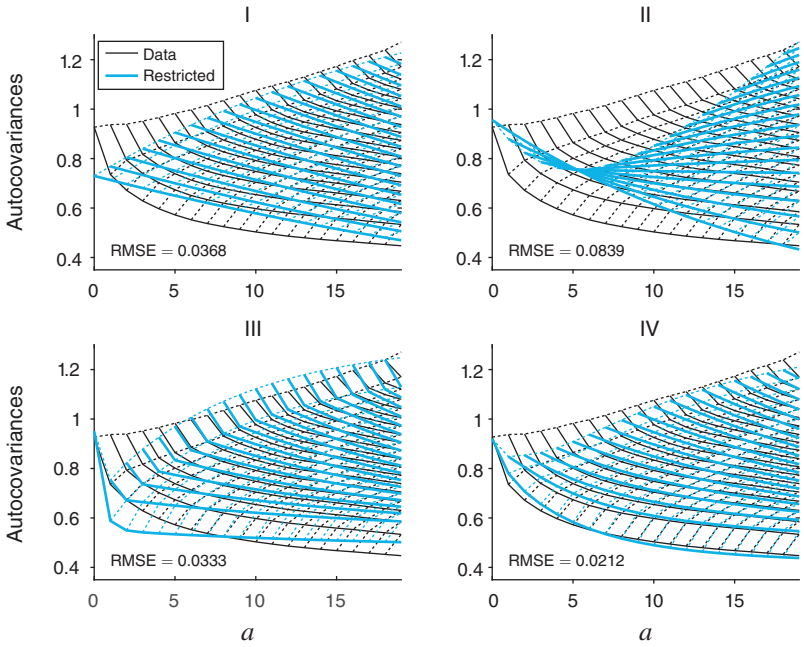

FIGURE 3. AUTOCOVARIANCE MATRICES: RESTRICTED MODELS

Notes: Cross-sectional covariance of log employment between age $a = h + j$ and age $h \leq a$ in the baseline and in the four restricted models (for the balanced firm panel estimates). In Model I ρ_w, σ_ε, and $\sigma_{\tilde{v}}$ are estimated, while imposing $\rho_u = \rho_v = \rho_w$ and $\sigma_\theta = \sigma_{\tilde{u}} = \sigma_z = 0$. In Model II ρ_u, σ_θ, and $\sigma_{\tilde{u}}$ are estimated, while imposing $\rho_w = \rho_v = \sigma_\varepsilon = \sigma_{\tilde{v}} = \sigma_z = 0$. Model III is the baseline with the restriction that $\rho_v = \sigma_{\tilde{v}} = 0$. Model IV is the baseline with the restriction that $\sigma_z = 0$. The figure also shows RMSE values for each of the restricted models, that of the baseline is RMSE $= 0.012$.

amounts to an AR(1) process with heterogeneous initial draws, but without heterogeneity in long-run steady states.[15] With this restriction, the model-implied autocovariances are almost linear in age, which conflicts with the nonlinear patterns in the data. The presence of ex ante heterogeneity thus relaxes the need for persistence coefficients which are very close to one.[16] In restricted Model II, we shut down all

[15] For example, Hopenhayn and Rogerson (1993) consider this process for productivity.

[16] Recent work by Gabaix (2009) and Luttmer (2011) suggests that in order to generate a power-law ergodic firm size distribution that is close to the data, a combination of permanent and persistent shocks may be necessary. This points to a potential trade-off between matching early life-cycle dynamics, as summarized by our autocovariance function, and long-run patterns such as the ergodic firm-size distribution. See online Appendix Section B for estimation results from a variant of our model with a unit root.

ex post shocks, allowing only for heterogeneous ex ante profiles (with only one initial condition). This version fails to match the monotone increase in dispersion with firm age seen in the data.

Restricted Model III illustrates why *both* parts of the ex ante component, u and v, are required to match the data. This version is the same as our baseline except that we shut down the transitory part v, and we re-estimate the remaining parameters. The presence of v enables the model to match the curvature of the autocovariance function, as it allows for different speeds of convergence to the long-run steady state employment levels. Finally, restricted Model IV shuts down the i.i.d. ex post shock z. It becomes clear that the presence of this shock somewhat improves the fit, by giving an extra kick to the dispersion of employment across firms, in line with the data, but without distorting the higher-order autocovariances.

While our baseline model provides a very good fit to the data, we estimate several extensions and alternatives in online Appendix Section B. These include, e.g., a generalized AR(1) process with a unit root similar to specifications in Gabaix (2009) or Luttmer (2011), an AR process with age-dependent dispersion of ex post shocks, and several dynamic panel data models akin to models in Arellano and Bond (1991), including a panel AR(2) model similar to the specification in Lee and Mukoyama (2015). Importantly, none of the alternatives improves on model fit without introducing more parameters, and our conclusions about the importance of ex ante heterogeneity remain unchanged across specifications.

E. *The Importance of Ex Ante and Ex Post Heterogeneity*

With the estimated model in hand, we can quantify the relative importance of ex ante profiles and ex post shocks for the cross-sectional dispersion in employment. This is done based on equation (2). With the lag length j set to zero, this equation provides a decomposition of the variance of size (log employment), at any given age a, into the contributions of the ex ante and ex post components. Figure 4 plots the fraction of the total variance that is accounted for by the ex ante component. Thick lines denote the age groups used in the estimation, i.e., age 0 to 19, whereas thin lines represent an extrapolation for firms at age 20 or above using the point estimates.[17]

The left panel of Figure 4 shows that for firms in the year of startup (age 0) the ex ante component accounts for about 85 percent of the cross-sectional variance in firm size. The remainder is due to ex post shocks that materialized in the first year. Considering older age groups, the contribution of ex ante heterogeneity declines, but remains high. At age 20, ex ante factors account for around 40 percent of the size variance among firms. In the data, more than 70 percent of the firms are 20 years old or younger. Our results show that, among these firms, ex ante factors are a key determinant of size. Increasing age toward infinity, the contribution of ex ante heterogeneity stabilizes at around 40 percent. Therefore, even among very old firms ex ante factors contribute to a large chunk of the dispersion in size.

[17] We have also computed confidence bands for this decomposition, but these are extremely narrow due to the very large number of data points used in the estimation and the resulting high precision of our point estimates.

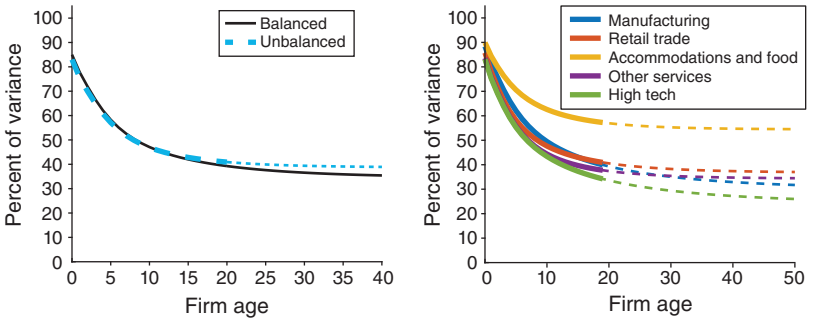

FIGURE 4. CONTRIBUTION OF EX ANTE HETEROGENEITY TO CROSS-SECTIONAL EMPLOYMENT DISPERSION

Notes: Contribution of the ex ante component, $\ln n_{i,a}^{EXA}$, to the cross-sectional variance of log employment, by age. Thin lines denote age groups not directly used in the estimation. The decomposition is based on equation (2) with $j = 0$. Left panel: economy wide. Right panel: within sectors.

F. *Results by Sector*

While our primary analysis controls for industry fixed effects to examine an average industry, in this subsection we present results for a selected group of sectors. We find that our results hold also *within* sectors.[18] Specifically, we re-estimate the employment process based on the empirical autocovariance structure by sector and compute again the contribution of ex ante factors to the cross-sectional size dispersion by age. It turns out that the specified employment process fits the data well also at the sectoral level, with the Root Mean Squared Error (RMSE) varying between 0.01 (e.g., retail trade) and 0.02 (e.g., high-tech). For comparison the RMSE for the economy-wide analysis is 0.012.

The right panel of Figure 4 shows the employment variance decomposition, for a number of sectors. While there is some variation across sectors, ex ante heterogeneity broadly emerges as a dominant source of heterogeneity also within sectors. In online Appendix Section C, we report results for additional sectors and obtain the same findings. These results suggest that ex ante heterogeneity is not only relevant for macroeconomic studies of firm dynamics, but also for industry-specific analysis. In the remainder of this paper, however, we will focus on the economy-wide data.

II. Structural Model

To learn about the implications of our findings for the aggregate economy, in this section we estimate a structural macroeconomic model with firm dynamics. This framework has several advantages relative to the statistical model in Section I. First, the structural model accounts for selective entry and exit. Second, the structural model allows us to compute aggregates. Third, micro-founding firm decisions allows us to analyze how various frictions (e.g., imperfect information, adjustment costs, or financial frictions) affect the observed patterns in the data.

[18] The results reported here are for balanced panels of firms. Results for unbalanced panels and for establishments can be found in online Appendix Section C. We also estimate the model for the high-tech sector that spans multiple NAICS sectors.

In the remainder of the paper, we use the estimated structural model for three distinct purposes. First, we revisit and extend our previous results regarding the importance of ex ante heterogeneity for firm-level performance (this section). Second, we show that the presence of ex ante heterogeneity in growth profiles can dramatically change the impact of distortions at the firm level on the macro economy (Section III). Third, we use our framework to provide new insights on the timing and sources of the decline in business dynamism observed over the past decades (Section IV).[19]

A. *The Model*

We consider a closed general equilibrium economy with heterogeneous firms and endogenous entry and exit, as in Hopenhayn and Rogerson (1993). Following Melitz (2003) and others, each firm is monopolistically competitive and faces a demand schedule which is downward-sloping in its price. To model heterogeneity across firms, we embed an idiosyncratic process with the same structure as in Section I, thereby allowing for differences in both ex ante profiles and ex post shocks.

Households.—The economy is populated by an infinitely lived representative household who owns the firms and supplies a fixed amount of labor in each period, denoted by $\bar{N}$. Household preferences are given by $\sum_{t=0}^{\infty} \beta^t C_t$, where $\beta \in (0,1)$ is the discount factor. The term C_t is a Dixit-Stiglitz basket of differentiated goods given by

$$C_t = \left(\int_{i \in \Omega_t} \varphi_{i,t}^{\frac{1}{\eta}} \, c_{i,t}^{\frac{\eta-1}{\eta}} \, di \right)^{\frac{\eta}{\eta-1}},$$

where Ω_t is the measure of goods available in period t, $c_{i,t}$ denotes consumption of good i, η is the elasticity of substitution between goods, and $\varphi_{i,t} \in [0, \infty)$ is a stochastic and time-varying demand fundamental specific to good i. We consider a stationary economy from now on and simplify notation by dropping time subscripts. Note, however, that variables with an i subscript will still vary over time because of shocks to the good-specific demand fundamental.

The household's budget constraint is given by $\int_{i \in \Omega} p_i c_i di = W\bar{N} + \Pi$, where p_i denotes the price of good i, W denotes the nominal wage, and Π denotes firm profits. Utility maximization implies a demand schedule given by $c_i = \varphi_i (p_i / P)^{-\eta} C$, where P is a price index given $P \equiv (\int_{i \in \Omega} \varphi_i p_i^{1-\eta} di)^{\frac{1}{1-\eta}}$, so that total expenditure satisfies $PC = \int_{i \in \Omega} p_i c_i di$.

Incumbent Firms.—There is an endogenous measure, Ω, of incumbent firms, each of which produces a unique good. Firms are labeled by the goods they produce $i \in \Omega$. The production technology of firm i is given by $y_i + f = n_i$, where y_i is the output of the firm, n_i is the amount of labor input (employment), and f is a fixed cost of operation common to all firms, denominated in units of labor. It follows that firms

[19] Throughout the analysis we report results for firms. Estimates for establishments are shown in online Appendix Section G.

face the following profit function: $\pi_i = p_i y_i - W n_i$. Additionally, given the market structure, each firm faces a demand constraint given by

$$y_i = \varphi_i (p_i/P)^{-\eta} C, \tag{3}$$

which is the demand schedule of the household combined with anticipated clearing of goods markets, which implies $c_i = y_i$.

At the beginning of each period, a firm may be forced to exit exogenously with probability $\delta \in (0,1)$. If this does not occur, the firm has the opportunity to exit endogenously and avoid paying the fixed cost. If the firm chooses to remain in operation, it must pay the fixed cost and in turn it learns its demand fundamental φ_i. Given its production technology and demand function, the firm sets its price p_i (and implicitly y_i, n_i, and π_i) to maximize the net present value of profits. The price-setting problem is static and the firm sets prices as a constant markup over marginal costs W, i.e., $p_i = (\eta/(\eta - 1))W$.

We let labor be the numéraire so that $W = 1$, and define the real wage $w \equiv W/P$ as the price of labor in terms of the Dixit-Stiglitz consumption basket C. Using this result, we can express profits as $\pi_i = \varphi_i w^{-\eta} C \chi - f$, where $\chi \equiv ((\eta - 1)^{\eta-1}/\eta^{\eta})$, and labor demand as $n_i = \varphi_i (\eta/(\eta - 1))^{-\eta} w^{-\eta} C + f$. Note that fluctuations in the demand fundamental directly map into the firms' employment levels.

The demand fundamental φ_i is a function of an underlying exogenous Markov state vector, denoted $\mathbf{s}_i$. The value of a firm at the moment the exit decision is taken, denoted V, can now be expressed as

$$V(\mathbf{s}_i) = \max\Big\{ E\big[\pi(\mathbf{s}_i') + \beta(1 - \delta) V(\mathbf{s}_i') \,|\, \mathbf{s}_i\big], 0 \Big\}.$$

In the equation above, $\mathbf{s}_i'$ denotes the value of the state realized after the continuation decision. Accordingly, we can express the profit, output, employment, and exit policies as $\pi_i = \pi(\mathbf{s}_i')$, $y_i = y(\mathbf{s}_i')$, $n_i = n(\mathbf{s}_i')$, and $x_i = x(\mathbf{s}_i)$, respectively.

Firm Entry.—Firm entry is endogenous and requires paying an entry cost f^e, denominated in units of labor. After paying the entry cost at the beginning of a period, the firm observes its initial level of $\mathbf{s}_i$, at which point it becomes an incumbent. Note that this means that the firm will choose to exit immediately, and therefore never produce, if $V(\mathbf{s}_i) = 0$. Free entry implies the following condition:

$$wPf^e = \int V(\mathbf{s}) G(d\mathbf{s}),$$

where G is the distribution from which the initial levels of $\mathbf{s}_i$ are drawn.

Aggregation and Market Clearing.—Let $\mu(\mathbf{S})$ be the measure of producing firms in $\mathbf{S}$. Given the exit policy, $\mu(\mathbf{S})$ satisfies

$$\mu(\mathbf{S}') = \int \big[1 - x(\mathbf{s})\big] F(\mathbf{S}'|\mathbf{s}) \big[(1 - \delta)\mu(d\mathbf{s}) + M^e G(d\mathbf{s})\big],$$

where M^e denotes the measure of entrants and $F(\mathbf{S}'|s)$ is consistent with the transition law for $\mathbf{s}_i$. The total measure of active firms is given by $\Omega = \int \mu(d\mathbf{s})$. Labor

market clearing implies that total labor supply equals total labor used for production, for the fixed cost, and for the entry cost:

$$\bar{N} = \int y(\mathbf{s}')\mu(d\mathbf{s}') + \int f\left[1 - x(\mathbf{s})\right]\left[\mu(d\mathbf{s}) + M^e G(d\mathbf{s})\right] + M^e f^e.$$

Stochastic Driving Process.—In line with the reduced-form analysis we allow for the following exogenous idiosyncratic process for the demand fundamental $\varphi_{i,a}$:

$$\begin{aligned}
\ln\varphi_{i,a} &= u_{i,a} + v_{i,a} + w_{i,a} + z_{i,a}, \\
u_{i,a} &= \rho_u u_{i,a-1} + \theta_i, && u_{i,-1} \sim iid\left(\mu_{\tilde{u}}, \sigma_{\tilde{u}}^2\right), \quad \theta_i \sim iid\left(\mu_\theta, \sigma_\theta^2\right), && |\rho_u| \leq 1, \\
v_{i,a} &= \rho_v v_{i,a-1}, && v_{i,-1} \sim iid\left(\mu_{\tilde{v}}, \sigma_{\tilde{v}}^2\right), && |\rho_v| \leq 1, \\
w_{i,a} &= \rho_w w_{i,a-1} + \varepsilon_{i,a}, && w_{i,-1} = 0, \quad \varepsilon_{i,a} \sim iid\left(0, \sigma_\varepsilon^2\right), && |\rho_w| \leq 1, \\
z_{i,a} &\sim iid\left(0, \sigma_z^2\right),
\end{aligned}$$

where we momentarily (re-)introduce the age subscript a, for clarity. In addition to its permanent type θ_i, the firm-level state $\mathbf{s}_{i,a}$ is composed of the components of the demand fundamental, $u_{i,a}$, $v_{i,a}$, $w_{i,a}$, and $z_{i,a}$. The above process implies that the level of demand faced by a firm is determined by both an idiosyncratic ex ante profile, captured by $u_{i,a}$ and $v_{i,a}$, as well as ex post shocks, which enter via $w_{i,a}$ and $z_{i,a}$.

In the model, the ex ante component reflects the profile for product demand expected immediately after entry, but prior to observing any ex post shocks. In the baseline specification, we assume that the ex ante components are observable immediately after paying the entry cost, f_e. By contrast, each period's ex post demand shocks are observable only after paying the operational cost, f, in that period. Therefore, in this frictionless model employment is based on the current level of demand, while the decision to exit takes into account the entire future demand path, which depends on both ex ante and ex post factors. Later on, we will consider extensions to the model that relax the assumptions of perfect information about ex ante components as well as those of frictionless adjustment.

Relation to Statistical Model.—As briefly noted in Section I, the statistical model is a special case of our structural baseline. The two coincide when the fixed cost of operation is zero ($f = 0$). In this case, the log of firm level employment in the structural model is given by $\ln n_i = \ln\varphi_i + \ln\xi$, where φ has the same structure as in the statistical model and $\xi \equiv (\eta/(\eta - 1))^{-\eta} w^{-\eta} Y$ is a constant. Moreover, without operational costs the structural model features no endogenous firm exit as is imposed in the statistical model. It follows that the two are observationally equivalent. Accordingly, the conceptual distinction between ex ante and ex post heterogeneity in the statistical model can be understood not only from a purely statistical perspective, but also from the perspective of the firms in the structural model with $f = 0$.

B. *Parametrization and Model Fit*

We now match the model to our data for firms. Before doing so, we set three parameters a priori, assuming a model period of one year, which corresponds to the frequency of our data. First, the discount factor is set to $\beta = 0.96$, which implies an

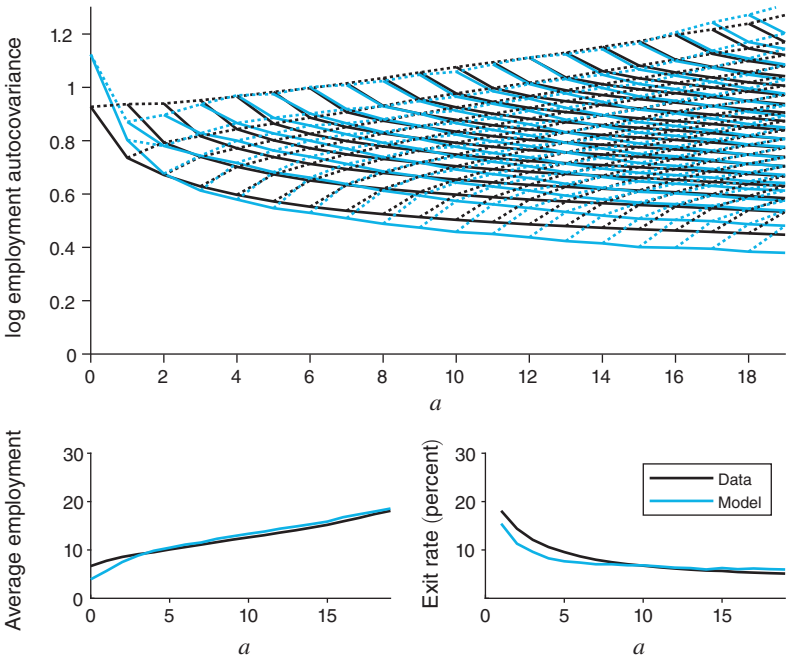

FIGURE 5. TARGETED MOMENTS: DATA AND STRUCTURAL MODEL

Notes: Top panel: autocovariances of log employment between age $a = h + j$ and age $h \leq a$ in the data and the model, for a balanced panel of firms surviving up to at least age $a = 19$. Bottom left panel: average employment by age a (unbalanced panel). Bottom right panel: exit rate by age a.

annual real interest rate of about 4 percent. Second, we set the elasticity of substitution between goods to $\eta = 6$, which is in the range of values common in the literature. Third, we set the entry cost f_e such that the ratio of the entry cost to the operational fixed cost is $f_e/f = 0.82$, following estimates of Barseghyan and DiCecio (2011).

The remaining parameters are set by matching moments in the data. Details of the numerical solution and simulation procedure are provided in online Appendix Section D.1. Again, we target the 210 covariance moments from the upper triangle of the autocovariance matrix of logged employment, by age, for a balanced panel of firms surviving up to at least age nineteen. Now, however, we also target the age profiles of the exit rate and average size (in an unbalanced panel), amounting to an additional 39 moments. In doing so, we assume that all shock innovations are drawn from normal distributions and we normalize the level parameters $\mu_{\tilde{u}}$ and $\mu_{\tilde{v}}$ to zero. In contrast to the reduced-form setup, we further assume that $\rho_v = \rho_w$, which eases the computational burden substantially because it reduces the number of state variables as firms no longer need to keep track of $w_{i,t}$ and $v_{i,t}$ separately.[20]

Figure 5 illustrates how the model fits the data. The upper panel shows the autocovariance matrix, while the lower left and right panels show the size and exit profiles by age, respectively. Overall, the model provides a good fit of the three sets of empirical moments (249 altogether), considering that the model consists of only

[20] Table 1 shows that the reduced-form estimates of these persistence parameters are close to each other. Imposing this restriction has only a small cost in model fit, increasing the RMSE from 0.0120 to 0.0171.

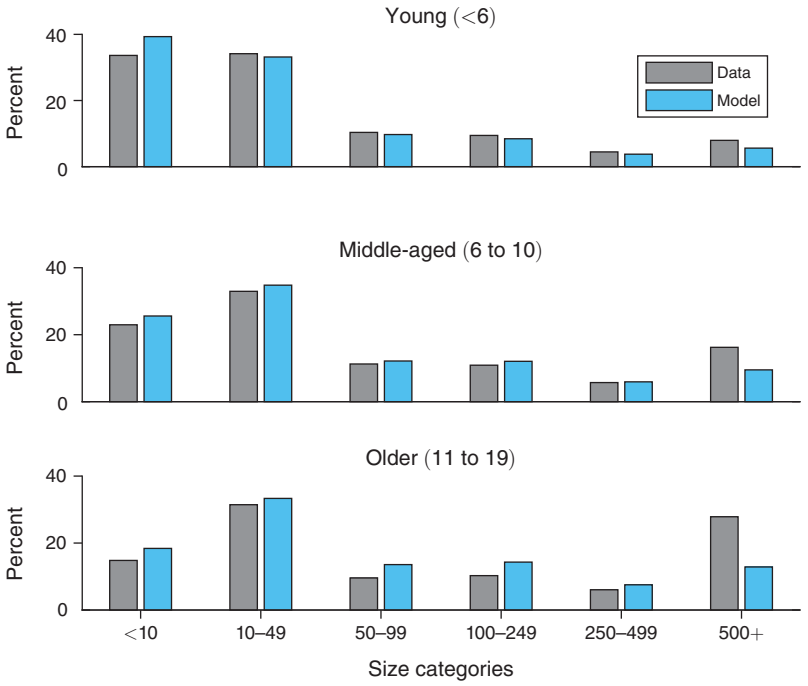

Figure 6. Employment Shares of Different Age/Size Bins: Model versus Data

Notes: Employment shares by firm age and size (employment). Values are expressed as percentages of total employment in firms between 0 to 19 years old, both in the data and the model. Data are obtained from the Business Dynamics Statistics, an aggregated and publicly available version of the LBD over the corresponding time period.

10 parameters. Additionally, we consider how the model fits the employment distribution by age and size, which is not directly targeted. Figure 6 shows employment shares of different age/size bins, in the model and in the data. Overall, the model fits this distribution well.[21]

The associated parameter values for our benchmark model are shown in Table 2. The fixed cost is estimated to be 0.54, which is about one-half of the wage of a single employee. The exogenous exit rate is estimated to be about 4.1 percent. Thus, a substantial fraction of firms exits for reasons unrelated to their fundamentals. However, Figure 5 makes clear that there is also a substantial amount of endogenous exit, as the overall exit rate in the model varies between 15.5 percent at age zero to 5.8 percent at age nineteen.

The remaining parameters are somewhat difficult to interpret individually, especially since the parameter values are for the unconditional distributions, whereas the equilibrium distributions are truncated by selection. However, online Appendix Section D.4 provides an analysis of the sources of identification of the parameters of the process. Importantly, similar to the results in the statistical model, also in the more complex structural model important identifying information about the dispersion of ex ante differences across firms is obtained from the long-horizon autocovariances.

[21] The only exception is the employment share of very large old firms which is somewhat understated in the model compared to the data. However, online Appendix Section E.2 shows that recalibrating the model and explicitly targeting the firm size distribution does not change our results.

TABLE 2—PARAMETER VALUES

Parameter		Value
Set a priori		
β	discount factor	0.96
η	elasticity of substitution	6.00
f^e	entry cost	0.44
Used to target moments		
f	fixed cost of operation	0.539
δ	exogenous exit rate	0.041
μ_θ	permanent component θ, mean	−1.762
σ_θ	permanent component θ, standard deviation	1.304
$\sigma_{\tilde{u}}$	initial condition u_{-1}, standard deviation	1.572
$\sigma_{\tilde{v}}$	initial condition v_{-1}, standard deviation	1.208
σ_ε	transitory shock ε, standard deviation	0.307
σ_z	noise shock z, standard deviation	0.203
ρ_u	permanent component, persistence	0.393
ρ_v	transitory component, persistence	0.988

Notes: Top three parameters are calibrated as discussed in the main text. The remaining parameters are set such that the model matches the empirical autocovariance of employment and the age profiles of average size and exit rates from age 0 to 19.

C. *The Importance of Ex Ante Heterogeneity Revisited*

Before moving to the main results on aggregate implications, we briefly revisit and extend the conclusions drawn from the statistical model. An advantage of the structural model is that it allows us to study the sources of employment dispersion while accounting for endogenous selection of firms. Related to this, the structural model enables us to consider the importance of ex ante heterogeneity not only for dispersion of firm size, but also for exit rates. Finally, the structural model allows us to study explicitly the importance of firms with ex ante high growth potential for average firm size (as opposed to size dispersion). We study these three outcomes in turn.

Employment Dispersion.—In the structural model, firm-level employment is given by $n_i = \xi \varphi_i^{EXA} \varphi_i^{EXP}$, where $\varphi_i^{EXA} = e^{u_i + v_i}$ is the ex ante component of demand, $\varphi_i^{EXP} = e^{w_i + z_i}$ is the ex post component, and ξ is defined above as $\xi \equiv ((\eta - 1)/\eta)^\eta w^{-\eta} Y$. In contrast to the statistical model, however, the ex ante and ex post component are no longer orthogonal to each other, due to a correlation induced by endogenous firm selection. This occurs because firms with relatively poor ex ante conditions can survive only if they were exposed to favorable ex post shocks and vice versa.[22] Accounting for this correlation, we instead decompose the variance of logged employment as

$$\begin{aligned} (4) \qquad \operatorname{var}[\ln n_i] &= \operatorname{var}\left[\ln \varphi_i^{EXA}\right] + \operatorname{var}\left[\ln \varphi_i^{EXP}\right] + 2\operatorname{cov}\left[\ln \varphi_i^{EXA}, \ln \varphi_i^{EXP}\right] \\ &= \operatorname{cov}\left[\ln \varphi_i^{EXA}, \ln n_i\right] + \operatorname{cov}\left[\ln \varphi_i^{EXP}, \ln n_i\right], \end{aligned}$$

[22] In the statistical model, shocks are assumed to be distributed independently and therefore $\operatorname{cov}[\ln \varphi_i^{EXA}, \ln \varphi_i^{EXP}] = 0$.

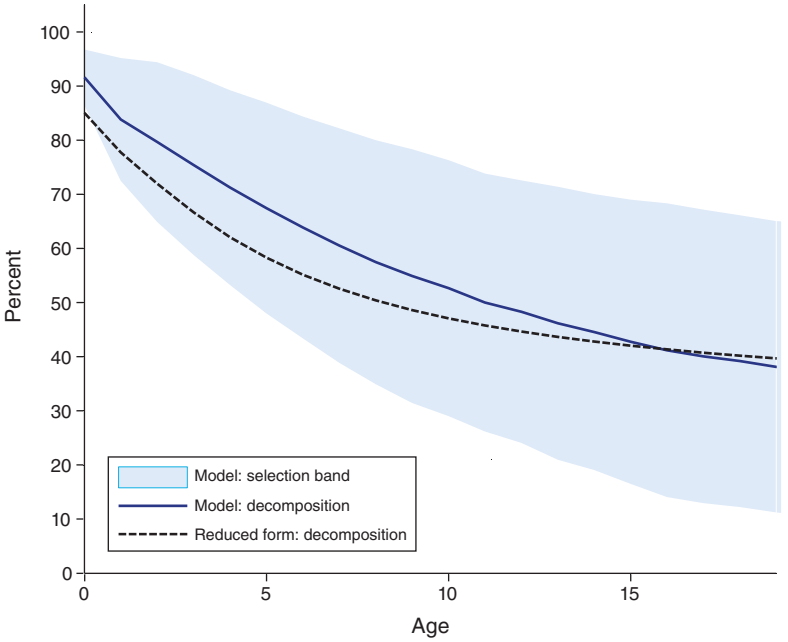

FIGURE 7. CONTRIBUTION OF EX ANTE HETEROGENEITY TO CROSS-SECTIONAL EMPLOYMENT DISPERSION

Notes: Contributions of ex ante heterogeneity to the total cross-sectional variance of log employment by age. *Reduced form* refers to the estimates from Figure 4 (left panel), *Model: covariance decomposition* is the decomposition based on the second line in equation (4). The shaded areas (*Model: selection band*) is constructed based on the first equality in equation (4) by attributing, in turn, the term $2\text{cov}(\ln \varphi_i^{EXA}, \ln \varphi_i^{EXP})$ fully to the ex ante component and to the ex post component.

where the second line evenly splits the covariance term in the first line between the ex ante and ex post components.

Figure 7 depicts the contribution of ex ante heterogeneity in the structural model (solid line), i.e., $\text{cov}[\ln \varphi_i^{EXA}, \ln n_i] / \text{var}[\ln n_i]$, together with the reduced-form decomposition (dashed line).[23] The figure also plots a "selection band" based on attributing, in turn, the covariance term in the first line of the equality in equation (4) either fully to the ex ante component or fully to the ex post component. While the structural model reestablishes our earlier conclusion that ex ante heterogeneity is a key source of size dispersion, it also highlights the importance of firm selection. The widening selection band indicates that selection has an increasingly important impact on the cross-sectional dispersion of firm size as firms age.

Exit Rates.—The previous paragraphs show that firm selection is important in our analysis. In what follows, we document that also firm selection is to a large extent driven by heterogeneity in ex ante components. Toward this end, we run a counterfactual simulation in which we use the firms' baseline decision rules but

[23] The slight difference reflects the fact that the structural model fits more moments, compared to the statistical one, and therefore provides a somewhat different fit to the autocovariance matrix.

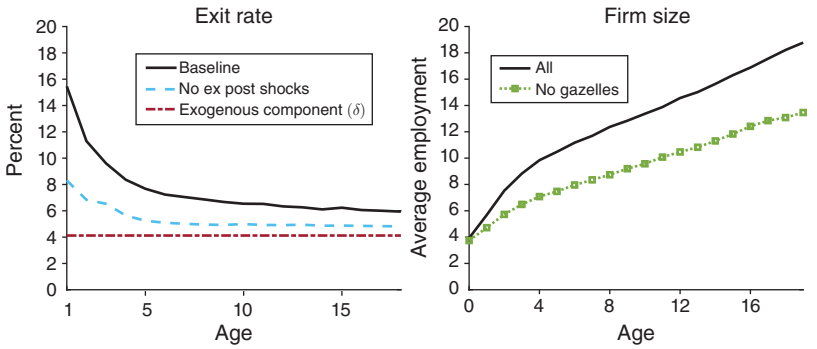

FIGURE 8. EXIT RATES AND AVERAGE FIRM SIZE BY AGE

Notes: Left panel: exit rates by age in the baseline model, in the counterfactual economy with selection only on ex ante profiles, and in the counterfactual economy with only exogenous exit, i.e., exogenous rate δ. Right panel: average size by age in the baseline model among all firms, and among all firms except for gazelles.

we completely shut down ex post shocks to demand, i.e., we set $\sigma_\varepsilon = \sigma_z = 0$.[24] We do, however, preserve exogenous exit shocks. The resulting average exit rate profile is therefore informative about the extent to which firms' exit is driven by ex ante characteristics. For example, firms may have declining ex ante demand profiles because of favorable initial condition coupled with a poor long-run growth potential. Such firms will find it economically viable to operate in the initial years, but not later on.

The results of this counterfactual simulation are presented in the left panel of Figure 8. The difference between the baseline exit profile and the constant exogenous exit rate is the endogenous component of the exit rate. With no ex post shocks, the exit rate is lower but it retains its declining pattern with age. Interpreting the difference between the baseline exit rate and that in the counterfactual simulation without ex post shocks as the amount of endogenous exit driven by selection on ex ante profiles, the figure suggests a quantitatively important role of ex ante characteristics for firm selection. Specifically, between 30 and 45 percent of overall endogenous exit is driven by selection on ex ante profiles. Even among older firms there is still selection on exit ante profiles, as some ex ante profiles decline very gradually.

High-Growth Firms.—Finally, we document that ex ante heterogeneity is not only important for firm selection, but also for firm growth among continuing firms. In what follows we specifically focus on high-growth firms, labeled as "gazelles," which have obtained much attention in the recent literature.[25] We start by defining gazelles as those startups with an ex ante projected growth rate of at least 20 percent

[24] This is equivalent to allowing exit to depend only on the ex ante profile, rather than the ex ante profile and the moving average of ex post shocks. These counterfactuals are also partial equilibrium simulations in the sense that we do not recompute the equilibrium and we keep aggregate demand fixed.

[25] See Guzman and Stern (2015) for a study of the predictability of high-growth outcomes in firms and Haltiwanger et al. (2016) for an analysis of the importance of high-growth firms for aggregate outcomes.

annually, over the first five years, and an expected employment level of at least 10 workers at some point during their lifetimes.[26,27]

The results of our classification show that gazelles account for only 5.4 percent of all startups. To gauge their impact on average firm growth, we conduct a similar counterfactual exercise as with firm selection. Specifically, we recompute the growth profile but this time excluding gazelles. The right panel of Figure 8 shows that without gazelles average size is considerably lower and the difference remains large up to at least age 19. At that age, average size is more than 25 percent lower than in the baseline.

Therefore, ex ante factors are important not only for firm selection, but also for firm growth among continuing businesses. While our baseline model deliberately abstracts from many interesting features of a more realistic economic environment, we show our main conclusions extend to more complex environments.

D. *Extensions*

We examine the robustness of the model results with respect to information frictions and flexible labor supply. In Section III, we will consider versions of the model with adjustment costs and financial frictions.

Information Frictions.—Information frictions play an important role in some prominent firm dynamics model, such as the seminal work by Jovanovic (1982). In our baseline model, however, firms have perfect information about the components of the shock process. One may wonder to what extent relaxing this assumption would affect the results and the interpretation of the documented empirical patterns.

To investigate these issues, we conduct two exercises (see online Appendix Section E.1). First, we consider the equilibrium generated from the model with perfect information, but we take the perspective of an outside observer who can never perfectly see the states, but rather learns about them in an optimal Bayesian way. The results suggest that one can learn about ex ante profiles extremely quickly, with most of the uncertainty being resolved in the first year upon entry.

Second, we solve a version of the model in which firms themselves have imperfect information and can only observe the fully underlying states one year after entry.[28] The conclusions from this model turn out to be very similar to those from the baseline with perfect information.[29]

[26] Defining gazelles using not only growth rates but also size excludes firms which grow quickly in percentage terms but nevertheless always stay small in terms of employed workers.

[27] While our definition of gazelles is in line with the literature, we classify firms according to their *ex ante* profiles at startup. By contrast, the existing literature has classified firms based on ex post realizations, since ex ante profiles are not directly observable. Using ex post realizations, however, it then follows almost by construction that gazelles contribute disproportionately to aggregate job creation because they are the firms that grew a lot. Online Appendix Section E.4 shows how our classification maps into that based on an ex post definition of gazelles. Moreover, while our classification is based on employment, Appendix Section E.4 also offers results for definitions based on firm value which relate more closely to some papers in the literature (see, e.g., Guzman and Stern 2015).

[28] Allowing firms to observe the state after one year eases the computational burden. Note however that the first exercise suggests that the bulk of the information friction is resolved after one year.

[29] Another interesting possibility is that agents might receive advance information on ex post shocks, as in the literature on news shocks in macroeconomics, see Beaudry and Portier (2004). If some of the information is already known upon entry, the importance of ex ante heterogeneity would be even larger than we estimate.

Flexible Labor Supply.—Our baseline model assumes fixed labor supply. In online Appendix Section E.3, we analyze a version of the model with flexible labor supply. For most of our baseline results, the introduction of flexible labor supply has no consequences, owing to our calibration strategy which targets the life-cycle profile of firm size. While aggregate outcomes do depend on labor supply, we also find that these implications are limited.

III. Macroeconomic Implications

A natural question is: what is missed by ignoring the sources of firm heterogeneity? In this section, we explore the extent to which the presence of ex ante heterogeneity across firms matters for our understanding of the macroeconomy. In the literature, firm dynamics models are often used as laboratories to quantitatively study the impact of firm-level frictions on the macro economy. Hopenhayn and Rogerson (1993), who examine the aggregate effects of a firing tax, may be the most famous early example. We provide examples which show that the outcome of such exercises can depend crucially on the nature of firm growth.

To this end, we contrast our baseline model with rich ex ante heterogeneity to a restricted version, often used in existing studies, in which the underlying shock process has an AR(1) structure. When comparing these economies, we ensure that both have near equivalent observable heterogeneity in terms of firm size and serial correlation of employment.

We then introduce two distinct micro-level frictions in each economy and study how the frictions' aggregate effects differ. In particular, we consider the effects of nonconvex adjustment costs and the effects of financial constraints on operation. These types of frictions have a long tradition in the firm dynamics literature with similar versions being used, for instance, to analyze how firing taxes affect the misallocation of resources or how financial frictions impact firm entry and exit decisions.

Our aim is to provide intuitive examples of how the richness of the micro-level shock process can matter for aggregate outcomes. Therefore, we choose the formulation of the frictions in a relatively simple way. We defer a range of robustness and extensions, as well as a consideration of other frictions, to the online Appendix.

The Restricted Model.—Our restricted model is a widely used AR(1) process with noise, which we obtain by setting $\rho_u = \rho_v = \rho_w = \rho$ and fixing $\theta_i = \mu_\theta$ and $u_{i,-1} = 0$ in the baseline. These restrictions imply that the underlying process for firm-level demand, $\ln \varphi_{i,a} = u_{i,a} + v_{i,a} + w_{i,a} + z_{i,a}$, evolves as

$$u_{i,a} + v_{i,a} + w_{i,a} = \mu_\theta + \rho\big(u_{i,a-1} + v_{i,a-1} + w_{i,a-1}\big) + \varepsilon_{i,a},$$

where $\varepsilon_{i,a} \sim N(0, \sigma_\varepsilon^2)$, $v_{i,-1} \sim N(0, \sigma_v^2)$, and $z_{i,a} \sim N(0, \sigma_z^2)$.[30] Given these restrictions, it is necessary to reparametrize the model. We do so by matching the same targets as in the baseline with the exception of the autocovariance matrix.

[30] Note that with the exception of now allowing for $\sigma_z > 0$, these restrictions are the same as those of model I in Section ID.

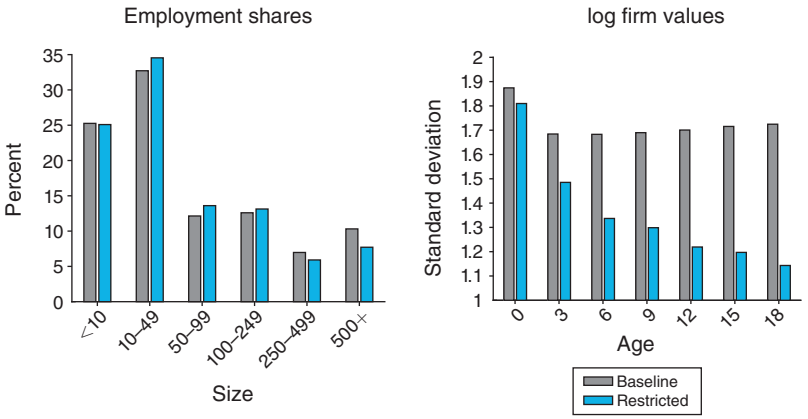

FIGURE 9. MODEL COMPARISON: BASELINE AND RESTRICTED VERSION

Note: Left panel shows employment shares by firm size and the right panel shows standard deviations of log firm values by firm age in the *Baseline* and *Restricted* models.

Instead, we follow the literature (see, e.g., Hopenhayn and Rogerson 1993) and target ρ_n and σ_n from the following regression:

$$\ln n_{i,a} = \bar{n} + \rho_n \ln n_{i,a-1} + \eta_{i,a},$$

where $\eta_{i,a}$ is a residual with mean zero and standard deviation σ_n. That is, we in effect match the unconditional serial correlation of employment, ignoring the strong age dependence of this correlation revealed by the autocovariance structure. Online Appendix Section D.2 shows that the model fit, including the implied (untargeted) autocovariance matrix, turns out to be very similar to the corresponding statistical model in Section ID.[31]

Comparison of the Two Economies.—Before moving on to our two exercises, let us first discuss more broadly how the baseline model differs from the restricted version. The left panel of Figure 9 shows that the two models have essentially the same (untargeted) firm size distribution. However, the right panel of Figure 9 shows that the two models differ substantially when it comes to the distribution of firm values. In particular, the baseline economy is characterized by a highly dispersed distribution of firm values, driven by ex ante heterogeneity in growth profiles. This is not true for the restricted model, especially among older firms.

Intuitively, firms in the restricted model are moving to the same long-run size and thus their firm values, the net present values of future profits, are much more similar to each other. Firm values, however, are critical to firms' forward-looking decisions such as entry, growth, and exit. A wider dispersion of firm values generally implies that there are fewer "marginal" firms which are indifferent between, for

[31] Online Appendix Section D.2.1 shows that similar results are obtained when parametrizing the restricted model using exactly the same targets as the baseline framework, i.e., including the autocovariance matrix in place of the parameters of an AR(1) in log employment.

example, exiting and continuing or between adjusting to a shock or not. Therefore, even though the two economies have a very similar firm size distribution (a typical parametrization target in existing firm dynamics studies), they display very different aggregate properties. This is precisely what the two exercises below aim to highlight.

A. *Application 1: Adjustment Costs*

In the first exercise, we introduce a nonconvex adjustment cost to demand growth into both versions of the model, related to, e.g., Gourio and Rudanko (2014) or Foster, Haltiwanger, and Syverson (2016). In this setting, demand becomes endogenous.[32]

Specifically, we assume that whenever $\varphi' > \varphi$ the incumbent firm has two options: retaining its current level of demand, φ, or paying a cost κ and obtaining the new higher level of demand φ'. The adjustment cost may thus prevent firms from growing their demand and reaching their full potential. One can think this as a cost a firm needs to pay in order to seize a demand growth opportunity, related to for example marketing costs or organizational restructuring.[33] In both the baseline and the restricted model, the adjustment cost is calibrated such that the average cost paid by adjusting firms is 1 percent of their output: see online Appendix Section F.1 for details.

Panel A of Table 3 shows the long-run impact relative to the case when adjustment costs are absent, $\kappa = 0$. Let us begin with the restricted version of the model, which predicts substantial aggregate losses induced by the adjustment costs. Firms considerably decrease their demand accumulation which results in a strong decline in average firm size and an increase in firm exit. Given the assumption of a fixed labor supply, this leads to an increase in the number of firms. All these effects result in a decline in the wage and a drop in aggregate output of more than 3 percent. Similar results have been found by, e.g., Hopenhayn and Rogerson (1993).

By contrast, in the baseline model the macroeconomy is largely insensitive to the introduction of adjustment costs. There is a slight reduction in firm values, which puts downward pressure on the real wage. In equilibrium, because of fixed labor supply, firms are larger but fewer.[34] Intuitively, the presence of ex ante heterogeneity in the baseline model results in *aggregates* being heavily influenced by a small number of high-value firms with high growth potential. These firms, in turn, tend not to be discouraged by adjustment costs as they seize any opportunity to grow toward their long-run potential.

B. *Application 2: Financial Frictions*

Our second exercise considers a financial constraint on the operation of firms. This exercise relates to a growing literature on financial frictions, see, for instance,

[32] See also Arkolakis (2016), Luttmer (2011), Drozd and Nosal (2012), and Perla (2015). Similarly, there is a vast literature in which productivity is endogenous through innovation decisions.

[33] This formulation has the practical advantage that it does not introduce any additional state variables to the model. Intuitively, the firm chooses between staying at the current demand state and moving to the state dictated by the new draw. This decision to adjust does not depend on lagged employment.

[34] The compositional shift toward larger businesses that are less likely to exit reduces the average exit rate, even as endogenous exit *conditional* on type is virtually unchanged. See also Karahan, Pugsley, and Şahin (2019) who show the importance of compositional change for the aggregate exit rate.

Table 3—Aggregate Impact of Micro-Level Frictions (Percent Change)

	Output	Wage	Size	Exit	Firms
Panel A. Adjustment costs					
Restricted model	−3.0	−0.6	−23.8	+7.2	+28.0
Baseline model	+0.1	−0.0	+4.6	−0.9	−4.3
Panel B. Financial frictions					
Restricted model	−1.3	−0.8	+31.8	+3.8	−24.6
Baseline model	−3.4	−0.7	+82.3	+3.1	−46.6

Notes: Long-run impact of introducing adjustment costs and financial frictions in the baseline economy and in the restricted version of the model. In both economies adjustment costs (κ) amount to 1 percent of output among adjusting firms and the financial constraint (ζ) is set to zero. Reported values are relative to the baseline without frictions. *Output* refers to aggregate production, *Wage* is the real wage rate, *Size* is average firm size, *Exit* is the average exit rate, and *Firms* refers to the number of incumbent firms.

Buera, Koboski, and Shin (2011); D'Erasmo and Boedo (2012); and Midrigan and Xu (2014), although the precise specification of the friction differs across studies. Specifically, we assume that firms can hold a risk-free asset, denoted b_i, which pays a net real interest rate $r = (1/\beta) - 1$. However, the firm is subject to a borrowing limit:

$$b_i' \geq \zeta,$$

where $\zeta \leq 0$, and b_i' denotes end-of-period assets in the firm. If a firm cannot meet this limit, it is forced to exit. Upon entry, a firm receives an initial equity injection $\tilde{b}$, but subsequently no additional equity injections are possible. When a firm exits with positive assets, then these assets are returned to the owners (i.e., the representative household). If a firm exits with debt ($b_i' < 0$) then the owners must settle the remaining debt. In both the baseline and the restricted model, the financial friction is calibrated such that $\zeta = 0$, i.e., firms cannot borrow. Finally, we assume that firms enter without any initial equity, $\tilde{b} = 0$. Further details can be found in online Appendix Section F.1.

The financial friction creates inefficient exit without distorting other margins. Inefficient exit happens either when a firm hits the constraint despite having a positive economic value, or when an unconstrained firm chooses to exit because the possibility of the financial friction binding in the future depresses the firm's value below zero.[35]

Panel B of Table 3 shows the long-run impact of financial frictions relative to the frictionless baseline. In this case, the baseline economy exhibits a stronger output fall compared to the restricted version of the model. The main reason for this is that the number of firms affected by the friction is much larger (almost one-half compared to about one-quarter in the restricted version). This is again because of the wider dispersion of profits (and hence firm values) in the baseline model. In the baseline economy there are many more firms with poor profit performance early on

[35] In online Appendix Section F.1 we show that hiring decisions remain static and are unaffected, and that firms' saving does not depend on their assets. The possibility of hitting the constraint ensures that the shadow value of firm assets is always greater than the real interest rate, so firms optimally retain all earnings until exit.

in their lives, which nevertheless would be viable in the frictionless world because of their long-run potential. Specifically these high-growth-potential firms are being strongly affected by the friction. The above results show two examples in which the presence of ex ante heterogeneity in growth profiles significantly alters our understanding of the macroeconomic impact of micro-level frictions. Online Appendix Section F.1 considers other micro-level frictions. Again, the baseline and restricted models display very different aggregate properties.

IV. Changes in the Nature of Firm Growth

Finally, we study how the nature of firm growth might have changed over the last few decades. This is especially relevant in light of the observed decline in the "dynamism" of US businesses, see, e.g., Decker et al. (2016). Using our model we can study underlying drivers of changes in firm dynamics. In particular, we are able to assess changes in firm dynamics resulting from changes in the distribution of ex post shocks and ex ante growth profiles. We can also study the aggregate implications of such changes.

To analyze the changes in firm dynamics, we split our data into two subsamples. The "early sample" includes firms born between 1979 and 1985, while the "late sample" includes firms born between 1986 and 1993. Again, we follow all firms up to age 19. In what follows, we first document changes in the three sets of key moments, the autocovariance function, the average size profile, and the exit profile. Next, we re-estimate our model on the two subsamples and interpret the changes in the data through the lens of our model with a particular focus on gazelles.

A. *Changes in the Data*

Figure 10 plots the three sets of key moments in the two samples. The top panel shows that the autocovariance function of logged employment of firms (balanced panel) has remained remarkably stable over time. This suggests that the relative importance of ex ante and ex post heterogeneity has not changed much. The bottom right panel shows that exit rates have also remained stable, see also Pugsley and Şahin (2019).

What has changed, however, is the profile of average size by age, which is shown in the bottom left panel of Figure 10. Over time, this profile has flattened. At startup, average size is about 7 employees in both the early and the late sample. However, by age 19, average employment has declined by almost 25 percent from an average 22 workers in the early sample to 17 employees in the late sample. In addition, this divergence in size profiles sets in gradually with age.

B. *Estimating Changes in Firm Dynamics*

To investigate the observed changes in firm dynamics and their aggregate consequences, we first re-estimate the model on the two subsamples. The estimated parameter values and model fit are shown in online Appendix Section D.3. The difference in the distribution of ex ante profiles is most apparent when examining gazelles, which we have defined entirely by their ex ante characteristics.

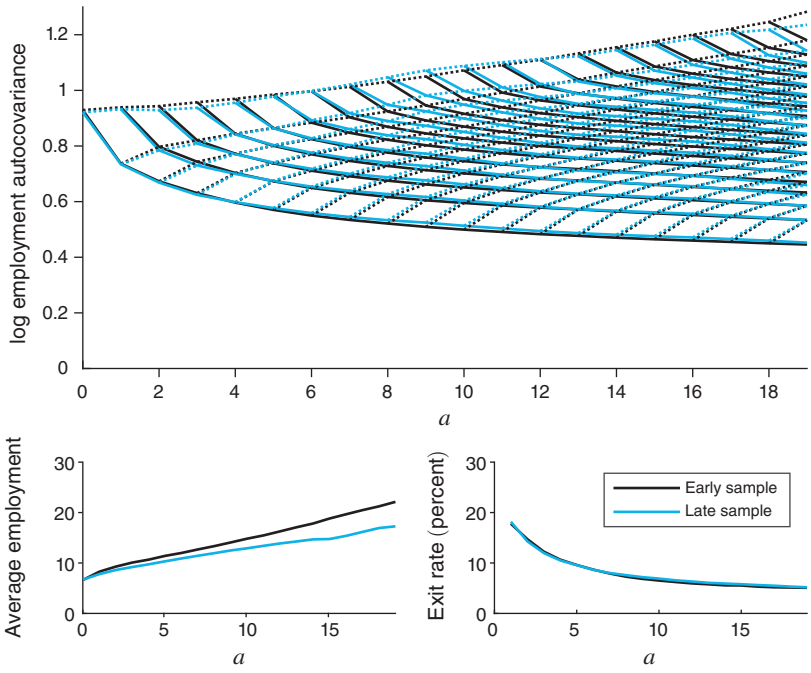

FIGURE 10. SPLIT-SAMPLE DATA MOMENTS

Notes: Top panel: autocovariances of log employment between age $a = h + j$ and age $h \leq a$ in the early and the late sample, for the balanced panel of firms surviving up to at least age $a = 19$. Bottom left: average employment by age a (unbalanced panel). Bottom right: exit rate by age a.

Are Gazelles Dying Out?—In what follows, we focus on (ex ante identified) gazelles and examine the differences in their number and quality between samples. Toward this end, we compute the fraction of gazelles in the population of firms in both subsamples. This is shown in the left panel of Figure 11. Among startups, the fraction of gazelles has declined from 6.4 percent in the early sample to 5.3 percent in the late sample. As firms age, the fraction of gazelles increases because gazelles are relatively unlikely to shut down compared to other firms with lower growth potential. Therefore, the gap in the share of gazelles widens with age between the two samples.

The right panel shows the average size profile of gazelles. In both subsamples, gazelles start with around 7 employees, but grow quickly to reach on average about 46 employees by age 5. Around age 10, however, the two subsamples diverge, and a reduction in the average size between the two subsamples becomes apparent. Thus, in the late sample gazelles on average do not grow as large as in the early sample.[36]

Aggregate Implications.—What are the aggregate implications of the decline in the presence of gazelles and in their growth profiles? Figure 12 plots the average

[36] Consistent with this finding, online Appendix Section D.3 also shows that the model generates a substantial decline in skewness of firm growth rates across the two subsamples. Moreover, these results hold also for survivors only and are therefore not primarily driven by selection.

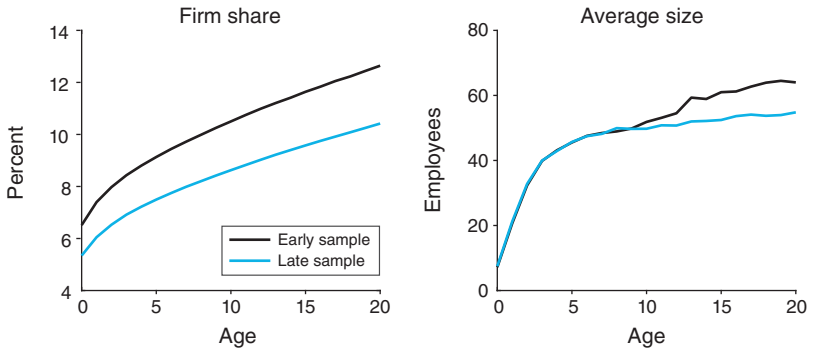

FIGURE 11. CHARACTERISTICS OF GAZELLES IN THE EARLY AND LATE SAMPLE (FIRMS)

Notes: Share of gazelles in the total number of firms (left panel) and the average size, by age, among gazelles (right panel). Gazelles are classified on an ex ante basis, as those startups with an ex ante growth rate of at least 20 percent annually, over the first five years, and an associated employment level that exceeds 10 at some point during this period.

size profile, which has flattened between the two subsamples. To assess the contribution of disappearing gazelles to this shift, we use the fact that at any age the average size among all firms is the sum of the average size of gazelles and non-gazelles, weighted by their respective firm share. We then construct a partial equilibrium counterfactual in which we recompute the average size in the early sample, but with the average size and firm share profiles of the gazelles in the late sample. The dashed line shows this counterfactual and suggests that changes associated with gazelles alone can account for roughly one-half of the decline in the average size profile. This is remarkable, since gazelles account for only about 5 percent of the startups.

Finally, we evaluate the implications for aggregate output by comparing equilibrium output between the early and late periods. This calculation, which accounts for general equilibrium effects, shows that aggregate output declines by 4.85 percent. Thus, seemingly small changes in the distribution of firms, such as the decline in the (already low) share of high-potential startups, as well as a reduction in their growth potential, emerge as important drivers of *aggregate* changes.[37,38]

Discussion.—Our results offer some new insights into an ongoing discussion of changes in US business dynamism (see, e.g., Decker et al. 2018, Guzman and Stern 2020). In particular, they speak to the declining importance of high-growth firms first documented by Decker et al. (2016). We find that the disappearance of such firms is related to ex ante factors, i.e., fewer "gazelles" are entering the US economy and diminished growth potential (relative to the early period) for those who do. Importantly, our results suggest that the seeds of the decline in dynamism might

[37] Within the model, this decline is entirely driven by a change in output per worker, i.e., labor productivity, since we keep labor supply fixed. Online Appendix Section E.3 shows results for the case of flexible labor supply.

[38] Shifts in the number of startups may also have macroeconomic effects (Sedláček 2020).

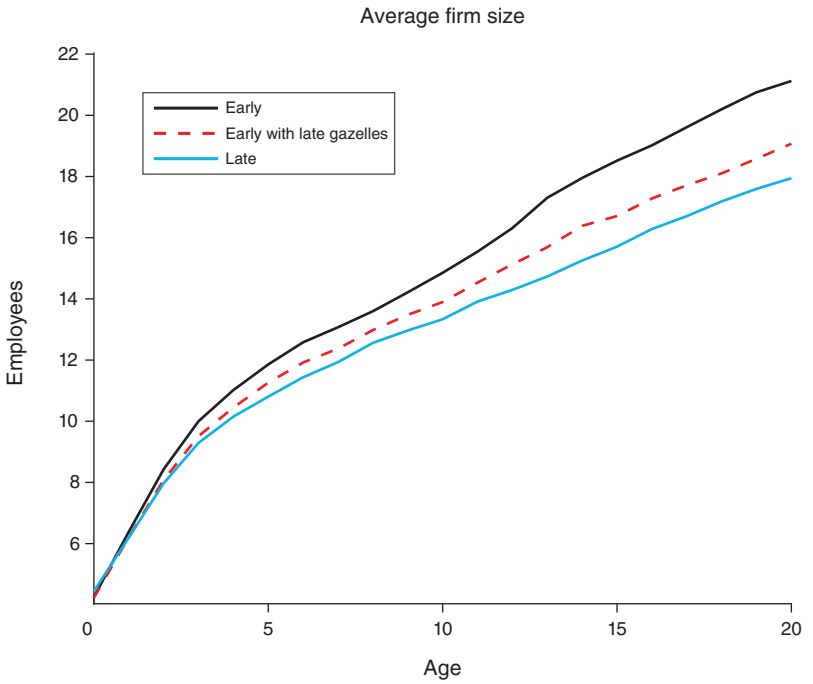

FIGURE 12. THE IMPACT OF DISAPPEARING GAZELLES

Notes: The figure plots the average size profile among all firms in the early sample and the late sample. It also plots a counterfactual average size profile for the early sample, computed by replacing firm share and average size profile of gazelles by their counterparts from the late sample.

have already been sown in the mid to late 1980s.[39] Finally, we find that the aggregate repercussions of the observed changes are quantitatively substantial.

V. Conclusions

We have used data on the population of US firms over several decades to better understand why some startups grow rapidly whereas others remain stagnant or exit quickly. To this end, we documented the autocovariance structure of employment and exploited this structure to estimate firm dynamics models, which allowed us to disentangle heterogeneous ex ante profiles from ex post shocks.

We found a dominant role for heterogeneous ex ante profiles, which capture future potential present at the moment of startup. Much of the firm size distribution, firm dynamics, and the prevalence of high-growth startups, "gazelles," is determined by ex ante heterogeneity in growth profiles. Moreover, the presence of such heterogeneity shapes the behavior of the macroeconomy. Indeed, not accounting for the precise

[39] It is very well possible that the effects of such changes only become noticeable in the aggregates after more than a decade. Recall the flatting of the growth profile is particularly pronounced after age 10. Moreover, aggregate effects due to incoming cohorts of startup accumulate over time. Note also that our results do not exclude the possibility that a decline in dynamism is also accompanied by further change in the responsiveness of firms to (ex post) shocks. In fact, they complement Decker et al. (2018) who find evidence for declining responsiveness in among recent cohorts.

nature of firm growth has the potential to dramatically change the macroeconomic predictions of firm dynamics models. Finally, having in mind recent concerns about the disappearance of gazelles, we have investigated potential changes in the nature of firm growth over time. Re-estimating our model using this information, we found a decline in the presence and growth potential of "gazelles" in the population of startups, with important repercussions for aggregate output.

An intriguing question left for future research is whether there is a connection between the demise of gazelle startups and the decline in the aggregate labor share of income, which also started in the late 1980s. For example, Autor et al. (2017) suggest that the decline in the labor share was due to an increase in product market concentration, giving rise to "superstar firms." Alternatively, the late 1980s were also times of large fiscal reforms which may have affected firm dynamics, see, e.g., Sedláček and Sterk (2019). Finally, our results also highlight the need to further study the role of startup conditions, and the individuals who become entrepreneurs and their decisions taken before or at the time of startup.

REFERENCES

Abbring, Jaap, and Jeffrey Campbell. 2005. "A Firm's First Year." Tinbergen Institute Discussion Paper 05-046/3.

Abowd, John M., and David Card. 1989. "On the Covariance Structure of Earnings and Hours Changes." *Econometrica* 57 (2): 411–45.

Akcigit, Ufuk, Harun Alp, and Michael Peters. 2017. "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries." NBER Working Paper 21905.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58 (2): 277–97.

Arkolakis, Costas. 2016. "A Unified Theory of Firm Selection and Growth." *Quarterly Journal of Economics* 131 (1): 89–155.

Arkolakis, Costas, Theodore Papageorgiou, and Olga A. Timoshenko. 2018. "Firm Learning and Growth." *Review of Economic Dynamics* 27: 146–68.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2017. "The Fall of the Labor Share and the Rise of Superstar Firms." Unpublished.

Barseghyan, Levon, and Riccardo DiCecio. 2011. "Entry Costs, Industry Structure, and Cross-Country Income and TFP Differences." *Journal of Economic Theory* 146 (5): 1828–51.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2013. "Cross-Country Differences in Productivity: The Role of Allocation and Selection." *American Economic Review* 103 (1): 305–34.

Beaudry, Paul, and Franck Portier. 2004. "An Exploration into Pigou's Theory of Cycles." *Journal of Monetary Economics* 51 (6): 1183–216.

Belenzon, Sharon, Aaron K. Chatterji, and Brendan Daley. 2017. "Eponymous Entrepreneurs." *American Economic Review* 107 (6): 1638–55.

Bento, Pedro, and Diego Restuccia. 2019. "The Role of Nonemployers in Business Dynamism and Aggregate Productivity." NBER Working Paper 25998.

Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin. 2011. "Finance and Development: A Tale of Two Sectors." *American Economic Review* 101 (5): 1964–2002.

Cabral, Luís M. B., and José Mata. 2003. "On the Evolution of the Firm Size Distribution: Facts and Theory." *American Economic Review* 93 (4): 1075–90.

Campbell, Jeffrey R., and Mariacristina De Nardi. 2009. "A Conversation with 590 Nascent Entrepreneurs." *Annals of Finance* 5 (3–4): 313–40.

Chamberlain, Gary. 1984. "Panel Data." In *Handbook of Econometrics*, Vol. 2, edited by Zvi Griliches and Michael D. Intriligator, 1274–318. Amsterdam: North-Holland.

DeBacker, Jason, Vasia Panousi, and Shanthi Ramnath. 2018. "A Risky Venture: Income Dynamics within the Non-Corporate Private Business Sector." Unpublished.

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2016. "Where Has All the Skewness Gone? The Decline in High-Growth (Young) Firms in the U.S." *European Economic Review* 86: 4–23.

Decker, Ryan A., John C. Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2018. "Changing Business Dynamism and Productivity: Shocks vs. Responsiveness." NBER Working Paper 24236.

D'Erasmo, Pablo N., and Hernan J. Moscoso Boedo. 2012. "Financial Structure, Informality and Development." *Journal of Monetary Economics* 59 (3): 286–302.

Drozd, Lukasz A., and Jaromir B. Nosal. 2012. "Understanding International Prices: Customers as Capital." *American Economic Review* 102 (1): 364–95.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2016. "The Slow Growth of New Plants: Learning about Demand?" *Economica* 83 (329): 91–129.

Gabaix, Xavier. 2009. "Power Laws in Economics and Finance." *Annual Review of Economics* 1: 255–93.

Gourio, Francois. 2008. "Estimating Firm-Level Risk." Unpublished.

Gourio, Francois, and Leena Rudanko. 2014. "Customer Capital." *Review of Economic Studies* 81 (3): 1102–36.

Guvenen, Fatih. 2009. "An Empirical Investigation of Labor Income Processes." *Review of Economic Dynamics* 12 (1): 58–79.

Guvenen, Fatih, and Anthony A. Smith Jr. 2014. "Inferring Labor Income Risk and Partial Insurance from Economic Choices." *Econometrica* 82 (6): 2085–129.

Guzman, Jorge, and Scott Stern. 2015. "Where Is Silicon Valley?" *Science* 347 (6222): 606–09.

Guzman, Jorge, and Scott Stern. 2019. "Measuring Founding Strategy." Unpublished.

Guzman, Jorge, and Scott Stern. 2020. "The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 32 US States, 1988–2014." *American Economic Journal: Economic Policy* 12 (4): 212–43.

Haltiwanger, John, Ron Jarmin, Robert Kulick, and Javier Miranda. 2016. "High Growth Young Firms: Contribution to Job, Output and Productivity Growth." US Census Bureau Center for Economic Studies Paper CES-WP-16-49.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda. 2013. "Who Creates Jobs? Small versus Large versus Young." *Review of Economics and Statistics* 95 (2): 347–61.

Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60 (5): 1127–50.

Hopenhayn, Hugo, and Richard Rogerson. 1993. "Job Turnover and Policy Evaluation: A General Equilibrium Analysis." *Journal of Political Economy* 101 (5): 915–38.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein. 2016. "Quantifying the Sources of Firm Heterogeneity." *Quarterly Journal of Economics* 131 (3): 1291–364.

Hsieh, Chang-Tai, and Peter J. Klenow. 2014. "The Life Cycle of Plants in India and Mexico." *Quarterly Journal of Economics* 129 (3): 1035–84.

Hurst, Erik, and Benjamin Wild Pugsley. 2011. "What Do Small Businesses Do?" *Brookings Papers on Economic Activity* 2: 73–118.

Jovanovic, Boyan. 1982. "Selection and the Evolution of Industry." *Econometrica* 50 (3): 649–70.

Karahan, Fatih, Benjamin Pugsley, and Ayşegül Şahin. 2019. "Demographic Origins of the Startup Deficit." NBER Working Paper 25874.

Lee, Yoonsoo, and Toshihiko Mukoyama. 2015. "Productivity and Employment Dynamics of US Manufacturing Plants." *Economics Letters* 136: 190–93.

Luttmer, Erzo G. J. 2007. "Selection, Growth, and the Size Distribution of Firms." *Quarterly Journal of Economics* 122 (3): 1103–44.

Luttmer, Erzo G. J. 2011. "On the Mechanics of Firm Growth." *Review of Economic Studies* 78 (3): 1042–68.

MaCurdy, Thomas E. 1982. "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis." *Journal of Econometrics* 18 (1): 83–114.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–725.

Midrigan, Virgiliu, and Daniel Yi Xu. 2014. "Finance and Misallocation: Evidence from Plant-Level Data." *American Economic Review* 104 (2): 422–58.

Perla, Jesse. 2015. "Product Awareness, Industry Life Cycles, and Aggregate Profits." Unpublished.

Pugsley, Benjamin Wild, and Ayşegül Şahin. 2019. "Grown-up Business Cycles." *Review of Financial Studies* 32 (3): 1102–47.

Schoar, Antoinette. 2010. "The Divide between Subsistence and Transformational Entrepreneurship." In *Innovation Policy and the Economy*, Vol. 10, edited by Josh Lerner and Scott Stern, 57–81. Chicago: University of Chicago Press.

Sedláček, Petr. 2020. "Lost Generations of Firms and Aggregate Labor Market Dynamics." *Journal of Monetary Economics* 111: 16–31.

Sedláček, Petr, and Vincent Sterk. 2017. "The Growth Potential of Startups over the Business Cycle." *American Economic Review* 107 (10): 3182–210.

Sedláček, Petr, and Vincent Sterk. 2019. "Reviving American Entrepreneurship? Tax Reform and Business Dynamism." *Journal of Monetary Economics* 105: 94–108.

Sterk, Vincent ⓕ Petr Sedláček ⓕ Benjamin Pugsley. 2021. "Replication Data for: The Nature of Firm Growth." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E121021V1.