THE NATURE OF FIRM GROWTH*

Benjamin W. Pugsley
University of Notre Dame

Petr Sedláček
University of Oxford & CEPR

Vincent Sterk
University College London & CEPR

September 28, 2018

Abstract

Only half of all startups survive past the age of five and surviving businesses grow at vastly different speeds. Using micro data on employment in the population of U.S. businesses, we estimate that the lion's share of these differences is driven by ex-ante heterogeneity across firms, rather than by ex-post shocks. Using a macroeconomic model with firm dynamics, we then study how ex-ante differences shape the distribution of firm size, "up-or-out" dynamics, and the associated gains in aggregate output. "Gazelles"—a small subset of startups with particularly high growth potential—emerge as key drivers of these outcomes. Analyzing changes in the distribution of ex-ante firm heterogeneity over time reveals that the birth rate and growth potential of gazelles has declined, creating substantial aggregate losses.

Keywords: Firm Dynamics, Startups, Macroeconomics, Administrative Data
JEL Codes: D22, E23, E24

*First version December 2017. We thank Costas Arkolakis, Eric Bartelsman, Christian Bayer, Mark Bils, Richard Blundell, Vasco Carvalho, Steven Davis, Chris Edmond, Doireann Fitzgerald, Xavier Gabaix, Urban Jermann, Greg Kaplan, Pawel Krowlikowski, Erzo Luttmer, Ezra Oberfield, Fabien Postel-Vinay, Michael Siemer, Kjetil Storesletten, Emily and Robert Swift as well as seminar and conference participants at the Barcelona GSE summer forum, the 2017 CAED conference in Seoul, Cambridge University, University of California Davis, CIRPÉE/CIRANO/CIREQ Montréal, Cornell University, Duke University, Federal Reserve Bank of Cleveland, Federal Reserve Bank of San Francisco, Federal Reserve Board, University of Melbourne, Monash University, Ifo Munich, University of Kent, University of Nottingham, University of Sydney, Tilburg University, 2017 SED meetings in Edinburgh, University of Queensland, 2018 NBER Macro Perspective meetings, and 2018 Minnesota Macro workshop for helpful comments. We also are grateful for excellent research assistance by Harry Wheeler. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

1 Introduction

High-growth firms are widely seen as pivotal contributors to economic prosperity, if only for the large number of jobs that they create, see e.g. Haltiwanger, Jarmin, Kulick, and Miranda (2016). But what is it that distinguishes such firms from others that stay small throughout their lives? One view 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 in the growth profiles of startups. Some types of startups are poised for growth, for example due to a highly scalable technology or business idea, whereas others are destined to stay small. Although both views seem plausible, there is little empirical evidence on the relative importance of the two in shaping firm dynamics.

While their origins are not yet fully understood, firm dynamics have long been recognized in the literature as a key determinant of macroeconomic outcomes (Hopenhayn and Rogerson (1993), Melitz (2003), Klette and Kortum (2004)). More recently, Decker, Haltiwanger, Jarmin, and Miranda (2016) have documented a downward trend in the skewness of firm growth rates, and put forward the idea that a disappearance of high-growth firms might have driven the slump in U.S. employment and productivity growth, observed over the last decade. However, the origins and implications of this trend are still unclear. On the one hand, the U.S. may no longer offer a fertile ground for entrepreneurs to create high-potential startups, founded on ambitious business models. Clearly, this would have important repercussions for the U.S. macro economy. On the other hand, the trend may reflect a mere change in the distribution of ex-post shocks faced by individual firms, which might not leave lasting effects on the economy.

This paper uses the Longitudinal Business Database (LBD), an administrative panel covering nearly all private employers in the United States from 1976 to 2012, to dissect the firm growth process and its changes over time. We follow startups for twenty years after they enter and estimate the extent to which observed differences across firms are driven by ex-ante heterogeneity and to what extent they are formed by ex-post shocks.[1] We do so using both a reduced-form statistical model and a structural macroeconomic model with firm dynamics, both of which allow for

[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).

heterogeneous ex-ante profiles as well as different types of ex-post shocks. While the reduced-form approach has the benefit of simplicity and yields analytical formulas which help us understand the identification of the key parameters, the structural firm dynamics framework offers many complementary advantages. In particular, it accounts for endogenous selection of firms. Moreover, it allows us to understand how firm heterogeneity impacts the aggregate economy and how various frictions, such as imperfect information or adjustment costs, affect our results. Finally, using this framework we can distinguish between types of startups with different ex-ante growth and survival profiles, analyze how their prevalence has changed over time, and quantify the aggregate consequences of such changes.

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). Perhaps surprisingly, the literature on firm dynamics does not have a similar tradition. To the best of our knowledge, the basic autocovariance structure of employment by age has not been systematically documented. Instead, the firm dynamics literature has emphasized the age profiles of average size and exit, see e.g. Haltiwanger, Jarmin, and Miranda (2013), Hsieh and Klenow (2014) and Akcigit, Alp, and Peters (2017).[2] While we also target these important moments in the structural model, we highlight the wealth of additional information that is embodied in the autocovariance structure.

A key finding of our study is that ex-ante heterogeneity accounts for a large share of the cross-sectional dispersion in employment, conditional on age. In the first year after entry, ex-ante heterogeneity accounts for more than ninety percent of the cross-sectional dispersion in employment. Even after *twenty* years, ex-ante factors still explain about forty percent of the cohort's employment dispersion.

This finding relates to several earlier studies. Abbring and Campbell (2005) use data on newly created bars in Texas to estimate an industry model and find that pre-entry decisions account for about 40 percent of the variation of sales in the first year. Campbell and De Nardi (2009) and Hurst and Pugsley (2011) present survey evidence that many nascent entrepreneurs do not expect their business to grow large.[3]

[2] Cabral and Mata (2003) also document the evolution of the skewness of the size distribution with age.

[3] Schoar (2010) makes a distinction between "subsistence" and "transformational" entrepreneur-

Sedláček and Sterk (2017) document strong cohort effects in firm-level employment, depending on the state of the business cycle in the year of entry.

The structural macroeconomic model with firm dynamics that we employ in our analysis follows the tradition of Hopenhayn (1992), Melitz (2003), and Luttmer (2007), and features endogenous entry and exit and general equilibrium forces. We introduce a multi-dimensional idiosyncratic process into this framework, to allow not only for persistent and transitory ex-post shocks, but also for heterogeneity in ex-ante growth and survival profiles. This relatively rich process aligns with the reduced-form evidence and is necessary to obtain a good fit with the empirical autocovariance structure. Therefore, our empirical evidence points towards models allowing for ex-ante differences in growth across firms, along the lines of e.g. Luttmer (2011).

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. Most importantly, ex-ante differences still emerge as the key source of heterogeneity.

After taking the structural model to the data, we show that ex-ante heterogeneity is not only an important determinant of size dispersion, but also of the well-documented "up-or-out" dynamics. That is, the fact that many young firms shut down while surviving businesses grow quickly is in large part driven by ex-ante heterogeneity. The impact of this materializes via selection on ex-ante growth profiles: firms with little growth potential exit, allowing firms with high potential to blossom. Indeed, we find that selection on ex-ante heterogeneity, as well as its interaction with ex-post shocks, makes the age profile of average size substantially more upward-sloping. Associated with this steeper slope is a large gain in aggregate output. By contrast, ex-post shocks alone create only small selection effects and hence by themselves matter little for aggregate output.

We also examine specifically the contribution of startups with high growth potential, known as "gazelles" in the literature. The model allows us to back out the distribution of ex-ante growth profiles, which we exploit to identify a subset of high-

ship in this regard. Guzman and Stern (2015) and Belenzon, Chatterji, and Daley (2017) also show that firm growth is partly predictable based on observable characteristics at the time of startup.

potential startups with projected annual growth of more than twenty percent in the first five years after entry.[4] We find that such *ex-ante* gazelles account for only about five percent of all startups. Nonetheless, they contribute greatly to aggregate job creation being responsible for much of the positive slope of the age profile of average firm size and to the associated gains in aggregate output.

Finally, we use the model to understand the sources and aggregate consequences of an apparent structural change in the growth dynamics of U.S. firms. After splitting our data sample in half, we find that both the autocovariance matrix and the age profile in firm exit have remained stable. However, what has changed is the profile of average size by age. This profile has flattened, implying less growth on average in the recent sample. This finding relates to the cross-country evidence presented by Hsieh and Klenow (2014) who document that the average size profile in India and Mexico is much flatter than in the U.S., and associate this with large differences in aggregate productivity.

To study the underlying changes and their implications directly, we re-estimate the model on the two subsamples. We can then evaluate how the growth profiles of startups have changed over time and what aggregate implications these changes entail. The results show that the prevalence of ex-ante gazelles in the population of startups has substantially declined over time, confirming the concerns raised by Decker, Haltiwanger, Jarmin, and Miranda (2016). In addition, we find that the growth profile of gazelles has flattened beyond age fifteen. These changes together account for about half of the flattening of the average size profile across all firms, despite the fact that gazelles make up only a small fraction of all startups. Importantly, the aggregate output losses implied by this change in firm dynamics between the two samples amount to about 4.5 percent, with larger losses to follow if the observed trend continues.

Supporting evidence for the conclusion that changes in ex-ante factors are a key driver of the observed decline in business dynamism is contained in the cohort structure of the firm size distribution. In particular, if the nature of firm growth has changed over time because of ex-ante factors, we should observe different patterns across different *cohorts* of startups rather changing patterns among all firms across time. Indeed, dissecting the firm size profile by startup cohorts reveals that firms

[4] Previous analysis of gazelles, e.g., Birch and Medoff (1994), were only able to identify gazelles by their realized growth paths.

entering since the late 1980's have a substantially lower growth profile compared to earlier cohorts.

Our findings provide a new avenue for understanding aggregate outcomes. A vast and growing literature seeks to understand macroeconomic performance by studying, at the micro level, how existing firms respond to changes in economic conditions and government policies. Our results, by contrast, assign a crucial role to the types of new startups that are being created. There is still much to be learned about how the distribution of startup types affects –and is affected by– the economic environment.

The remainder of this paper is organized as follows. Section 2 presents the data, the reduced-form model, and initial estimates of the importance of ex-ante heterogeneity for size dispersion. Section 3 describes the structural firm dynamics model and the parametrization procedure. Baseline results from the structural model are presented in Section 4, after which Section 5 presents the results from the split-sample analysis. Finally, Section 6 concludes.

2 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. Using a statistical 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 is formed by ex-post shocks. We begin by describing our data set and the central piece of empirical evidence used in the estimation: the autocovariance function of logged employment, at both the establishment- and 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.

2.1 Data

The analysis is based on administrative micro data on employment in the United States, taken from the from Census Longitudinal Business Database (LBD). The data cover almost the entire population of employers over the period between 1979 and 2012. We construct a panel of log employment at the establishment- and firm-level in the year of startup (age zero) up to age nineteen.[5,6] 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 4-digit level. In order to streamline the discussion, we will use the term "business" whenever we refer to both establishments and firms.

2.2 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 panels) and autocorrelations (right panels). The figure presents this information for both establishments (top panels), and for firms (bottom panels). Finally, since businesses may exit at any age, we display patterns for a balanced panel (solid line) that includes only businesses that survive for at least 20 years and for an unbalanced panel (dashed line) that includes all businesses in our data set.[7] Interestingly, business exit affects essentially only the cross-sectional employment dispersion by age; the autocorrelations are remarkably similar across the balanced and unbalanced panels.

Let us first focus on the cross-sectional standard deviations by age, shown in the left panels. Standard deviations are between 1 and 1.4 log points for both establishments and firms, 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 panels. The latter indicates that the observed fanning out of the size distribution with age is not purely driven by selective exit of certain businesses.[8]

The right panels of Figure 1 depict the cross-sectional correlations of logged employment between age a and an earlier age $h < a$. Keeping h fixed, the autocorre-

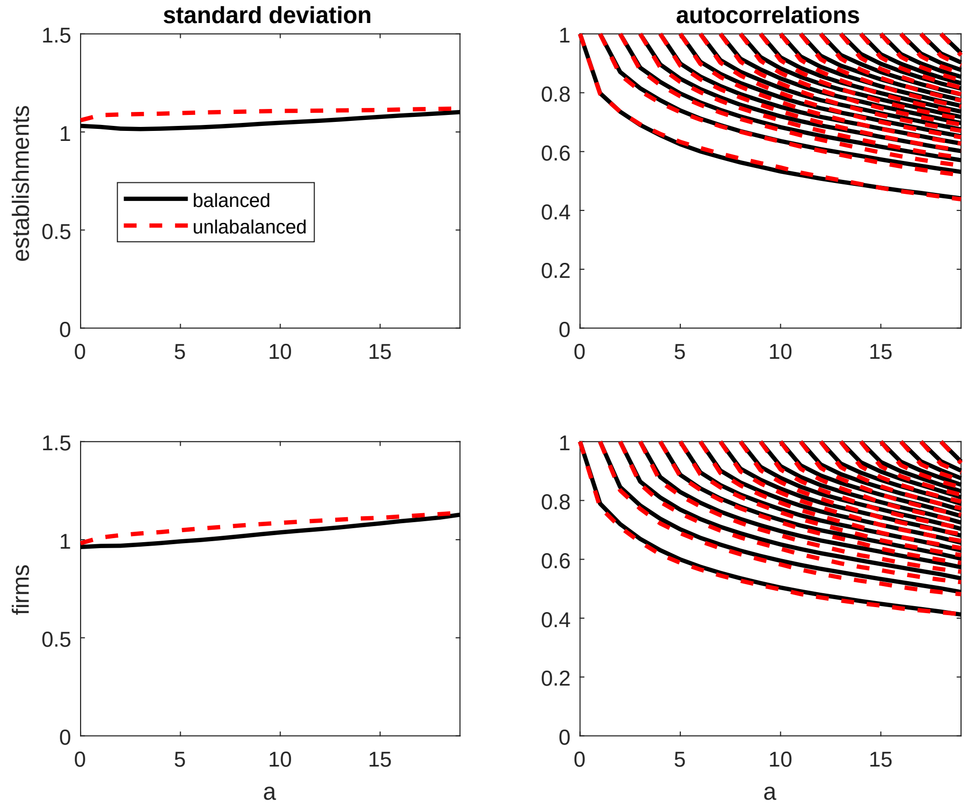
[5]Establishments are the physical units of a firm, located at a specific addresses. A firm can consist of one or multiple establishments. The data are a snapshot taken in the month of March of each year. The age of an establishment is computed as the current year, minus the first year an establishment came into existence. The age of a firm is computed as the age of its oldest establishment.

[6]For brevity, we omit a full description of the panel construction from the LBD microdata from the main text. Please refer to Appendix A for a detailed description of the LBD, the establishment- and firm-level longitudinal links and the construction of the autocovariance matrix.

[7]We focus primarily on the balanced panel of firms for our main results, although we present both here. In Appendix A.2 we describe in more detail the differences in their measurement as well as present the full autocovariance matrices for both panels.

[8]The exception to this pattern is the flat age profile of cross-sectional dispersion for establishments below age five in the balanced panel.

Figure 1: Standard deviations and autocorrelations of log employment by age



Note: The left panels show cross-sectional standard deviations of log employment by age (a) for establishments (top left panel) and firms (bottom left panel). The right panels show cross-sectional correlations of log employment between ages a and age $h \leq a$ for establishments (top right panel) and firms (bottom right panel). “Balanced” refers to a panel of establishments (firms) which survived at least up to age 19, while “unbalanced” refers to a panel of all establishments (firms).

lations decline with age a. For instance, while the autocorrelation between logged employment at ages zero and ten is 0.55, the autocorrelation between ages zero and nineteen is 0.44. 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 zero and age nine is 0.56, whereas the corresponding correlation between age ten and nineteen 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.

2.3 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 includes both sources of heterogeneity. The model nests as special cases reduced-form representations of several prominent structural firm dynamics models, such as those of Hopenhayn and Rogerson (1993) and Melitz (2003), while at the same time being flexible enough to fit the observed autocovariance structure well. Appendix B.3 estimates numerous alternative model specifications (including conventional panel data models) showing that our specification, which is grounded in existing firm dynamics models, 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 business i at age a and consider the following process for this variable:

$$\begin{aligned} \ln n_{i,a} &= \ln n_{i,a}^{EXA} + \ln n_{i,a}^{EXP}, && && (1) \\ \ln n_{i,a}^{EXA} &= u_{i,a} + v_{i,a}, && \text{(ex-ante component)} \\ \ln n_{i,a}^{EXP} &= w_{i,a} + z_{i,a}, && \text{(ex-post component)} \end{aligned}$$

where

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

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

In the above process, $\ln n_{i,a}^{EXA} = u_{i,a} + v_{i,a}$ captures the *ex-ante* profile, which is governed by three business-specific parameters that are drawn independently just prior to startup, i.e. at age $a = -1$. The parameter θ_i is a permanent component which is allowed to accumulate gradually with age at a speed governed by ρ_u. The second and third parameter, $u_{i,-1}$ and $v_{i,-1}$, represent two initial conditions. The former allows for the possibility that the path of the ex-ante component starts away from zero. The latter, which is allowed to die out at its own speed ρ_v, enables the curvature of the ex-ante profile to vary over the lifecycle.

If $\rho_u < 1$, in the long run the ex-ante component reaches a steady state level given by $\ln n^{EXA}_{i,\infty} = \theta_i/(1-\rho_u)$. Since this level differs across businesses, the process admits heterogeneity in long-run steady states. Moreover, since initial conditions differ across businesses, we allow for heterogeneity in the paths from initial employment towards the steady states. Finally, since the process includes two separate initial conditions, each with their own persistence parameter, it allows businesses to gravitate towards their steady-state levels at different speeds. The implied ex-ante growth profiles therefore allow for rich heterogeneity.[9]

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 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 modelled 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 $\sigma_u = \sigma_\theta = \sigma_z = 0$ and $\rho_v = \rho_w$. 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 $\sigma_u = \sigma_v = \sigma_z = 0$ and $\rho_u = 0$, 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 $\sigma_z > 0$.[10] Our baseline process also

[9]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 towards some expected long-run size, the ex-ante terms would instead characterize heterogeneous growth rates from some initial size.

[10]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

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 (forthcoming).[11]

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

$$\begin{aligned}\mathrm{Cov}\left[\log n_{i,a}, \log n_{i,a-j}\right] = & \underbrace{\left(\sum_{k=0}^{a} \rho_u^k\right)\left(\sum_{k=0}^{a-j} \rho_u^k\right)\sigma_\theta^2 + \rho_u^{2(a+1)-j}\sigma_{\tilde{u}}^2 + \rho_v^{2(a+1)-j}\sigma_{\tilde{v}}^2}_{\text{ex-ante components}} \\ & \underbrace{+\, \sigma_\varepsilon^2 \rho^j \sum_{k=0}^{a-j} \rho_w^{2k} + \sigma_z^2 \mathbf{1}_{j=0}}_{\text{ex-post components}} .\end{aligned} \tag{2}$$

This result is derived in Appendix B.1. The autocovariance function is a non-linear function of the persistence and variance parameters of the components of the

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 appendix B.3). Our results thus caution against the use of standard panel data estimators when applied to employment dynamics of young businesses.

[11] For further discussion, please refer to Appendix B.3 where we consider a number of alternative statistical models both as special cases and further generalizations of equation (1).

underlying process.[12] 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 $|\rho_u|, |\rho_v|, |\rho_w| < 1$ 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:

$$\lim_{a \to \infty} \text{Cov}\left[\ln n_{i,a}, \ln n_{i,h}\right] = \frac{1 - \rho_u^{h+1}}{\left(1 - \rho_u\right)^2} \sigma_\theta^2.$$

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 parameters of the process using a minimum distance procedure, as proposed by Chamberlain (1984). Specifically, we minimize the sum of squared deviations of the upper triangular parts of the autocovariance matrix implied by the process, from its counterpart in the data.[13] Because the size of the LBD ensures each that element of the empirical autocovariance matrix is precisely estimated, we assign equal weights to all elements in the estimation procedure. Throughout, our results apply to the balanced panel data set, although they are similar using the unbalanced panel.[14]

Figure 2 shows that the model fit is very good for both establishments and firms, correctly capturing the convexly declining pattern of the autocovariances in the lag length, given the initial age h, and the concavely increasing pattern in age given the lag length $j > 0$. Finally, the model fits the non-monotonic pattern in cross-sectional dispersion by age.

The corresponding parameter estimates are shown in Table 1. A key feature of

[12]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.

[13]For brevity, we defer a detailed discussion of the estimation procedure to Appendix B.2.

[14]For reference, we include the estimated parameters using the unbalanced panel in Appendix Table B.1 panel A.

Table 1: Parameter estimates from reduced-form model

	Estabs	*Firms*
ρ_u	0.2059	0.2183
	(0.0015)	(0.0018)
ρ_v	0.8415	0.8323
	(0.0010)	(0.0014)
ρ_w	0.9489	0.9625
	(0.0003)	(0.0003)
σ_θ	0.6031	0.5545
	(0.0014)	(0.0015)
$\sigma_{\widetilde{u}}$	2.0461	1.7425
	(0.0174)	(0.0145)
$\sigma_{\widetilde{v}}$	0.7378	0.6951
	(0.0017)	(0.0021)
σ_ε	0.2554	0.2548
	(0.0004)	(0.0004)
σ_z	0.2623	0.2716
	(0.0006)	(0.0006)
$RMSE$	0.0100	0.0120

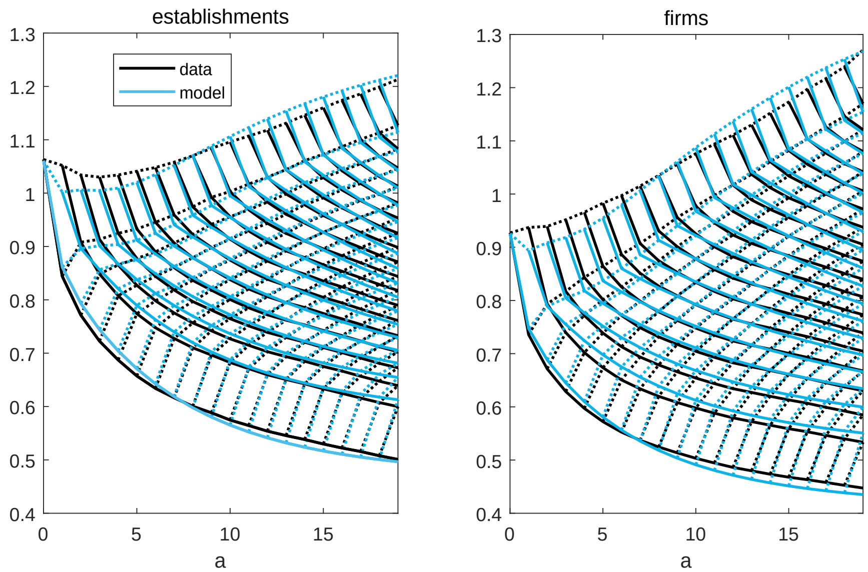
Note: Equally-weighted minimum distance estimates of Equation (1) for both establishments and firms using the balanced panel. See Appendix Table B.1 panel A for estimates using unbalanced panel. $RMSE$ is the square root of the mean squared error, which is the average of the 210 squared residuals of the model generated covariance element relative to the data.

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 0.76 for establishments and 0.71 for firms. These values are substantial when considering that the overall cross-sectional dispersion of twenty year old businesses is about 1.4 (see Figure 1).

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 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 considering the empirical fit of four restricted versions of the model, depicted in Figure 3.

Restricted models I and II (top row) illustrate, respectively, why both permanent

Figure 2: Autocovariance matrices: statistical models versus data



Note: Autocovariance 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 and establishments using the balanced panel.

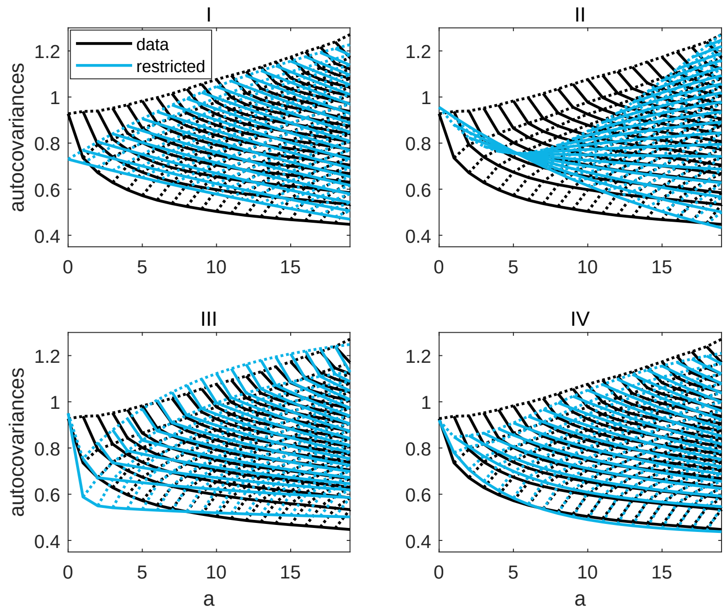
ex-ante heterogeneity and ex-post shocks are needed to match the data. In restricted model I, we show a popular specification in the literature on firm dynamics models, which essentially amounts to an AR(1) process with hetreogeneous initial draws but without heterogeneity in long-run steady states.[15] Because of the latter, all ex-ante heterogeneity vanishes in the long run. We achieve this by imposing $\rho_v = \rho_w$ and $\rho_u = \sigma_\theta = \sigma_{\widetilde{u}} = \sigma_z = 0$, and re-estimating ρ_w, σ_w and $\sigma_{\widetilde{v}}$. Under this specification, the high long-run autocovariances demand very persistent ex-post shocks. This results in the model-implied autocorrelations being almost linear in age, which conflicts with the non-linear patterns in the data.[16] The presence of ex-ante heterogeneity thus relaxes the need for the persistence parameters to be close to one. Indeed, as seen in Table 1, the data strongly reject the presence of a unit root in our sample.[17]

[15]For example, Hopenhayn and Rogerson (1993) consider this process for productivity. In the baseline version of their model without firing costs, the AR(1) for productivity results in an AR(1) for employment.

[16]Note that $\rho_w = 1$ would introduce a random walk component, consistent with Gibrat's law. However, violation of Gibrat's law in the data has been documented in the literature, in particular among new and young firms, see e.g. Haltiwanger, Jarmin, and Miranda (2013).

[17]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

Figure 3: Autocovariance matrices: restricted models



Note: Autocovariance of log employment between age $a = h + j$ and age $h \leq a$ in the baseline (for the balanced firm panel estimates), and in the four restricted models. In Model I ρ_w, σ_w and $\sigma_{\widetilde{v}}$ are estimated, while imposing $\rho_v = \rho_w$ and $\rho_u = \sigma_\theta = \sigma_{\widetilde{u}} = \sigma_z = 0$. In Model II ρ_u, σ_θ and $\sigma_{\widetilde{u}}$ are estimated, while imposing $\rho_w = \rho_v = \sigma_\epsilon = \sigma_{\widetilde{v}} = \sigma_z = 0$. Model III is the baseline with the restriction that $\rho_v = \sigma_{\widetilde{v}} = 0$. Model IV is the baseline with the restriction that $\sigma_z = 0$.

In restricted model II, we shut down all ex-post shocks, allowing only for heterogeneous ex-ante profiles (with only one initial condition). We do so by imposing $\rho_w = \rho_v = \sigma_\epsilon = \sigma_{\widetilde{v}} = \sigma_z = 0$, and re-estimating ρ_w, σ_w and $\sigma_{\widetilde{v}}$. Again, we obtain a poor fit. In particular, this version fails to match the pattern of increasing employment dispersion with firm age, as present in the data.

Restricted model III illustrates why *both* transitory ex-ante components, u and v, are required to match the data. This version is the same as our baseline except that we set $\rho_v = \sigma_{\widetilde{v}} = 0$, and re-estimate the remaining parameters. The bottom left panel of Figure 3 makes clear that 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 explores the role of the iid ex-post shock z. In this version, we re-estimate the model imposing $\sigma_z = 0$. The presence of z somewhat improves the fit of the model, by giving an extra kick to the dispersion of employ-

firm-size distribution. See Appendix B.3 for estimation results from a variant of our model with a unit root.

ment 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 Appendix B.3. 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 improve on model fit without introducing more parameters, and our conclusions about the importance of ex-ante heterogeneity remain unchanged across specifications.

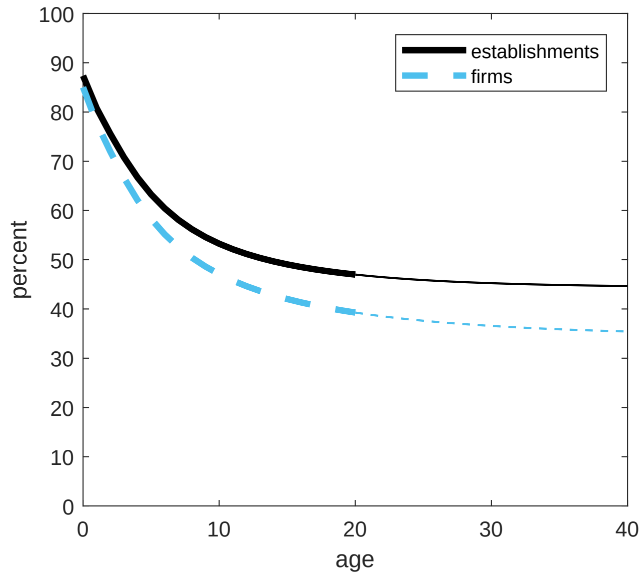
2.5 The importance of ex-ante and ex-post heterogeneity

With the estimated model in hand, we can quantify the relative importance of ex-ante profiles and ex-post shocks for the cross-section 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 zero to nineteen, whereas thin lines represent an extrapolation for businesses at age 20 or above using the point estimates.[18]

Figure 4 shows that for businesses in the year of startup, that is at age zero, the ex-ante component accounts for about 85 percent of the cross-sectional variance in 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 twenty, ex-ante factors account for 47 percent of the size variance among establishments, and around 40 among firms. In the data, more than seventy percent of the businesses are twenty years old or younger. Our results show that, among these businesses, ex-ante factors are a key determinant of size. Increasing

[18]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. This is also reflected in very small standard errors around the point estimates for the parameters, as can be observed from Table 1.

Figure 4: Contribution of ex-ante heterogeneity to cross-sectional employment dispersion



Note: 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 $j = 0$.

age towards infinity, the contribution of ex-ante heterogeneity stabilizes at around 45 percent for establishments and 35 percent for firms. Therefore, even among very old businesses ex-ante factors contribute to a large chunk of the dispersion in size.

As discussed above, Appendix B.3 provides results for a wide range of alternative models. Importantly, regardless of the specified process, ex-ante characteristics explain a significant fraction of early life-cycle employment dispersion. In addition, for the processes that also match the autocovariance structure well, they attribute nearly identical shares of long-run variance to ex-ante characteristics, reinforcing the findings from our baseline model.

3 Structural model

In this section we estimate a structural macroeconomic model with firm dynamics, which has several advantages relative to the reduced-form approach in Section 2. First, the structural model accounts for selective entry and exit. Indeed, much of Section 4 focuses on understanding to what extent ex-ante determined characteristics drive the process of firm selection. Second, the structural model allows us to compute aggregates. We are therefore able to quantify the importance of ex-ante heterogeneity across firms for the *aggregate* economy. Third, micro-founding the process of firm selection allows us to analyze how various frictions impact the observed patterns in the data. Specifically, we show that our results remain intact even in the presence of imperfect information, or different forms of adjustment costs.

Finally, having understood how ex-ante heterogeneity impacts firm growth, selection and in turn the aggregate economy, we use our framework to shed light on the observed changes in firm dynamics, and the associated aggregate outcomes, during the past decades. Throughout the analysis we report results for firms. Estimates for establishments are shown in Appendix C.6.

3.1 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 the reduced-form analysis, 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 $\overline{N}$. Household preferences are given by $\sum\limits_{t=0}^{\infty} \beta^t C_t$, where $\beta \in (0,1)$ is the discount factor. C_t is a Dixit-Stiglitz basket of differentiated goods given by:

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

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.

The household's budget constraint is given by $\int_{i\in\Omega} p_i c_i = W\overline{N} + \Pi$, 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 $c_i = \varphi_i \left(p_i/P\right)^{-\eta} C$, where

P is a price index given $P \equiv \left(\int_{i\in\Omega} \varphi_i p_i^{1-\eta}\right)^{\frac{1}{1-\eta}}$, so that total expenditure satisfies $PC = \int_{i\in\Omega} p_i c_i$.

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 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

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

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:

$$p_i = \frac{\eta}{\eta - 1} W.$$

We let labor be the numeraire 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 $\chi \equiv \frac{(\eta-1)^{\eta-1}}{\eta^\eta}$, and labor demand as $n_i = \varphi_i \left(\frac{\eta}{\eta-1}\right)^{-\eta} w^{-\eta} C + f$. 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:

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

In the above equation $\mathbf{s}_i'$ 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. 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 firms in $\mathbf{S} \in \mathcal{S}$, where is $\mathcal{S}$ is the Borel $\sigma-$algebra generated by $\mathbf{s}$. Given the exit policy, $\mu(\mathbf{S})$ satisfies:

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

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:

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

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

$$
\begin{array}{lclllll}
\ln \varphi_{i,t} & = & u_{i,t} + v_{i,t} + w_{i,t} + z_{i,t} & & & \\
u_{i,t} & = & \rho_u u_{i,t-1} + \theta_i, & u_{i,-1} \sim iid(\mu_{\tilde{u}}, \sigma^2_{\tilde{u}}) & \theta_i \sim iid(\mu_\theta, \sigma^2_\theta) & |\rho_u| < 1 \\
v_{i,t} & = & \rho_v v_{i,t-1}, & v_{i,-1} \sim iid(\mu_{\tilde{v}}, \sigma^2_{\tilde{v}}) & & |\rho_v| < 1 \\
w_{i,t} & = & \rho_w w_{i,t} + \varepsilon_{i,t}, & w_{i,-1} = 0 & \varepsilon_{i,a} \sim iid(0, \sigma^2_\varepsilon) & |\rho_w| < 1 \\
z_{i,t} & \sim & iid(0, \sigma^2_z), & & &
\end{array}
$$

where we re-introduced time indices for clarity. Note that the firm-level state is given by $\mathbf{s}_{i,t} = [u_{i,t}, v_{i,t}, w_{i,t}, z_{i,t}]$. The above process implies that the level of demand faced by a firm is determined by both a idiosyncratic ex-ante profile, captured by $u_{i,t}$ and $v_{i,t}$, as well as ex-post shocks, which enter via $w_{i,t}$ and $z_{i,t}$.

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. In what follows, we consider extensions to the model that relax the assumptions on the perfect information about ex-ante components as well as those on frictionless labor adjustment.

3.2 Discussion and extensions

An imperfect-information perspective. The baseline model assumes full information, in the sense that firms immediately observe stochastic innovations to the components of the shock process. One may wonder to what extent economic agents, including the firm owners, actually have this much information, and to what extent this affects firms' decisions and the interpretation of the empirical patterns that we have documented.

To investigate these issues, we conduct two exercises, shown in Appendix C.3. First, we document that under the baseline estimates, optimal Bayesian updating enables one to learn about the ex-ante profiles extremely quickly. In fact, most of the uncertainty about ex-ante profiles is resolved in the first year upon entry. Nevertheless, this may still distort firms' decisions. Therefore, as a second exercise we consider a version of the model where firms have imperfect information about their ex-ante profiles, similar to e.g. Jovanovic (1982). While in this version, ex-post shocks are attributed a somewhat larger role, our main conclusions remain unchanged.[19]

Adjustment costs. Ex-post demand shocks are a standard feature of firm dynamics models, since demand conditions may change for various reasons that are beyond the control of the firm. Considering ex-ante heterogeneity across firms also has a strong tradition in the literature. While in certain models ex-ante heterogeneity across firms materialize immediately (see e.g. Melitz, 2003), other studies consider a gradual accumulation of difference, for instance through customer base accumulation (see e.g. Arkolakis, 2016; Luttmer, 2011; Drozd and Nosal, 2012; Gourio and Rudanko, 2014; Perla, 2015).

Our baseline model allows for so-called "passive" accumulation of ex-ante differences through the accumulation of the fixed effect θ_i. Importantly, the parameters governing this process will be estimated. Nevertheless, in Appendix C.4 we consider endogenous adjustment costs as an alternative. Specifically, in addition to passive demand accumulation, firms can choose to invest into "active" demand accumulation as in e.g. Foster, Haltiwanger, and Syverson (2016). [20] Importantly, while incorporating endogenous adjustment costs affects the parameter estimates, our main results remain essentially unchanged.

The process of firm selection. As in any firm dynamics model with endogenous entry and/or exit, a key channel via which heterogeneity may impact on aggregate outcomes is the process of firm selection. Since we integrate a multi-dimensional idiosyncratic process into the model, selection occurs along several different competing margins. Importantly, there is no one-to-one mapping between a particular value of demand and the survival probability of the respective firm. For example, a currently small and unprofitable startup may survive with high probability if it has sufficiently

[19]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.

[20]Our baseline model also abstracts from differences in technologies, another form of heterogeneity often considered in the firm dynamics literature. However, given that we match our model to employment data, our model is observationally equivalent to one with heterogeneity in TFP. Moreover, Hottman, Redding, and Weinstein (2016) and Foster, Haltiwanger, and Syverson (2016) have recently investigated the relative importance heterogeneity in demand versus technology. They conclude that demand factors are a major driver of heterogeneity in the data.

promising long-run growth potential and only faces poor initial conditions or ex-post shocks. A large part of our analysis is devoted to thoroughly understanding the sources, and (aggregate) consequences, of the process of firm selection.

3.3 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 annual real interest rate of about four 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 Appendix C.1. Again, we target 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 the age profiles of the exit rate and average size (in an unbalanced panel). In doing so, we assume that all shock innovations are drawn from normal distributions and we normalize the level parameters μ_u and μ_v to zero. In contrast to the reduced-form setup, we further assume that $\rho_v = \rho_w$, which eases the computational burden substantially.[21]

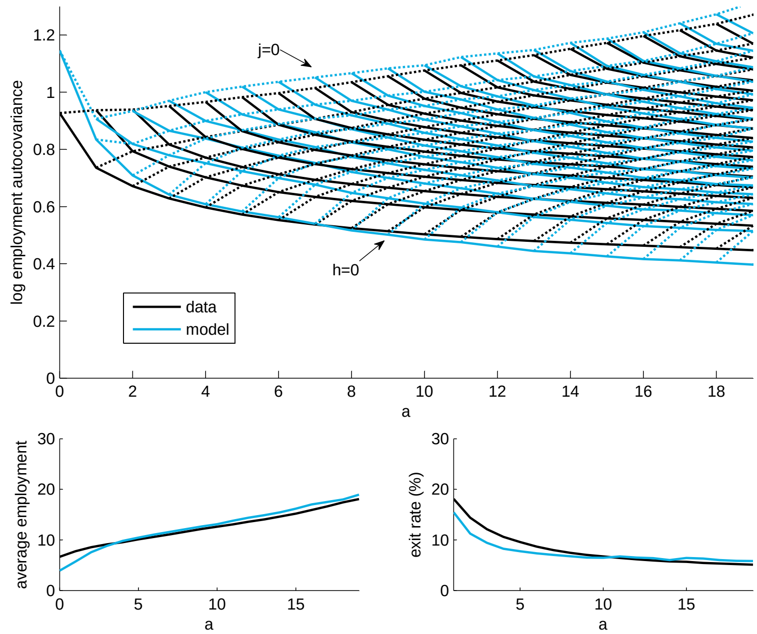
Figure 5 illustrates how the model fits the data. The upper panel shows the autocovariance matrix, while the lower left and right panels show the age size and exit profiles, respectively. Overall, the model provides a good fit of the three sets of empirical comments, considering that 10 parameters are used to target 249 moments.

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.[22] The only exception is the employment share of very large old

[21]This restriction reduces the number of state variables as firms no longer need to keep track of $w_{i,t}$ and $v_{i,t}$ separately. 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.

[22]The model also provides a similarly good fit of the fractions of firms in each of these bins (not

Figure 5: Targeted moments: data and structural model (firms)



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.

firms which is somewhat understated in the model compared to the data. In order to analyze whether this impacts our results, Appendix C.5 re-calibrates the model while in addition to our baseline targets we also explicitly target the firm size distribution. Importantly, our conclusions remain unaltered.

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 half 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. Below, however, we will quantify the model's implications for the importance of ex-ante heterogeneity and make a

shown).

Table 2: Parameter values (firms)

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, st. dev.	1.304
$\sigma_{\tilde{u}}$	initial condition u_{-1}, st. dev.	1.572
$\sigma_{\tilde{v}}$	initial condition v_{-1}, st. dev.	1.208
σ_ϵ	transitory shock ϵ, st. dev.	0.307
σ_z	noise shock z, st. dev.	0.203
ρ_u	permanent component, persistence	0.393
ρ_v	transitory component, persistence	0.988

Note: 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.

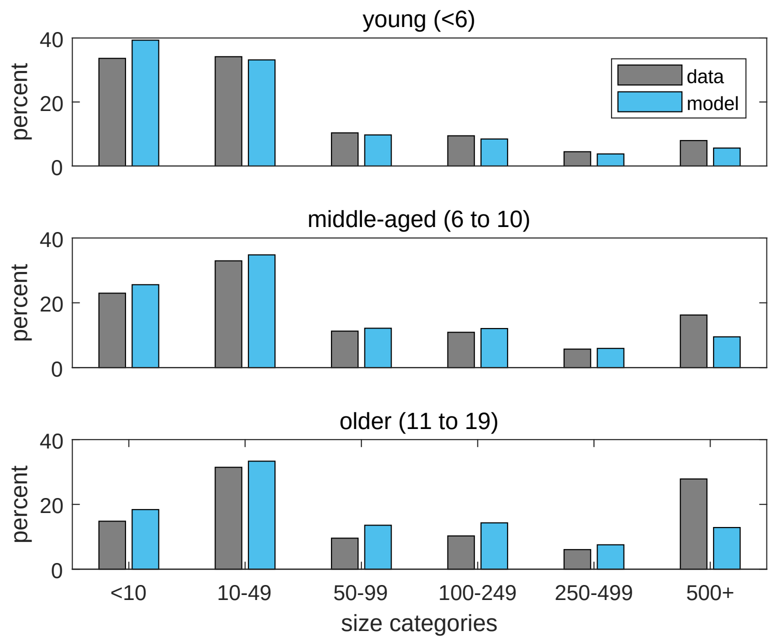
direct comparison to the reduced-form model along this dimension.

4 Importance of ex-ante vs. ex-post heterogeneity

In this section we use the structural model to study the importance of ex-ante heterogeneity for firm dynamics and the aggregate economy. We focus particularly on the process of firm selection and whether it should be thought of as a process that sifts out firms with high ex-ante potential or rather as a process that reflects idiosyncratic business risk. This distinction will also be especially important in Section 5 to understand the secular trends in firm dynamics.

We begin this section by analyzing how our quantification of the importance of ex-ante heterogeneity may be impacted by the presence of endogenous firm selection. We do so first by showing its impact on the variance-decomposition of employment dispersion by firm age introduced for the statistical model in Section 2.5. As a second step, we analyze the drivers of the process of firm selection and show that to a large extent firms endogenously shut down because of their ex-ante characteristics rather

Figure 6: Employment shares of different age/size bins: model versus data (firms)



Notes: Employment shares by firm age and size (employment). Values are expressed as percentages of total employment in firms between 0 to 19 year old firms, 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.

than because of unfavorable ex-post shocks. Finally, given that ex-ante characteristics turn out to be important drivers of the process of firm selection, we analyze their influence on firm growth and aggregate outcomes. Differences in ex-ante characteristics emerge as quantitatively important sources of average life-cycle profiles and aggregate gains from selection.

4.1 Selection and the importance of ex-ante heterogeneity

We revisit the importance of ex-ante heterogeneity for the cross-sectional dispersion in employment, conditional on age where now our structural model allows us to quantify the influence of selection. Defining $\chi \equiv ((\eta - 1)/\eta)^{\eta} w^{-\eta} Y$, the employment level of firm i can be expressed as:

$$n_i = \chi \varphi_i^{EXA} \varphi_i^{EXP}, \tag{4}$$

where $\varphi_i^{EXA} = e^{u_i + v_i}$ is the ex-ante component of demand and $\varphi_i^{EXP} = e^{w_i + z_i}$ is the ex-post component. As in the reduced-form exercise, we can now compute the contribution of ex-ante heterogeneity to the cross-sectional variance of employment by shutting down variation in φ_i^{EXP}. In contrast to the reduced-form model, however, the ex-ante and ex-post component are no longer orthogonal, due to a, typically negative, correlation induced by endogenous 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. Accounting for this correlation, we decompose the variance of logged employment as:

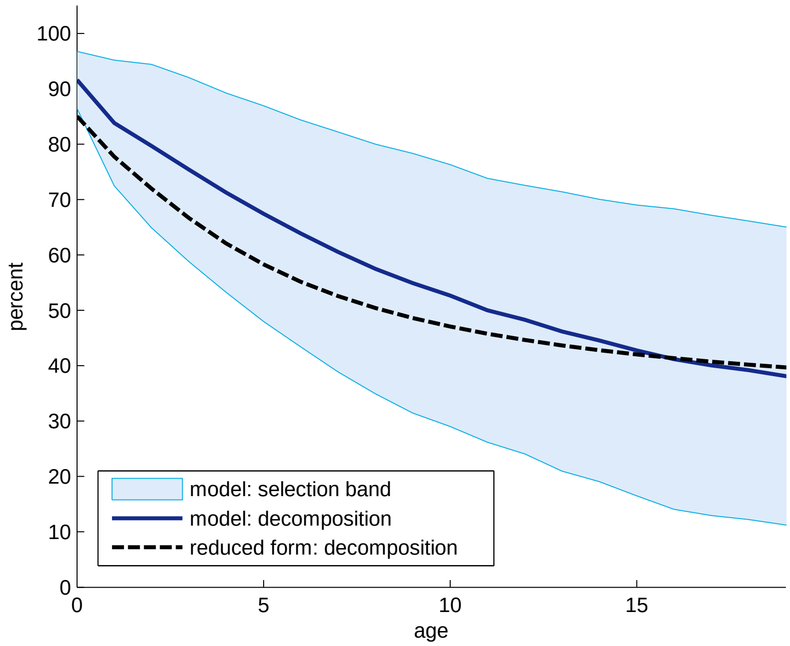
$$\begin{aligned} Var\left(\ln n_i\right) &= Var(\ln \varphi_i^{EXA}) + Var(\ln \varphi_i^{EXP}) + 2Cov(\ln \varphi_i^{EXA}, \ln \varphi_i^{EXP}), \\ &= Cov(\ln \varphi_i^{EXA}, \ln n_i) + Cov(\ln \varphi_i^{EXP}, \ln n_i). \end{aligned} \tag{5}$$

In the reduced-form model, the covariance term $Cov(\ln \varphi_i^{EXA}, \ln \varphi_i^{EXP})$ in the first equality is zero, due to the assumption of independently distributed shocks. However, in the structural model, selection induces a non-zero covariance term which, as mentioned above, tends to be negative. We therefore decompose the variance according to the second equality in Equation (5).[23] Figure 7 depicts the contribution of ex-ante heterogeneity in the structural model (solid line), i.e. $Cov(\ln \varphi_i^{EXA}, \ln n_i)/Var\left(\ln n_i\right)$, together with the reduced-form decomposition (dashed line). Both decompositions attribute a similarly large fraction of size dispersion to ex-ante heterogeneity, at any age. The slight differences between the structural model's solid line and the dashed line for the results of the reduced-form decomposition, especially for younger firms, reflects the effects of endogenous selection.

Figure 7 also plots a "selection band" based on the first equality in Equation (5). This band is constructed by attributing, in turn, the covariance term either fully to the ex-ante component or fully to the ex-post component. This gives us a sense of how much selection matters in the model. While the structural model re-establishes 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. Given its importance, we now analyze the sources and consequences of the process of firm selection.

[23]Note that when $Cov(\ln \varphi_i^{EXA}, \ln \varphi_i^{EXP}) = 0$, it holds that $Var(\ln \varphi_i^{EXA}) = Cov(\ln \varphi_i^{EXA}, \ln n_i)$ and $Var(\ln \varphi_i^{EXP}) = Cov(\ln \varphi_i^{EXP}, \ln n_i)$. The decomposition then exactly coincides with the one we used in the reduced-form analysis.

Figure 7: Contribution of ex-ante heterogeneity to cross-sectional employment dispersion (firms)



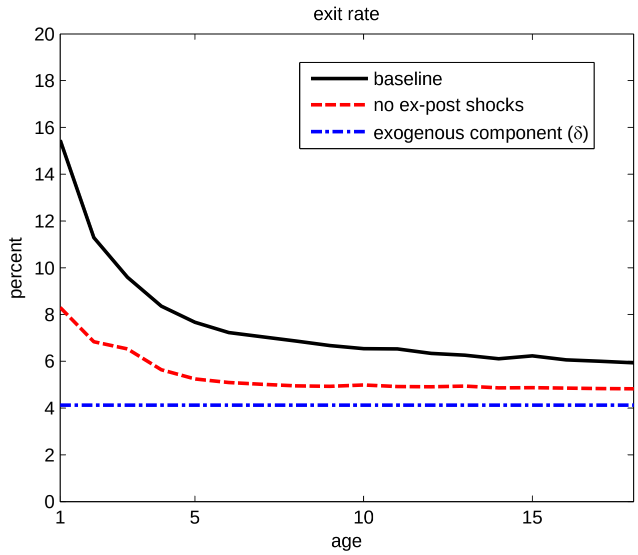
Note: Contributions of ex-ante heterogeneity to the total cross-sectional variance of log employment by age. "Reduced-form" refers to the estimates from Figure 4, "model: covariance decomposition" is the decomposition based on the second line in Equation (5). The shaded areas ("model: selection band") is constructed based on the first equality in Equation (5) by attributing, in turn, the term $2Cov(\ln \varphi_i^{EXA}, \ln \varphi_i^{EXP})$ fully to the ex-ante component and to the ex-post component.

4.2 Sources of firm selection

To what extent is firm selection the purge of businesses with low ex-ante growth potential, and to what extent is it a product of idiosyncratic business risk? In our structural model, firm selection is a multifaceted process that is affected both by a firm's current fundamentals and by its expectations about how these fundamentals will evolve in the future. This evolution is in turn driven by the relative importance of the ex-ante profile and ex-post shocks.

Therefore, to quantify the importance of ex-ante heterogeneity for exit decisions, we run a counterfactual simulation in which we use the firms' baseline decision rules but we completely shut down ex-post shocks to demand $\epsilon_{i,a} = z_{i,a} = 0$ for all i and a. We do, however, preserve exogenous exit shocks. The resulting average exit rate profile is therefore informative about the extent to which firms shut down because of idiosyncratic risk and to what extent firm exit is driven by ex-ante characteristics. For example, firms may have declining ex-ante demand profiles because of favorable

Figure 8: Exit rates (firms)



Note: exit rates by age in the baseline model, in the counterfactual economy with selection only on ex-ante profile, and in the counterfacual economy with only exogenous exit, i.e. exogenous rate δ.

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.

Figure 8 shows the age profile of the exit rate in this counterfactual with only ex-ante heterogeneity, together with the exit profile in the baseline model, and the exogenous component of the exit rate, δ. The difference between the baseline exit profile and the constant exogenous exit rate is the endogenous component of the exit rate, i.e. the part that is driven by selection on both ex-ante and ex-post heterogeneity.

As expected, the exit rate is lower without ex-post demand shocks. However, endogenous firm selection remains and the exit rate in the counterfactual declines with age, as in the baseline. We can interpret the difference between the exit rate without ex-post shocks and the exogenous exit rate δ as the amount of endogenous exit that is driven by selection on ex-ante profiles. Figure 8 then implies that, depending on age, between 30 and 45 percent of overall endogenous exit is driven by selection on ex-ante profiles. Thus, we find that the process of firm selection is to a large extent driven by ex-ante heterogeneity.

4.3 Ex-ante heterogeneity, firm growth and aggregate outcomes

Having shown that ex-ante profile heterogeneity plays an essential role in the process of firm selection, we now turn to examining its effects on firm growth. Specifically, we quantify the extent to which ex-ante heterogeneity shapes the age-profile of average firm size and in turn its effects on aggregate outcomes.

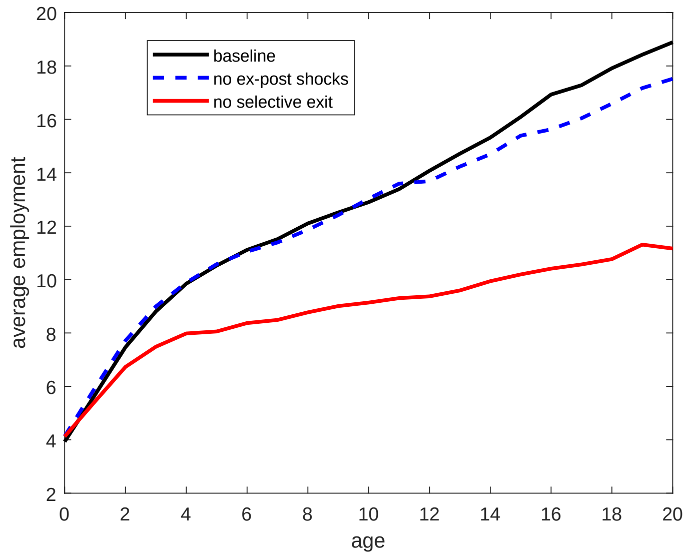
Firm growth. To examine the effect firm selection has on firm growth, we make use of the same counterfactual exit policy functions as in the previous subsection. First, we consider a counterfactual simulation in which firms are not allowed to shut down endogenously, i.e., they exit only at constant exogenous rate δ. The average growth profile of this cohort of startups is depicted in Figure 9 as "no selective exit". In contrast to this counterfactual, the "baseline" employment profile is much steeper. In the absence of firm selection, the average size of twenty year old firms is 50 percent lower compared to the baseline economy. This is exactly because unprofitable (small) firms are allowed to shut down in the baseline economy.

To what extent is this selection-induced gap between the two employment profiles driven by selection on ex-post shocks and to what extent is it determined by firms shutting down because of ex-ante fundamentals? To answer this, we conduct the same exercise as in Figure 8, but this time for firm size. In other words, we retain the baseline employment policy rules, but we use the exit policy determined only by ex-ante profiles.[24] The resulting average employment profile is depicted in Figure 9 as "no ex-post shocks".

The figure clearly shows that selection on ex-post shocks accounts for only a very small share of the overall gains from selection. We thus find that ex-ante heterogeneity is not only an important driver of dispersion in size, but also of the age profiles of exit and average size, especially among younger firms. Thus, up-or-out dynamics largely reflect the separation of firms with high and low long-run growth potential. An important driver of differences in up-or-out dynamics across countries or different time periods within a country might therefore reflect differences in the types of startups that enter the economy. We will return to this issue in the next section.

[24] Note that in this counterfactual is for the same set of firms as in the baseline, i.e. we retain the baseline exit rules which do depend on both ex-ante and ex-post shocks.

Figure 9: Average size and selection (firms)



Note: Average size, unbalanced panel and by age, in the baseline in three counterfactuals. See the main text for a description of these counterfactuals.

Aggregate output. We now explore some of the aggregate implications of our findings. For this purpose we use the same counterfactuals, in combination with the following expression for aggregate output:

$$Y = \Omega^{\frac{\eta}{\eta-1}} \chi^{\frac{1}{1-\eta}} \overline{n}^{\frac{\eta}{\eta-1}},$$

where $\overline{n}$ is the average size across all firms; see Appendix C.1 for a derivation. We re-compute average firm size in both of the counterfactuals described in the previous paragraphs (i.e. when there is no selective exit and when ex-post demand shocks are "switched off"). Using these values, we then compute aggregate output based on the above equation.[25] As is well known, the productivity gains from selection are large. With only exogenous exit, output is 38 percent lower than the baseline. However, shutting down exit on ex-post shocks, i.e., where selection depends on *only* the ex-ante profiles, reduces output by just 4 percent.

These results imply that up-or-out dynamics are indeed an important contributor to aggregate output. Moreover, a key factor driving these dynamics is selection based on firms' ex-ante growth profiles. By contrast, ex-post shocks alone matter

[25]Note again that these are partial-equilibrium counterfactuals since we do not recompute χ and Ω.

relatively little, especially at younger ages. Note further that our counterfactual exercises are based on distributions conditional on firm entry, i.e. based on demand fundamentals of firms which have already *decided* to begin operating. The impact of ex-ante heterogeneity would likely be even larger if selection before entry were to be included in the counterfactuals.

High-potential startups Our estimates show a large amount of heterogeneity in ex-ante profiles: some high-potential startups are on steep ex-ante age profiles of demand growth, whereas others are on flat or even downward-sloping age profiles.

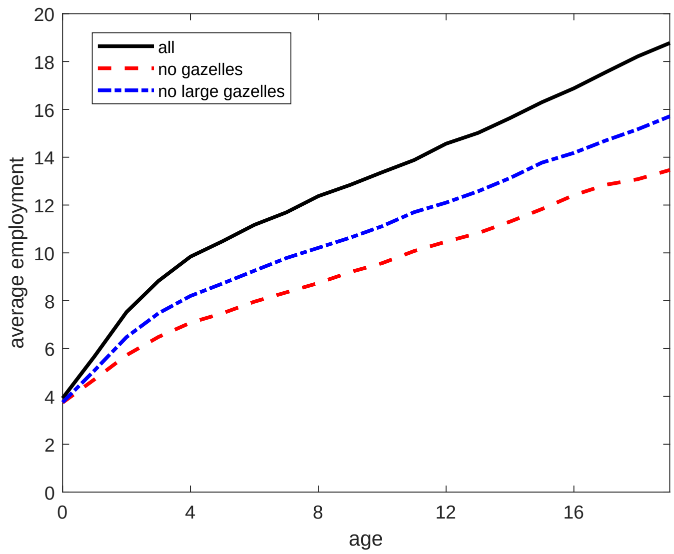
We now quantify the importance of high-growth startups. Such firms, labeled "gazelles" since Birch and Medoff (1994), have been emphasized in the literature as important engines of aggregate job creation. We classify firms according to their ex-ante growth profiles, i.e. those that firms would follow in the absence of ex-post shocks. We then define gazelles as those startups with an ex-ante projected growth rate of at least 20 percent annually, over the first five years, and an expected employment level that exceeds 10 workers at some point during their lifetimes.[26]

While our definition of gazelles is in line with the literature, we differ from existing studies in an important way: 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 observed in the data. Thus, the gazelles as defined in the literature include firms which at startup were not expected to grow very much, but ex post were hit by favorable shocks and grew as a result. It then follows almost by construction that gazelles contribute disproportionately to aggregate job creation. By contrast, in our definition firms that grow just because of favorable ex-post shocks are not counted as gazelles. A priori, it then becomes less clear whether gazelles will contribute disproportionately to job creation or not.

The results of our classification show that gazelles account for only 5.4 percent of all startups. However, their impact on the average size profile, and in turn on aggregate outcomes, is much larger. To understand this, we re-compute the average size profile leaving out the gazelles, see Figure 10. Without gazelles, average size is considerably lower and the difference remains large up to at least age twenty. At that age, average size is more than 25 percent lower than in the baseline. In a second

[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.

Figure 10: The importance of high-potential startups (firms)



Note: Average size, unbalanced panel and by age, in the baseline in two counterfactuals. See the main text for a description of these counterfactuals.

counterfactual we leave out only "large gazelles", which are defined as gazelles with a *startup* size of at least 10 workers. In this counterfactual, average size is about 15 percent lower at age twenty compared ot the baseline, despite large gazelles accounting for only about 1 percent of all startups.

These counterfactuals make clear that high-potential startups are indeed important contributors to aggregate output and employment. Moreover, it follows that seemingly small shifts in the distribution of ex-ante profiles of startups may have large consequences, as suggested also by Sedláček and Sterk (2017). Our results further provide a perspective on the findings of Hsieh and Klenow (2014), who report that average size profiles are much flatter in India and Mexico than in the United States. A flat profile can indicate that there are few startups that operate a high-potential business model, or that high-potential startups have relatively low chances of survival.

5 Changes in the firm dynamics process

The previous section documented that ex-ante heterogeneity is a crucial determinant of not only firm life-cycle dynamics but also of aggregate outcomes. In light of these findings, we now investigate the changes in firm dynamics observed over the past decades. Such changes have attracted attention in the light of the disappointing evolution of employment and productivity growth in the U.S. over the last ten to fifteen years. A disconcerting trend over the same period, noted by Decker, Haltiwanger, Jarmin, and Miranda (2016), is that the skewness of firm growth rates has declined, suggesting that high-growth firms are becoming increasingly rare.

Our framework enables us to identify the sources of the observed changes in firm dynamics. In particular, we are able to assess whether firm dynamics have changed as a result of changes in ex-post shocks affecting all firms, or whether the observed secular trends are the result of shifts in ex-ante heterogeneity determining the *type* of startups operating in the economy.

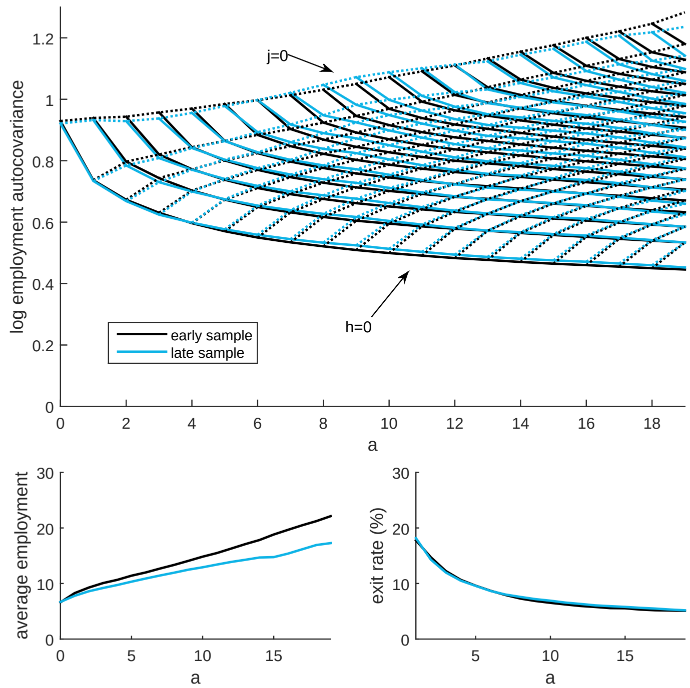
We analyze the changes in firm dynamics by splitting our data into an early sample, including firms born between 1979 and 1985, and a late sample with firms born between 1986 and 1993. Again, we follow all firms up to age 19. We first document changes in the three sets of key moments, the autocovariance 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. In particular, we study whether the population share and growth profiles of gazelles have changed, and how this has affected aggregate outcomes.

5.1 Changes in the data

Figure 11 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 (forthcoming).

What has changed, however, is the profile of average size by age, which is shown in the bottom left panel of Figure 11. 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 nineteen, 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.

Figure 11: Split-sample data moments (firms)



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 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.

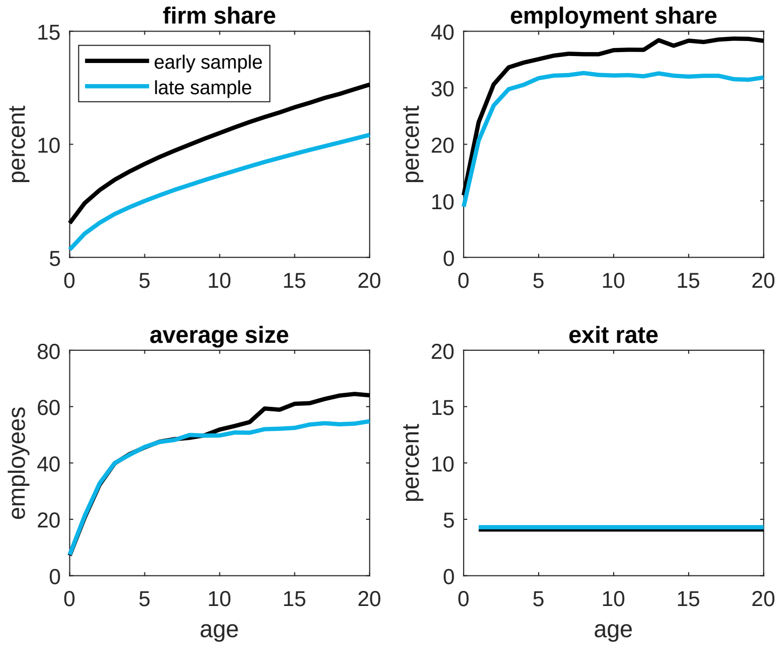
5.2 Are gazelles dying out?

To investigate the observed changes in firm dynamics and their aggregate consequences, we re-estimate the model on the two subsamples. The parameter values and model fit are shown in Appendix C.2.

Inspecting the estimated parameters alone, it is difficult to pinpoint whether the observed changes in firm dynamics should be attributed to different ex-post or ex-ante characteristics across the two subsamples. Therefore, in what follows, we focus on (ex-ante identified) gazelles and ask to what extent changes in their characteristics alone can explain the observed decline in business dynamism. This is justified also by the stability of the exit profile across the two subsamples which points to changes in the top of the firm distribution, rather than the bottom where firms exit relatively more frequently.

Towards this end, we compute the fraction of gazelles in the population of firms in both subsamples. This is shown in the left top panel of Figure 12. Among startups,

Figure 12: Characteristics of gazelles in the early and late sample (firms)



Note: Top panels: share of gazelles in the total number of firms and in total employment. Bottom panels: average size and exit rate profile of gazelles. 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.

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. At age twenty, the fraction of gazelles is 12.5 percent in the early sample but only 10.4 in the late sample.

A similar picture is painted by the top right panel which shows the employment shares, by age. Among startups at age zero, gazelles account for around 9 percent of employment in both the early and the late samples. However, a gap emerges between the two samples as firms age and start fulfilling their ex-ante growth potential.

The bottom left panel shows the average size profile of gazelles. In both sub-samples, gazelles start with around 7 employees, but grow quickly to reach on average about 46 employees by age five. Around age 10, however, the two sub-samples diverge, and a reduction in the average size between the two sub-samples becomes apparent. Thus, in the late sample gazelles on average do not grow as large as in the early sample. Finally, the exit profile, plotted in the bottom right panel, is essentially the same in both samples, as gazelles exit practically only for exogenous reasons.[27]

Our findings thus confirm the concerns that high-growth firms are becoming increasingly rare. While Decker, Haltiwanger, Jarmin, and Miranda (2016) document that the decline in the skewness of firm growth rates occurred around 2000 and primarily in the services, information and high-tech sectors, the sources of these secular changes remain to be identified. While our framework does not provide a definitive answer to this question, it does offer additional new insights. First, we document that the disappearance of gazelles is related to ex-ante factors, suggesting that high-growth firms are in fact dying out. Second, not only are there fewer gazelles, but those that nevertheless start up tend to expand less than high-growth firms of the past.

5.3 Aggregate implications

While the results clearly suggest a change in the characteristics of high-growth firms, it is unclear to what extent these changes alone can help explain the observed decline in dynamism at the aggregate level. After all, gazelles account for only a small share of all businesses.

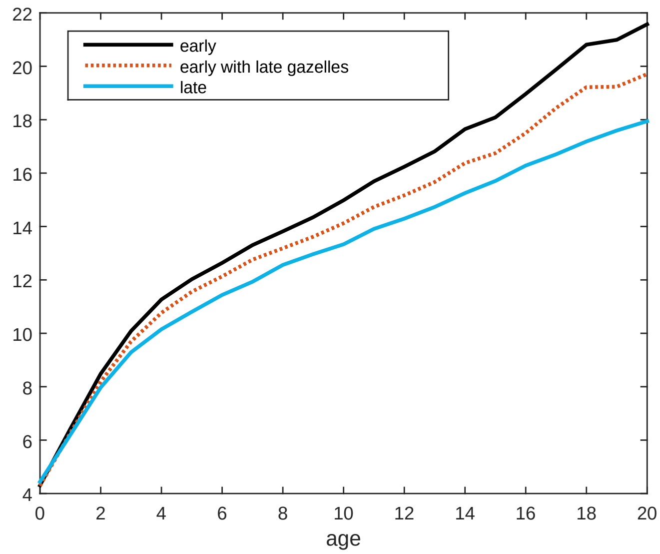
To investigate this question, Figure 13 plots the average size profile, in the estimated model over the two sub-samples. As noted before, this profile has flattened. To assess the contribution of disappearing gazelles to this shift, we use the fact that at any age the average size among all firms is the sum of the average size of gazelles and non-gazelles, weighted by their respective firm share. We then construct a counterfactual in which we re-compute 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 in Figure 12 plots this counterfactual. It shows that the change in the fraction of gazelles and their average size profile accounts for roughly half of the decline in the average size profile. This is remarkable, given that gazelles account for only about five percent of the startups.

Finally, we evaluate the implications for aggregate output. We find that between the two samples, aggregate output declines by 4.5 percent. Thus, seemingly small changes in the distribution of firms, such as the decline in the (already low) share of

[27]Consistent with this finding, the Appendix also shows that the model generates a substantial decline in skewness of firm growth rates across the two subsamples. This decline is generated entirely endogenously, since we assume symmetric, normally distributed shocks.

Figure 13: The impact of disappearing gazelles (firms)



Note: 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.

high-potential startups, as well as a reduction in their growth potential, emerge as important drivers of *aggregate* changes.[28,29]

5.4 Supporting empirical evidence

The above results suggest that changes in ex-ante factors are the key driver of the observed decline in business dynamism. In this subsection, we provide *model-free* evidence in support of this conclusion. In particular, if indeed the nature of firm growth has changed over time because of ex-ante factors, we should observe different patterns across different *cohorts* of startups rather changing patterns among all firms across time.

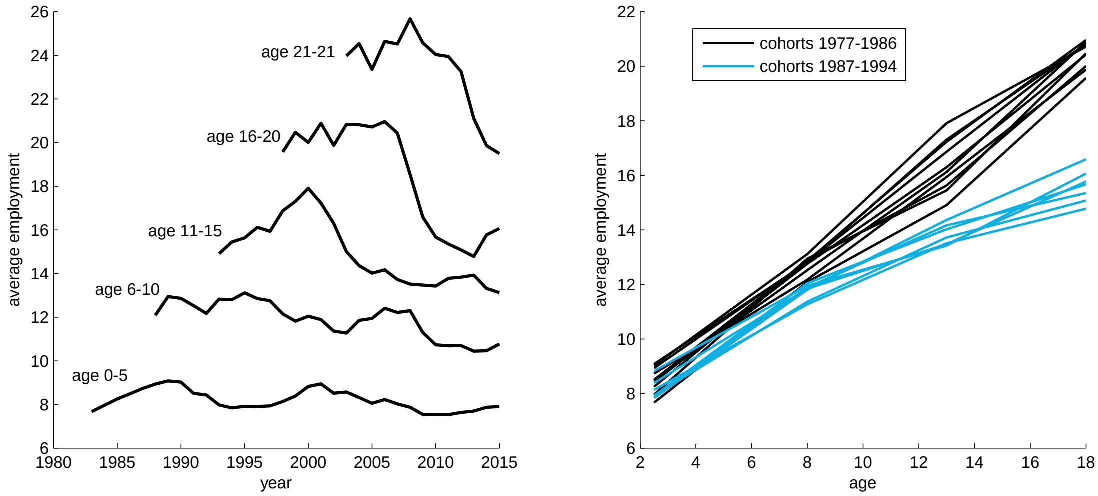
To better understand the flattening of the age profile, we consider in more detail how it occurred. Figure 14 plots average firm size in different five-year age bins.[30]

[28]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. In a model version with endogenous labor supply, there could be an associated decline in aggregate employment as well.

[29]Shifts in the number of startups may also have important macroeconomic consequences, see Sedláček (2015).

[30]The figure uses the Business Dynamics Statistics data, which is a publicly available aggregated version of the underlying LBD data set used in our estimations.

Figure 14: Flattening of the average size profile in the data (firms)



Note: The left panel plots, by year, average firm employment in different age bins: 0-5 years, 6-10 years, 11-15 years, 16-20 years, and 21-25 years. The right panel plots the same data, but now by overlapping 5-year cohorts, grouped by the birth year of the youngest firm in each cohort. Source: Business Dynamics Statistics.

The left panel plots these by year of observation. The figure clearly shows the decline in average size among older firms.[31] However, it also shows that the decline occurred in a staggered way, taking place later in older age bins. In particular, the average size time paths of the three oldest age categories clearly move in lock-step with five year gaps between them. This also makes clear that the flattening of the average size profile was set into motion before the Great Recession. Finally, note that average size declined also for firms $0-5$ and $6-10$ years of age by 3 and 15 percent, respectively.[32]

The right panel of Figure 14 plots the same data but now by cohort, defined by the birth year of the youngest firms in each age category. The figure shows very clearly that the flattening occurred by cohort. In addition, this change was not gradual, but it happened rather abruptly in the late 1980's. Specifically, cohorts born since the late 1980's had a much flatter average size profile compared to cohorts of firms born earlier. These patterns support our results that the flattening of the average size profile was an ex-ante phenomenon, rather than the result of changes in the character of ex-post shocks, which would likely affect *all* firms simultaneously.

[31]This finding is consistent with Adam and Weber (2017), who regress establishment size on an age trend, as well as industry dummies, and find that the age coefficient has declined over time. They then proceed to study the consequences for the optimal inflation rate.

[32]These values are based on averages in the first and last 15 years of the sample period.

The above insights point to potential future avenues of research attempting to identify the reasons behind the disappearance of gazelles. In particular, our results suggest that the disappearance of gazelles was set in motion already in the late 1980's, as opposed to the early 2000's when the change in skewness became most apparent. An intriguing connection may be made between the demise of gazelle startups and the decline in the aggregate labor share of income, which also started in the late 1980's. For example, Autor, Dorn, Katz, Patterson, and Reenen (2017) suggest that the decline in the labor share was due to an increase in product market concentration, giving rise to "superstar firms". Increased domination of incumbent superstar firms might have made it more difficult for high-potential startups to enter the economy. Or vice versa, a lack of competitive pressure from gazelle startups might have contributed to the increase in market concentration.

6 Conclusions

We have used data on the population of U.S. 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. Most of the dispersion in firm size, at a given age, is driven by such ex-ante potential. Moreover, we found that that ex-ante heterogeneity also drives much of the "up-or-out" dynamics observed in the data: high-potential startups, "gazelles", grow quickly and survive at high rates, whereas low-potential firms tend to exit quickly. These dynamics lead to substantial gains in aggregate output.

We have also investigated potential changes in the firm dynamics process, following up on recent concerns that high-growth firms are disappearing. We documented a dramatic flattening of the age profile of average size, among cohorts of firms born since the late 1980's. Re-estimating the 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.

Our results highlight the need for future research on which individuals become entrepreneurs and what decisions such aspiring entrepreneurs make *before or at* startup, as opposed to their behavior *after* the firm has become operational. While the macroeconomic implications of the latter have been studied extensively in the literature, much less is known about how institutional conditions change who becomes an entrepreneur and what types of firms are being created. Our results show that such changes can be of first-order importance for macroeconomic outcomes.

References

Abbring, J., and J. Campbell (2005): "A Firm's First Year," Tinbergen Institute Discussion Paper 05-046/3.

Abowd, J., and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2), 4111–445.

Adam, K., and H. Weber (2017): "Optimal trend inflation," mimeo.

Akcigit, U., H. Alp, and M. Peters (2017): "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries," NBER Working Paper 21905.

Arellano, M., and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and Application to Employment Equations," *Review of Economic Studies*, 58(2), 277–297.

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

Arkolakis, C., T. Papageorgiou, and O. Timoshenko (forthcoming): "Firm Learning and Growth," *Review of Economic Dynamics*.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. V. Reenen (2017): "The Fall of the Labor Share and the Rise of Superstar Firms," Working paper.

Barseghyan, L., and R. Dicecio (2011): "Entry Costs, Industry Structure, and Cross-Country Income and TFP Differences," *Journal of Economic Theory*, 146, 1828–1851.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2013): "Cross-Country Differences in Productivity: The Role of Allocation and Selection," *American Economic Review*, 103, 305–334.

Beaudry, P., and F. Portier (2004): "An exploration into Pigou's theory of cycles," *Journal of Monetary Economics*, 51(6), 1183–1216.

Belenzon, S., A. Chatterji, and B. Daley (2017): "Eponymous Entrepreneurs," *American Economic Review*, 107(6), 1638–1655.

Birch, D., and J. Medoff (1994): "Gazelles," in *Labor Markets, Employment Policy and Job Creation*, ed. by L. Solmon, and A. Levenson, pp. 159–168. Boulder: Westview Press.

Cabral, L., and J. Mata (2003): "On the Evolution of the Firm Size Distribution: Facts and Theory," *American Economic Review*, 93(4), 1075–1090.

Campbell, J., and M. De Nardi (2009): "A Conversation with 590 Nascent Entrepreneurs," *Annals of Finance*, 5(3), 313–340.

Chamberlain, G. (1984): "Panel Data," in *Handbook of Econometrics, Volume II*, ed. by Z. Griliches, chap. 22, pp. 1274–1318. Elsevier Science, Amsterdam, North-Holland.

Decker, R., J. Haltiwanger, R. Jarmin, and J. 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.

Drozd, L., and J. Nosal (2012): "Understanding International Prices: Customers as Capital," *American Economic Review*, 102(1), 364–395.

Fort, T. C., S. D. Klimek, et al. (2016): "The Effects of Industry Classification Changes on US Employment Composition," Discussion paper, Tuck School at Dartmouth.

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

Gabaix, X. (2009): "Power laws in economics and finance," *Annu. Rev. Econ.*, 1(1), 255–294.

Gourio, F., and L. Rudanko (2014): "Customer Capital," *Review of Economic Studies*, 81(3), 1102–1136.

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

Guvenen, F., and A. Smith (2014): "Inferring Labor Income Risk and Partial Insurance from Economic Choices," *Econometrica*, 82(6), 2085–2129.

Guzman, J., and S. Stern (2015): "Where is Silicon Valley?," *Science*, 347(6222), 606–609.

Haltiwanger, J., R. Jarmin, R. Kulick, and J. Miranda (2016): "High Growth Young Firms: Contribution to Job, Output and Productivity Growth," US Census Bureau Center for Economic Studies Paper No. CES-WP-16-49.

Haltiwanger, J., R. Jarmin, and J. Miranda (2013): "Who Creates Jobs? Small vs. Large vs. Young," *The Review of Economics and Statistics*, 45(2), 347–361.

Hopenhayn, H. (1992): "Entry, Exit and Firm Dynamics Long Run Equilibrium," *Econometrica*, 60(5), 1127–1150.

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

Hottman, C., S. Redding, and D. Weinstein (2016): "Quantifying the Sources of Firm Heterogeneity," *Quarterly Journal of Economics*, 131(3), 1291–1364.

Hsieh, C.-T., and P. J. Klenow (2014): "The Life Cycle of Plants in India and Mexico," *Quarterly Journal of Economics*, 129, 1035–1084.

Hurst, E., and B. Pugsley (2011): "What Do Small Businesses Do?," *Brookings Papers on Economic Activity*, 2, 73–118.

Jarmin, R. S., and J. Miranda (2002): "The longitudinal business database," Discussion paper, Census Bureau.

Jovanovic, B. (1982): "Selection and the Evolution of Industry," *Econometrica*, 50(3), 649–670.

Klette, T. J., and S. Kortum (2004): "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 112.

Lee, Y., and T. Mukoyama (2015): "Productivity and Employment Dynamics of US Manufacturing Plants," *Economics Letters*, 136, 190–193.

Luttmer, E. (2007): "Selection, Growth, And the Size Distribution of Firms," *Quarterly Journal of Economics*, 122(3), 1103–1144.

——— (2011): "On the Mechanics of Firm Growth," *Review of Economic Studies*, 78, 1042–1068.

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

Melitz, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.

Perla, J. (2015): "Product Awareness, Industry Life Cycles, and Aggregate Profits," mimeo.

Pugsley, B. W., and A. Şahin (forthcoming): "Grown-up Business Cycles," *The Review of Financial Studies*, p. hhy063.

Pugsley, B. W., and H. Wheeler (2018): "Longitudinal Linking of Enterprises in the LBD and SSL," Discussion paper, CES TN-2018-02.

Rouwenhorst, K. G. (1995): "Asset Pricing Implications of Equilibrium Business Cycle Models," in *Frontiers of Business Cycle Research*, ed. by T. Cooley, pp. 294–330. Princeton University Press.

Schoar, A. (2010): "The Divide between Subsistence and Transformational Entrepreneurship," in *Innovation Policy and the Economy, Volume 10*, pp. 57–81. NBER Chapters.

Sedláček, P. (2015): "Lost Generations of Firms and Aggregate Labor Market Dynamics," Working paper.

Sedláček, P., and V. Sterk (2017): "The Growth Potential of Startups Over the Business Cycle," *American Economic Review*, 107(10), 3182–3210.