The Growth Potential of Startups over the Business Cycle*

Petr Sedláček
University of Bonn †

Vincent Sterk
University College London ‡

September 18, 2014

Abstract

This paper shows that employment of U.S. firm cohorts is strongly influenced by aggregate conditions at the time of their entry. Employment variations across cohorts are found to be persistent and largely driven by differences in average firm size, rather than the number of firms. To disentangle startup composition from post-entry choices, we estimate a general equilibrium firm dynamics model. We find that aggregate conditions at birth drive the vast majority of employment variation across cohorts through their effect on the proportion of startups with high growth potential. In the aggregate, startup conditions result in large slow-moving fluctuations in employment.

Keywords: Firm Dynamics, Heterogeneous Agents, Maximum Likelihood, DSGE

JEL Codes: E32, D22, L11, M13

*We thank Eric Bartelsman, Christian Bayer, Francisco Buera, Mariacristina De Nardi, Philip Jung, Leo Kaas, Patrick Kehoe, Javier Miranda, Fabien Postel-Vinay, Morten Ravn, Aysegul Sahin, Moritz Schularick, Kjetil Storesletten, Emily and Robert Swift as well as seminar participants in Adelaide, Bocconi, Edinburgh, Frankfurt, Konstanz, Manchester, Melbourne, Surrey, Sydney, Warwick and at the ANU, De Nederlansche Bank, the Ifo, and UCL for helpful comments. All errors are our own.

†Department of Economics, e-mail: sedlacek@uni-bonn.de.

‡Department of Economics and Centre for Macroeconomics, e-mail: v.sterk@ucl.ac.uk.

1 Introduction

The number of firm startups in the U.S. fell sharply during the Great Recession.[1] Given the importance of startups for aggregate job creation, this raises concerns about a long-lasting drag on aggregate employment and output. While such worries are valid, this paper shows that in addition we should be concerned with the roughly 2 million startups that *did* enter during the latest downturn. We document that recession-born firms tend to remain persistently smaller on average, even when the aggregate economy recovers. Underlying this pattern are changes in the types of startups with respect to their potential to grow large. Moreover, rather than fading out over time, decisions taken at the entry phase leave an increasingly large footprint on the macro-economy as startups age.

Using newly developed Business Dynamics Statistics (BDS) we follow cohorts of firms, starting from their year of entry. The data span all sectors in the U.S. private economy and cover the years from 1979 until 2011. We document three new stylized facts: (i) variations in entrant employment are large and pro-cyclical, (ii) to a great extent these variations persist as firm cohorts age (sharply contrasting the strong mean-reversion in aggregate employment) and (iii) the majority of variation in cohort-level employment is driven by changes in average firm size, rather than by the number of firms within a cohort.

The empirical patterns suggest that cohorts born at different stages of the business cycle are composed of different types of firms, giving rise to long-lasting effects. However, the composition of startups is unobserved and firm size fluctuations across cohorts are also driven by variations in post-entry decisions made by a given mix of firms. To disentangle the two and to quantify the impact of compositions changes, we *estimate* a general equilibrium firm dynamics model using both aggregate and cohort-level data. The model features heterogeneous firms, endogenous entry, post-entry growth subject to adjustment costs and aggregate uncertainty. Importantly, depending on aggregate conditions at the time of entry, startups choose a production technology affecting their ability to grow large.

[1] According to the Business Dynamic Statistics, the number of startups in 2009 was 30% below its pre-crisis level in 2006.

The estimated model implies that conditions in the year of birth drive at least 90 percent of the variation in employment across cohorts of a given age, even for older firms. The importance of the birth stage for long-term outcomes stems from the composition of cohorts with respect to firms' chosen production types. Initially, changes in the composition of startups have only a moderate effect since all firms enter relatively small. But as the cohort ages, a fraction of firms grows large and the composition effect on cohort-level employment becomes increasingly pronounced. Comparing for example the cohorts born in 2006 and 2008 our model predicts that, due to differences in composition alone, firms born in 2008 grow to be on average 13 percent smaller at age 5 (amounting to a reduction of 300 thousand jobs). By the age of 20, the predicted gap grows to almost 20 percent.

Using the model we also show that macroeconomic conditions at the startup phase are important for fluctuations in aggregate employment. In particular, the contribution of startup conditions to aggregate employment fluctuations evolves in a strikingly similar way to the *trend* component of the employment rate, often discarded in business cycle analysis. Our framework thus opens up an avenue for future research aiming to understand the drivers of macroeconomic fluctuations at a more complete range of frequencies. Furthermore, we show that general equilibrium effects associated with fluctuations in entry decisions are strong, particularly in the long run.

An important pre-requisite of our analysis is the estimation of the model using Maximum Likelihood. It is well known that solving heterogeneous firm models with aggregate uncertainty is a complex problem because the aggregate state includes entire distributions of firm-specific variables. A key methodological contribution of this paper is designing a novel computational strategy, based on first-order perturbation around firms' steady state growth paths, allowing the model to be solved quickly.[2] We identify changes in startup composition by exploiting the model's implication that a cohort's composition reveals itself through its increasing influence on average firm size as the cohort ages. The use of cohort-level data, observed at different points in time, is therefore paramount to our estimation.

[2]Campbell (1998) pioneered the use of perturbation methods to solve heterogeneous-agents models, replacing functional equations with quadrature approximations. Our setup avoids the need for such approximations because the economic state is large but finite-dimensional, preserving exact aggregation.

The model incorporates two novel features in order to suit our empirical focus. First, the model allows for heterogeneity in returns-to-scale in production. This contrasts standard firm dynamics models in which heterogeneity is introduced purely through variations in total factor productivity (TFP) across firms. A motivating factor behind our choice is that we apply the model to the entire cross-section of private employers in the economy rather than to confined industries, in which returns to scale are arguably more homogeneous.[3] Moreover, only a small degree of heterogeneity in returns to scale is sufficient for our model to fit the size distribution of firms in the data, conditional on age. That said, we show that a version of the model in which firms differ only in TFP delivers very similar results.

The second novel feature of our framework is the modeling of the firm entry phase. Aspiring entrants are free to *choose* any of the given number of business opportunities, each associated with a certain technology type. However, highly scalable businesses are relatively valuable, provoking intense competition for these opportunities among aspiring startups. Strong competition reduces the probability of successfully starting up and encourages entry of firms that have little potential to grow large, of which we see many in the data.[4] What arises is a natural equilibrium relation between the value of a firm and the number of startups of a particular technology type. This modeling framework is akin to the directed search literature (see e.g. Moen, 1997) and similar in spirit to models of innovation and research and development (see e.g. Klette and Kortum, 2004).

The composition of startups, therefore, responds endogenously to aggregate shocks through their differential impact on the values of firms of different types. In addition, startup composition is affected directly by a shock to the distribution of business opportunities across firm types, resembling shocks in models of vintage technologies (see e.g. Campbell, 1998; Gilchrist and Williams, 2000). The set of microfoundations one can give to this shock is, however, more general. In the Appendix we show how various

[3]Even so, Holmes and Stevens (2012) provide evidence of substantial heterogeneity in returns to scale even within narrowly defined industries. Basu and Fernald (1997) provide evidence in favor of heterogeneity in returns to scale across sectors.

[4]In 2007, the fraction of firms with 10 or fewer employees among firms between 21 and 25 years of age was about two thirds. This is also consistent with empirical evidence that many starting entrepreneurs have low growth expectations, see Campbell and De Nardi (2009) and Hurst and Pugsley (2011).

frictions – related to uncertainty and financial and product markets – map into outcomes which are observationally equivalent to our setup. The advantage of our approach is that we are able to quantify the overall impact of composition changes. Nonetheless, we explore to what extent proxies of these frictions comove with our estimated measure of startup composition and find that product market frictions may be an additional driver.

The empirical results complement the analysis in Haltiwanger, Jarmin, and Miranda (2013), who emphasize the importance of young firms for aggregate job creation on average. Cyclical patterns in firm entry are studied in Campbell (1998) and Lee and Mukoyama (2013) who analyze the behavior of entering and exiting firms in the manufacturing sector. Unlike these studies, we exploit the newly developed BDS data to *follow* cohorts of firms as they age, enabling us to investigate how their later job creation is affected by aggregate conditions at the time of their birth.[5]

We are not the first to study firm dynamics and variations in entrant size. However, in contrast to existing studies we use our general equilibrium firm dynamics model as an *empirical tool* to uncover an unobservable state of the aggregate economy: the distribution of entrant types with respect to their growth potential. Moreover, our analysis focuses on the aggregate economy rather than industry-level outcomes. Our framework builds on a rich literature of structural firm dynamics models. A workhorse model of firm dynamics (without aggregate uncertainty) is presented in Hopenhayn and Rogerson (1993). Abbring and Campbell (2004) estimate a partial equilibrium firm dynamics model without aggregate uncertainty and find that pre-entry scale decisions are important for the variation in sales across firms. Campbell (1998) and more recently Lee and Mukoyama (2013) and Clementi and Palazzo (2014) use the Hopenhayn-Rogerson framework to study the role of entry and exit decisions in the propagation of aggregate shocks in the manufacturing industry.

[5] Earlier studies using BDS data include Moscarini and Postel-Vinay (2012) and Fort, Haltiwanger, Jarmin, and Miranda (2013) who study the cyclical sensitivities of large versus small and younger versus older firms, but do not focus on entrants or cohorts. Decker, Haltiwanger, Jarmin, and Miranda (2013) use BDS data to document a downward trend in the pace of business dynamism, and find that a secular decline in the number of startups accounts for much of this trend decline. Also related is Bartelsman, Haltiwanger, and Scarpetta (2009) who use a cross-country data set to study average post-entry behavior of young firms.

The organization of the remainder of this paper is as follows. Section 2 describes the data and presents empirical stylized facts. The model and its parametrization are described in Sections 3 and 4, respectively. Section 5 presents the model results and elaborates on potential additional drivers of the compositional fluctuations uncovered by the estimated model. Concluding remarks are made in Section 6.

2 Empirical evidence

Startups are widely recognized to be important drivers of aggregate job creation on average, at least since Haltiwanger, Jarmin, and Miranda (2013). This section documents cyclical patterns in employment by young U.S. firms over time, without imposing any model structure. Our units of analysis are cohorts, that is, aggregates over firms born in the same year. We exploit newly developed Business Dynamics Statistics (BDS) data, described in the next subsection, which covers all sectors in the private economy. The findings can be summarized by three stylized facts:

Fact 1. *Entrant job creation is volatile and pro-cyclical.*

Fact 2. *Cohort-level employment is largely determined in the year of birth.*

Fact 3. *Average size is the main driver of variations in cohort-level employment with an increasing importance as cohorts age.*

We aim to contribute to a rich empirical literature that tries to understand the dynamics of firms and differences across firms, as surveyed in Bartelsman and Doms (2000) and Syverson (2011). In particular, our first stylized fact complements empirical evidence presented in Campbell (1998), and Lee and Mukoyama (2013), who find that the number and job creation of new plants is pro-cyclical in the manufacturing sector. Our analysis, by contrast, is not confined to a single industry and applies to firms rather than establishments.[6,7] To the best of our knowledge, our sec-

[6]Appendix D presents within-industry findings and makes a more detailed comparison of the cyclical patterns we find in the BDS relative to those of Lee and Mukoyama (2013) who use data from the Longitudinal Research Database provided by the Census Bureau.

[7]Appendix B shows that our results remain to hold also for establishments. An

ond and third stylized facts have no precedent in the empirical literature on firm job creation.

2.1 Data and definitions

The BDS database is based on administrative records and covers 98 percent of US private employment. This is an important advantage over alternative data sources, especially given our objective to study implications for aggregate outcomes. We use the available annual information on the number of firms and their job creation broken down into age categories, for the period 1979 until 2011.[8]

The available age breakdown in the BDS allows us to follow cohorts of new firms for up to five years after they enter the economy.[9] The BDS groups older firms into age categories spanning five years. However, we find that our stylized facts continue to hold for averages for firms 6-10 and 11-15 years of age. Additionally, Appendix C presents a robustness check using establishment-level micro-data underlying the BDS, which we use to construct data for (1-year) cohorts up to the age of 15. The documented aggregate patterns hold far beyond the age of five, reinforcing our results.

We introduce the following notation. Let $M_{a,t}$ be the number of firms in a cohort of age a in year t. Following the BDS notation, startups enter with age $a = 0$. Similarly, let $N_{a,t}$ be the employment level of a cohort of firms of age a in year t. The employment level of a given cohort is measured as cumulative net job creation since birth, i.e. $N_{a,t} = \sum_{i=0}^{a} NJC_{i,t-a+i}$, where $NJC_{a,t}$ is the net number of jobs created in firms of age a in year t.[10]

establishment is defined as a single physical location where business is conducted or where services or industrial operations are performed. A firm is a business organization consisting of one or more establishments that were specified under common ownership or control.

[8]The data represents a snapshot taken in March of each year. The data starts in 1977 but we drop the initial two years following Moscarini and Postel-Vinay (2012), who cast doubt on the quality of the initial two years.

[9]Appendix A.3 shows that our patterns are not driven by a particular cohort by analyzing a two year moving average of the data.

[10]Alternatively, one could use the employment stock data presented in the BDS. These employment numbers do not equal the sum of net job creation because the net job creation data is cleaned from observed entrants that are not believed to be true startups, while the employment data is not. BDS documentation states that: "...it may be determined that an establishment's entry/exit as shown by the data is not credible. These establishments are excluded from the change calculations in a given year"

Table 1: Correlations of entrant job creation with business cycle indicators

	linear trend		HP filter		levels	
	e-rate	GDP	e-rate	GDP	e-rate	ΔGDP
NJC entrants	0.59	0.68	0.31	0.39	0.59	0.47

Notes: The table reports correlation coefficients between the variables in the columns and job creation of entrants for various de-trending methods. "e-rate" stands for employment rate, defined as 1 minus the unemployment rate (the correlation with the employment-to-population ratio is between 0.92 and 0.98, depending on the detrending method). "ΔGDP" is the growth rate of real GDP. All variables are logged prior to detrending and entrant job creation is treated in the same way as the given business cycle indicator.

2.2 The cyclicality of startup job creation

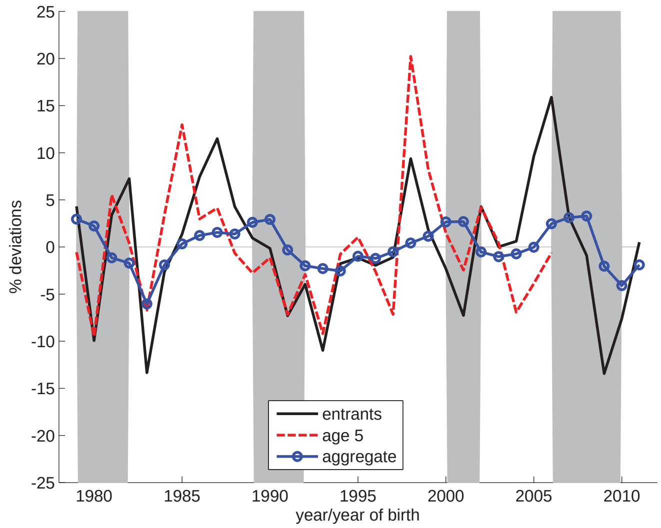
To visualize the cyclicality of cohort-level employment, Figure 1 displays HP-filtered employment levels between 1979 and 2011 of (i) cohorts of startups, (ii) cohorts of five year old firms, where the time series is shifted back to the year of their birth and (iii) the aggregate economy.[11] Several patterns stand out. First, fluctuations in cohort-level employment are large (more than 3 times larger than that of aggregate employment). Also, the cohort-level volatility does not appear to diminish with age. Second, job creation by entrants and aggregate employment move together and drop during recession years, indicated by shaded areas.

Table 1 presents more formal measures of cyclicality by correlating employment in startups with several business cycle indicators and using several detrending methods. In all cases the correlations are positive and statistically significant, confirming the strong pro-cyclical nature of entrant job creation.

(http://www.census.gov/ces/dataproducts/bds/faqs.html). Thus, the net job creation data are superior, at least for our purpose. Nevertheless, Appendix A.2 shows that our results are robust to other ways of constructing employment levels.

[11]All variables are logged prior to de-trending. Because our analysis deals with time series of different lengths (e.g. information on five year old firms starts only in 1984), we always de-trend the given data over the longest possible sample with the earliest starting point in 1979. Throughout the paper, the smoothing parameter in the HP filter is set to 100 to leave no obvious cyclical pattern. Appendix A.1 confirms robustness of our results to alternative detrending methods. Where possible, aggregate variables are averages over March-to-March periods, consistent with the BDS timing.

Figure 1: Cohort-level employment at age 0 and 5 by year of birth and aggregate employment by year



Notes: Cohort-level and aggregate employment are plotted in percentage deviations from an HP-trend. Shaded areas are the NBER recessions. Source: BDS, BLS.

2.3 The year of birth and cohort-level employment

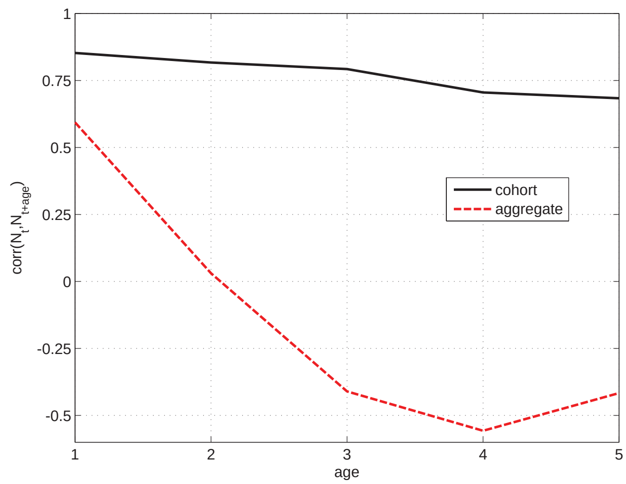
To quantify the persistence of cohort-level employment, we correlate entrant employment in year t with the employment level in $t + a$ in the same cohort. Figure 2 plots the correlation coefficients for ages $a = 1$ to $a = 5$ both for the individual cohorts, as well as for aggregate employment.[12] While cohort-level employment at birth and 5 years into existence are highly correlated, with a correlation coefficient of 0.68, its aggregate counterpart displays no persistence after two years. Thus, there is no evidence of convergence across cohorts. Turning our attention to older firms (grouped in 5-year bins), we find that the persistence is at least as strong. In particular, for cohorts in the age category 11-15 we find that the correlation of their current employment level with the level 10 years earlier is 0.89 (not plotted).[13]

A formal alternative to the reported autocorrelations is presented in

[12]While cohort-level employment does not display a trend, aggregate employment does and therefore we choose to HP-filter both time series. For the aggregate series we simply correlate period t employment with employment in years $t + a$.

[13]To be able to compute this number we use the employment level reported in the BDS, rather than net job creation. We use the same de-trending methods as before.

Figure 2: Autocorrelations



Notes: Correlation coefficients of employment in year $t = 0$ and in year $t + age$, with $age = 1, 2, 3, 4, 5$, at both the level of a cohort born in period $t = 0$ and at the aggregate level. Source: BDS, BLS.

Appendix A.4, which estimates panel regression of cohort-level employment at different ages on entrant job creation and explicitly controls for age and also current entrant employment (a measure of current aggregate startup conditions). The implied elasticities of employment at different ages with respect to employment in the year of entry correspond closely with the autocorrelation coefficients reported in Figure 2.

2.4 Decomposing cohort-level employment variation

The previous paragraphs established that the number of jobs created by cohorts of startups upon entry largely persists into later years of their existence and that the fluctuations in these numbers are large. We now investigate whether the observed variation of cohort-level employment is driven primarily by variation in the number of firms within the cohort (extensive margin), or by persistent differences in the growth potential of startups (intensive margin).

To this end, we decompose the natural logarithm of cohort-level employment as:

$$\ln N_{a,t} = \ln S_{0,t-a} + \sum_{j=1}^{a} \ln \gamma_{j,t-a+j} + \ln M_{0,t-a} + \sum_{j=1}^{a} \ln \delta_{j,t-a+j},$$

where $S_{a,t}$ is average firm size within the cohort, $M_{a,t}$ is again the number of firms, $\gamma_{j,t} \equiv \frac{S_{j,t}}{S_{j-1,t-1}}$ denotes average size growth and $\delta_{j,t} \equiv \frac{M_{j,t}}{M_{j-1,t-1}}$ denotes the average firm survival rate. Based on the above expression, the variance of employment can be decomposed as:

$$\mathrm{var}(\hat{N}_{a,t}) = \underbrace{\mathrm{cov}(\hat{N}_{a,t}, \hat{S}_{0,t-a}) + \sum_{j=1}^{a} \mathrm{cov}(\hat{N}_{a,t}, \hat{\gamma}_{j,t-a+j})}_{\text{intensive margin}} + \underbrace{\mathrm{cov}(\hat{N}_{a,t}, \hat{M}_{0,t-a}) + \sum_{j=1}^{a} \mathrm{cov}(\hat{N}_{a,t}, \hat{\delta}_{j,t-a+j})}_{\text{extensive margin}} + \eta_t,$$

where a hat indicates deviations from an HP-filter trend of a logged variable and η_t is a residual term coming from the detrending method.[14] The first two terms on the right-hand side jointly capture the contribution of the intensive margin (average size) to the total variance. The first term individually captures the contribution of average size in the year of entry alone. The third and fourth terms capture the contributions of the extensive margin.

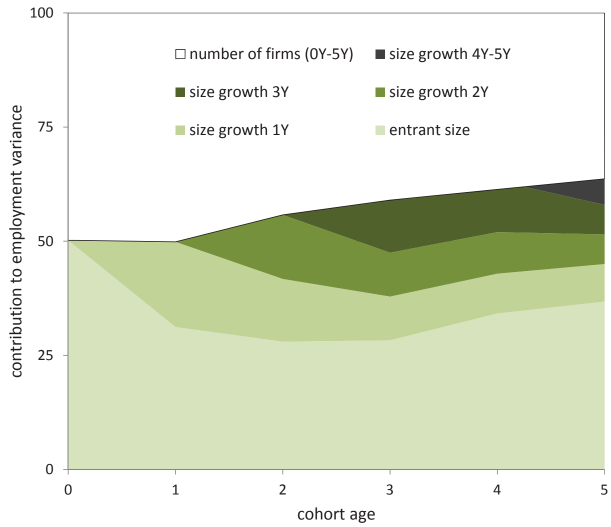
The importance of the intensive margin is made clear by Figure 3, which displays the variance decomposition by age. The total shaded area represents the contributions of average size variations to cohort-level employment fluctuations, at different ages. The white area accounts for the contribution of variation in the number of firms.[15] It is clear that the contribution of average firm size variation is increasing as the cohort ages (accounting for about 50% at birth and 64% at age 5).[16] Extending the

[14] In our case, the residual η is negligible, not exceeding 0.01%.

[15] The vast majority of the contribution of the number of firms is due to the number of startups. The contribution of changes in the number of firms after startup, i.e. fluctuations in firm survival rates, account for only 3% on average for firms aged 1 to 5.

[16] The intensive margin also becomes increasingly important for the *average level* of cohort employment as the cohort ages (i.e. $\ln S_{0,t-a} + \sum_{j=1}^{a} \ln \gamma_{j,t-a+j}$ increases relative to $\ln M_{0,t-a} + \sum_{j=1}^{a} \ln \delta_{j,t-a+j}$). Especially in the initial years following birth, many firms exit whereas those that continue grow rapidly on average. These patterns have been labeled "up-or-out" dynamics by Haltiwanger, Jarmin, and Miranda (2013) and will play an important role in the structural model we estimate in the next section. In partic-

Figure 3: Contribution of average size to employment variation



Notes: Contributions of average firm size at different ages and the number of firms to the variation in cohort-level employment as a percentage of its total variation. Source: BDS.

analysis to older firms reveals that average size remains very important in determining fluctuations in cohort-level employment, accounting for 70% among 11 to 15 year old firms (not plotted).

Within the total shaded area in Figure 3, different shades break down the contribution of the intensive margin by age, with the lightest shade denoting startup size. Clearly, the contribution of startup size variation is large and does not die out as the cohort ages, but plateaus at around 35%.

Although startup size emerges as a persistently important contributor to employment variations across cohorts, we need to be careful in interpreting it as a measure of compositional effects. Changes in entrant size may represent fluctuations in post-entry decisions by a given mix of firm types, rather than composition. Conversely, post-entry employment growth of a cohort is affected by its composition. We revisit the variance decomposition in Section 5.1.1 using our estimated model and precisely quantify the contribution of composition effects.

The established stylized facts paint a picture in which firm characteristics at the entry stage are crucial in determining a cohort's potency to

ular, they will imply that compositional variations across cohorts become increasingly pronounced with age.

create jobs, both initially and later in its life. One can think of several plausible explanations for why the composition of entrants may fluctuate over the business cycle. One possibility is that during recessions job creation within newborn cohorts declines because of a reallocation of activity between sectors. Appendix D documents, however, that our stylized facts continue to hold, with a few exceptions, also within sectors.[17]

Another possibility is that our findings are driven by fluctuations in the entry of very small firms. Several studies emphasize the role of entrepreneurship as a way to escape unemployment ("necessity entrepreneurs"). Such businesses are likely to remain very small.[18] However, Appendix E shows that the vast majority of employment variation of five year old firms is driven by large firms, rather than small ones.

3 The Model

The empirical evidence presented in the previous section suggests that fluctuations in the composition of startups are important for cohort-level employment in later years. However, the data alone does not allow us to quantify the importance of composition fluctuations, since we do not directly observe the distribution of cohorts with respect to firm types. For the same reason, the empirical facts can provide only limited information on the aggregate implications of decisions made at the entry stage.

To address these issues, this section proposes a general equilibrium model of firm dynamics in which startups can choose the technology type affecting the scalability of their aspired businesses. The model further features endogenous firm entry, labor adjustment costs and several sources of aggregate uncertainty. To quantify the contribution of composition effects and post-entry employment choices for the evolution of cohort-level employment in later years, we estimate the aggregate shocks using information about average size from the BDS. While post-entry shocks have a transitory impact on average firm size, the impact of changes in the com-

[17]The exceptions are the mining sector and transportation, communication, and public utilities in which entrant job creation is counter- and a-cyclical, respectively. Also, the extensive margin dominates employment variations in retail trade and construction.

[18]See e.g. Hurst and Pugsley (2011) and Poschke (2012). However, note that the BDS data do not include self-employed individuals.

position of startups increases as the cohort ages. Therefore, using data on both average firm size at birth and at a later age enables us to sharply identify fluctuations in the composition of cohorts of entrants.

The estimation procedure also delivers predicted values of the distribution of firms with respect to age and type in each period. The knowledge of this distribution enables us to investigate the aggregate implications of firm decisions made at the entry stage. We thus actively use the large aggregate state of our model in the quantitative analysis.

The model is designed for an application to cohort-level and aggregate data, rather than for analyzing individual firms. We therefore abstract from firm-specific technology shocks and associated endogenous exit after entry. Instead, we model an age-dependent exit rate calibrated using BDS data. Clearly, such an assumption is a simplification as exit rates are known to vary over time and to be related to firm productivity (see e.g. Bartelsman and Doms, 2000). Therefore, in Appendix G we show that allowing for stochastic variation in exit rates consistent with the data does not substantially affect our results.[19]

The model economy is populated by an infinitely-lived representative household and a continuum of heterogeneous firms. All agents have rational expectations. Firms and households trade on a goods market and a labor market, both of which are perfectly competitive.[20]

3.1 Existing firms

There is an endogenous mass of heterogeneous firms which produce a homogeneous good. Before describing the entry decision, which is key to our analysis, we lay out how incumbent firms behave. While all firms in the economy use labor as the only factor of production, the production technology itself differs across firms. In particular, there is a finite number of technology types, indexed by $i = 1, 2, ..., I$. Existing firms grow only gradually towards their optimal sizes due to costs related to adjusting their

[19]This is supported by the variance decomposition in Subsection 2.4 which implied that variation in exit rates explains on average only 3% of fluctuations in cohort-level employment for firms aged 1-5 years.

[20]Firm dynamics models with more detailed descriptions of the labor market include e.g. Elsby and Michaels (2013), Kaas and Kircher (2011), and Sedláček (2014).

employment levels. Finally, firms face an exogenous, but age-dependent probability of shutting down, denoted ρ_a. By symmetry, all firms of the same type and age make the same decisions and we therefore index them only by technology type and age.

Technology types differ in the degrees of returns to scale and/or total factor productivity. In particular, a firm of age a and with technology type i is characterized by the following production function

$$y_{i,a,t} = A_t z_i n_{i,a,t}^{\alpha_i},$$

where A_t is an exogenous and stochastic aggregate TFP variable with mean one, $n_{i,a,t}$ is the firm's level of employment, z_i is a technology-specific TFP parameter, and α_i is a technology-specific returns to scale parameter. In our quantitative simulations we parameterize $\alpha_i \in (0, 1)$ for each technology type i, i.e. returns to scale are decreasing. As a result, there exists a type-specific "optimal size" beyond which further growth is undesirable.

Firms are subject to costs of adjusting labor, $Q_t\zeta(\Delta n_{i,a,t})$, where $\zeta(.)$ is strictly increasing and strictly convex and Δ is the first difference operator.[21] Q_t is an aggregate shock variable with mean one. Since firms in the model are typically on an upward growth path, we label Q_t an "expansion cost shock". Given that firm expansion is a form of investment, Q_t resembles an investment-specific technology shock that features prominently in the DSGE literature and is sometimes interpreted as a stand-in for time-varying financial frictions.

The functional form we specify for $\zeta(.)$ implies a negative relationship between firm age and employment growth, which is widely documented in the BDS data.[22] Such a cost specification, which is independent of current size, has been used in many models, see e.g. Cooper, Haltiwanger, and Willis (2007) and Kaas and Kircher (2011) for an extensive discussion.

[21]It is assumed that firms that shut down do not pay adjustment costs and that new-born firms have an initial employment level of zero.

[22]See e.g. Haltiwanger, Jarmin, and Miranda (2013). In the quantitative application we choose a cost that is quadratic in the absolute change in employment. Alternatively, one could specify cost to be convex in employment growth, but such a specification implies a roughly constant growth rate as firms age.

Firms maximize the expected present value of profits:

$$V_{i,a}(n_{i,a-1,t-1},\mathcal{F}_t)=\max_{n_{i,a,t}}\left[\begin{array}{c} y_{i,a,t}-W_t n_{i,a,t}-Q_t\zeta(\Delta n_{i,a,t}) \\ +(1-\rho_a)\,\mathbb{E}_t\Lambda_{t,t+1}V_{i,a+1}\left(n_{i,a,t},\mathcal{F}_{t+1}\right)\end{array}\right], \quad (1)$$

where $V_{i,a}(n_{i,a-1,t-1},\mathcal{F}_t)$ is the asset value of a firm of type i and age a, $\mathbb{E}_t$ is the conditional expectations operator and $\mathcal{F}_t$ is the aggregate state to be described later. W_t is the economy-wide wage rate, $\Lambda_{t,t+1}$ is the firm's stochastic discount factor between period t and $t+1$. The first-order necessary condition for the firm's optimal choice of labor can be written as

$$W_t+Q_t\zeta_{n_{i,a,t}}(\Delta n_{i,a,t})=\alpha_i\frac{y_{i,a,t}}{n_{i,a,t}}+\beta\Lambda_{t,t+1}(1-\rho_a)\mathbb{E}_tQ_{t+1}\zeta_{n_{i,a,t}}(\Delta n_{i,a+1,t+1}), \quad (2)$$

where $\zeta_{n_{i,a,t}}(.)$ denotes the derivative with respect to current employment. The condition simply equates the marginal costs of firm expansion to the marginal benefits. Marginal costs consist of the wage and the marginal expansion cost. Marginal benefits equal the sum of the marginal product of labor and the expected discounted marginal reduction in expansion costs to be paid next period.

3.2 Entry decisions

At the heart of the model are variations in the types of business opportunities seized by startups. We therefore model the selection of firm types at the entry phase as the outcome of endogenous *choices* made by potential startups. Given the striking prominence of small firms in the BDS data, it is important to design a setup in which many startups endogenously choose a business type that has little potential to grow large.[23] Our approach therefore differs from standard firm dynamics models along the lines of Hopenhayn and Rogerson (1993), in which a firm's technology type is randomly drawn from an exogenous distribution. Before we describe the specifics of our setup, let us first provide the intuition of the resulting equilibrium outcomes.

[23]In the data there is a large fraction of firms that remain small even after many years of continuous operation. For example, in 2011 about 60 percent of firms in the BDS older than 25 years had fewer than 10 employees.

We consider an equilibrium with positive entry in which the number of startups of type i, denoted $m_{i,0,t}$, is determined by the following relation:

$$m_{i,0,t} = \gamma_0 V_{i,0}(0, \mathcal{F}_t)^{\gamma_1} \psi_{i,t}, \tag{3}$$

where $\gamma_0, \gamma_1 > 0$ are functions of structural parameters unrelated to firm type. This relation states that the number of startups in type i is determined by two factors. The first is the value of a startup of this type, $V_{i,0}(0, \mathcal{F}_t)$. If a business cycle shock triggers an increase in $V_{i,0}(0, \mathcal{F}_t)$ relative to the values of other firm types, a larger fraction of startups is drawn into type i. Thus, composition shifts among startups arise *endogenously* via differential fluctuations in firm values across types. Note also that an overall increase in firm values increases the total number of entrants. The second factor determining the number of startups is a variable $\psi_{i,t}$, which fluctuates exogenously and stochastically.

The microeconomic foundations that lead up to the above equilibrium relationship are as follows. Starting up a firm requires the sacrifice of a cost $\chi > 0$, capturing initial costs of doing market research, formulating a business plan etc. Upon paying this cost, a potential entrant chooses one business opportunity from a finite measure of possibilities given by $\psi_{i,t}$. Each business opportunity allows for at most one successful startup.[24] It is assumed that potential entrants cannot coordinate on which business opportunities to select. That is, not all individual opportunities are seized whereas others are pursued by several aspiring startups. This results in $m_{i,0,t}$ being strictly smaller than both the number of business opportunities, $\psi_{i,t}$, and the number of startup attempts, denoted $e_{i,t}$. It follows that an attempted startup of a business type i is successful only with probability $\frac{m_{i,0,t}}{e_{i,t}}$. Unsuccessful startups exit before production takes place. This way of modeling firm entry is similar in spirit to models of innovation and research and development (see e.g. Klette and Kortum, 2004; Saint-Paul, 2002).

Free entry implies that in equilibrium the cost of a startup attempt

[24] At a deeper level, the exclusivity of business opportunities could arise from patents claimed by individual firms. Alternatively, exclusivity could be generated by market size limitations coupled with fixed costs in production. For tractability, we do not model these factors explicitly.

equals the expected benefits:

$$\chi = \frac{m_{i,0,t}}{e_{i,t}} V_{i,0,t}\left(0, \mathcal{F}_t\right), \text{ for } i = 1, 2, .., I, \tag{4}$$

From the above it follows that firm types associated with high values on average attract more entry attempts, lowering the success probability $\frac{m_{i,0,t}}{e_{i,t}}$. This in turn encourages entry of firm types with less potential to grow large, of which we see many in the data. In equilibrium aspiring entrants are indifferent between selecting any of the business opportunities, akin to models of directed search.

The coordination friction among aspiring startups is concisely summarized by an entry "matching" function, borrowed from the search and matching literature. This function relates the number of startups within each type to the respective number of startup attempts and business opportunities. It is assumed to be increasing in both arguments and to display constant returns to scale. In particular, $m_{i,0,t} = e_{i,t}^{\phi} \psi_{i,t}^{1-\phi}$ where $\phi \in (0, 1)$ is the elasticity with respect to the number of startup attempts.[25]

Returning to the reduced form relationship in (3), it is straightforward to verify that it holds with $\gamma_0 = \chi^{\phi/(\phi-1)}$ and $\gamma_1 = \frac{\phi}{1-\phi}$. Given that the matching function elasticity ϕ is the same across firm types, it controls not only the strength of composition effects but also the volatility of overall firm entry, which is helpful when we calibrate the model.

While the total measure of business opportunities, denoted by $\Psi = \sum_i \psi_{i,t} > 0$, is assumed to be constant, its composition with respect to technology types is allowed to vary stochastically over time. This happens according to an exogenous "composition" shock X_t. This shock is a technological fundamental and will be specified in detailed in Section 3.4.

However, the reduced-form Equation (3) is sufficiently general to nest a variety of alternative microfoundations for the composition shock. In particular, type-specific shocks to the entry cost are observationally equivalent in this equation. Appendix I explicitly describes three models with alternative frictions leading to such changes in the (effective) entry cost and Section 5.3 explores empirically their link to our estimated composition

[25] See Saint-Paul (2002) for a similar specification in the context of firms' research and development.

changes.

3.3 Households

There is a representative household which consists of a continuum of members, some of which supply labor on a perfectly competitive market. The household maximizes the expected present value of life-time utility, subject to its budget constraint:

$$\max_{\{C_t,N_t\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma}-1}{1-\sigma} - \nu Z_t N_t \right) \quad \text{s.t.}$$

$$C_t = W_t N_t + \Pi_t, \tag{5}$$

where C_t is the total amount of goods purchased by the household, N_t denotes total employment within the household, $\sigma > 0$ is the coefficient of risk aversion, $\nu > 0$ is a parameter capturing the disutility of labor, Z_t is a stochastic preference shock, Π_t denotes firm profits and $\beta \in (0,1)$ is the household's subjective discount factor. Following the indivisible labor models as in Rogerson (1985), we assume that utility is linear with respect to labor supply. Profits and the wage are taken as given by the household. The optimal employment choice takes on the familiar form:

$$W_t = Z_t \frac{\nu}{C_t^{\sigma}}. \tag{6}$$

The first-order condition makes clear that Z_t drives a wedge between the marginal product of labor and the households intratemporal marginal rate of substitution. Hence it has been labeled a "labor wedge" in the literature and is typically thought of as a shock that may capture time-varying labor market frictions.

3.4 Shock processes

Before implementing the model quantitatively, we fill in the final details of the shock processes. First, we specify precisely how the composition shock affects the distribution of business opportunities. Assume, without loss of generality, that technology types reflect the degree of returns to scale in increasing order, such that $i = 1$ is associated with the lowest and $i = I$ with the highest degree of returns to scale. Finally, let ι be the median technology type.[26] The measure of business opportunities of type i in period t is given by

$$\psi_{i,t} = X_t \overline{\psi}_i \qquad \text{if} \quad i < \iota, \tag{7}$$

$$\psi_{i,t} = \overline{\psi}_i \frac{\Psi - \overline{\psi}_\iota - X_t \sum_{j=1}^{\iota-1} \overline{\psi}_j}{\sum_{j=\iota+1}^{I} \overline{\psi}_j} \qquad \text{if} \quad i > \iota, \tag{8}$$

where a bar indicates steady state values and $\Psi = \sum_i \psi_{i,t}$. The composition shock thus shifts mass from the upper half of returns to scale technologies to the lower half, in proportion to the respective steady state levels.

We assume that all four aggregate shocks follow an AR(1) process:

$$J_t = 1 - \rho_J + \rho_J J_{t-1} + \epsilon_t^J, \tag{9}$$

where ρ_J is a persistence parameter and ϵ_t^J are i.i.d. innovations distributed normally with mean zero and standard deviation σ_J, for $J = A, Q, X, Z$.

3.5 Equilibrium

Let $\mathbb{N}$ be the set of natural numbers, including zero. Using that all firms of the same age and technology type take the same decisions, the aggregate resource constraint, the labor market clearing condition and the law of motion for the measure of firms by age and technology type can be written, respectively, as:

$$\sum_{i=1}^{I} \sum_{a \in \mathbb{N}} m_{i,a,t} \left(y_{i,a,t} - Q_t \zeta(\Delta n_{i,a,t}) \right) - \sum_{i=1}^{I} e_{i,t} \chi = C_t, \tag{10}$$

$$\sum_{i=1}^{I} \sum_{a \in \mathbb{N}} m_{i,a,t} n_{i,a,t} = N_t \tag{11}$$

$$m_{i,a,t} = (1 - \rho_{a-1})\, m_{i,a-1,t-1} \text{ for } a \in \mathbb{N}_{>0} \text{ and } i = 1, 2, .., I. \tag{12}$$

[26] In the quantitative simulations, the number of technology types is odd. However, allowing for an even number of technology types is straightforward.

The aggregate state consists of the measure of firms of each age-technology combination, the employment levels of these firms in the previous period, as well as the values of the stochastic aggregate shocks, i.e. $\mathcal{F}_t = \left[A_t, Q_t, X_t, Z_t, \{m_{i,a-1,t-1}, n_{i,a-1,t-1}\}_{i=1,..I,\ a\in\mathbb{N}_{>0}}\right]$. The system of model equations we use to solve for the equilibrium consists of equations (1)-(12). A formal definition of the recursive equilibrium is given in Appendix F.

4 Quantitative Implementation

We parameterize the model using a combination of Maximum Likelihood (ML) estimation and matching moments in the (BDS) data. The dynamic model is solved using a first-order perturbation method around the stationary equilibrium (i.e. around the steady state growth paths of firms). The following subsections first describes the calibration of parameters used to match moments and then the parameters estimated using ML.

The aggregate state of the model includes the entire firm distribution making the solution of the model challenging. Nevertheless, our proposed solution strategy enables us to solve the model relatively quickly. This is accomplished by imposing a maximum firm age of $K = 50$ years, which makes the aggregate state finite.[27] The solution method enables us to track the aggregate state *entirely*, given the approximated policy functions, instead of being forced to revert to iterative methods in the spirit of Krusell and Smith (1998) which rely on an approximation of the aggregate state. The solution method is explained in detail in Appendix F.

Even though the aggregate state consists of over 900 state variables and all variables and shocks have continuous support, the speed of our computational strategy allows us to estimate parameters. As a result, we also obtain implied time series of all the variables, including the entire firm distribution which we use in our analysis.

[27]As a robustness check, we investigate a version of the model with $K = 75$. The results are extremely similar to our benchmark parametrization.

4.1 Parameters calibrated to match moments

Following the frequency of the BDS data we set the model period to one year. While the values of individual parameters typically influence the behavior of the entire model, it is instructive to discuss them separately in relation to the specific moments we target. For clarity, we divide the calibrated parameters into three groups. First, parameters pertaining to the household, second parameters specific to the technology types, and third firm-level parameters that are common to all firm types. All model parameters are summarized in Table 2. More details on the estimation procedure and the parameter estimates are provided in Appendix F.3.

4.1.1 Household parameters

Household preferences are chosen in line with conventional values in the macro literature. The household's discount factor, β, is set to 0.96, corresponding to an annual real interest rate of four percent. The household's coefficient of relative risk aversion, σ, is set to one implying log utility with respect to consumption. The preference parameter ν is backed out from the household's first order condition for a given wage and total consumption. We target a steady-state wage such that, given all other parameters, the model matches a profit rate of 3% taken from Hornstein, Krusell, and Violante (2005).

4.1.2 Firm-type parameters

Model parameters that describe firm technology types are the total factor productivity parameters, z_i, the returns-to-scale parameters, α_i, and the steady-state measure of business opportunities in each technology type, $\overline{\psi}_i$. In our benchmark model we normalize all the firm-specific TFP parameters to one, preserving only heterogeneity in returns to scale. The main motivating factor behind our choice is that we apply the model to the entire cross-section of private employers in the economy, rather than to confined industries. Evidence in favor of heterogeneity in returns to scale across sectors can be found in Basu and Fernald (1997).

That said, there is compelling evidence of large productivity differences even within narrowly defined sectors (see e.g. Syverson, 2011, for a survey).

Table 2: Model parameters

	parameter	value	target/estimate
β	discount factor	0.96	annual interest rate 4%
σ	relative risk aversion coefficient	1	log-utility
υ	disutility of leisure	0.06	3% profit rate, Hornstein, Krusell, and Violante (2005)
ζ	expansion cost level	0.004	average entrant size = 6.1, BDS
ξ_0	exit rate level coefficient	0.050	exit rates by age, BDS
ξ_1	exit rate power coefficient	0.170	exit rates by age, BDS
χ	entry cost	0.930	entry costs = 0.73% of GDP, World Bank
Ψ	measure of business opportunities	0.090	$M = 1$, normalization
ϕ	elasticity in entry function	0.300	std(entry)/std(y)=2.5, BDS
ρ_A	TFP shock (A), persistence	0.896	
σ_A	TFP shock (A), standard deviation	0.011	
ρ_Q	expansion cost shock (Q), persistence	0.593	
σ_Q	expansion cost shock (Q), standard deviation	0.158	
ρ_X	composition shock (X), persistence	0.415	
σ_X	composition shock (X), standard deviation	9.1e-6	
ρ_Z	preference shock (Z), persistence	0.751	
σ_Z	preference shock (Z), standard deviation	0.012	

α_i	returns to scale						average size in BDS size classes (16-20Y)		
	0.890	0.932	0.946	0.956	0.963	0.968	0.972	0.976	0.988
$\frac{m_{i,0}}{e_i}$	probability of starting up a type i firm						firm shares in BDS size classes (16-20Y)		
	0.625	0.357	0.218	0.123	0.070	0.040	0.022	0.013	0.002

Notes: Model parameters and their respective targets or sources. Since the magnitude of the measure of business opportunities of type i firms is hard to grasp, we rather report the probabilities of successfully starting up a business of type i ($m_{i,0}/e_i$) conditional on paying the startup cost.

It has also been documented that observed productivity differences are informative about firm growth rates. Therefore, in Appendix H we consider a version of the model with TFP heterogeneity and show that the results are hardly affected.

The presence of heterogeneity in technology types implies a cohort-level size distribution of firms, which we can confront with the BDS data. We set the total number of technology types equal to the number of size groups available in the BDS database, where we group the three largest size categories into one, giving us $I = 9$ technology types.

We exclude production functions with increasing returns to scale. To pin down the returns to scale parameters, we target average firm size in the 9 size categories reported in the BDS data for firms aged between 16 and 20 years (averaged over the period $2000-2010$). The implied values for the returns to scale parameters are shown in the bottom part of Table 2. They range between 0.890 and 0.988, which is within the range of estimates of Basu and Fernald (1997).

To pin down the steady state measure of business opportunities in each technology type, $\overline{\psi}_i$, we match the distribution of the number of firms between 16 and 20 years old, over the nine size categories reported in the BDS data, again averaged over the period $2000 - 2010$. The implied probabilities of successfully starting a business of a given type can be found at the bottom of Table 2. These probabilities may be interpreted as the survival rates in the first year of a firm's existence. The values imply an average entrant survival rate of 43%. While not a calibration target, this value is essentially identical to the empirical one of 44% based on the Business Employment Dynamics (BED) database.[28]

4.1.3 Parameters common to all firm types

Parameters that are common across all technology types are the expansion cost function, $\zeta(.)$, the exogenous firm exit rate, ρ_a, the entry cost, χ, the mass of potential entrants, Ψ, and the elasticity of the number of startups

[28] Unlike the BDS, the BED has quarterly information (for establishments), starting in 1992Q3, allowing one to calculate the survival rate of firms younger than one year. In a given year, the survival rate is calculated as the number of firms reported to be younger than 1 year in March of the given year divided by the sum of the number of establishments which started up in the relevant March-to-March period.

with respect to firm values, ϕ. We assume the expansion cost is quadratic, $\zeta(\Delta n_t) = \frac{\zeta}{2}(n_t - n_{t-1})^2$, as in e.g. Kaas and Kircher (2011). Here, ζ is a level parameter which we calibrate to match the average size of entrants of 6.1 as in the BDS data. Given that we also match average sizes of $16-20$ year old firms, ζ essentially pins down the average firm growth rate. To capture the age-dependency of exit rates observed in the data, we let the exit probability be $\rho_a = \xi_0 + \frac{\xi_1}{a}$. For $a < K$, the parameters ξ_0 and ξ_1 are chosen to closely match the empirical exit rates conditional on age in the BDS.

The last parameters in this category pertain to firm entry. We set the entry cost χ such that total entry costs are equal to 0.73% of GDP which is the average value for the US economy in the years 2004 to 2010 as documented by the "Doing Business" database of the World Bank. The measure of business opportunities Ψ is set such that the total mass of firms in the economy, M, is normalized to 1 in the steady state. Finally, recall that the parameter ϕ commands the degree to which the number of entrants changes in response to the changes in firm values. We therefore set ϕ such that the model approximates the relative volatility of the (log) number of entrants with respect to (log) real GDP as observed in the data.

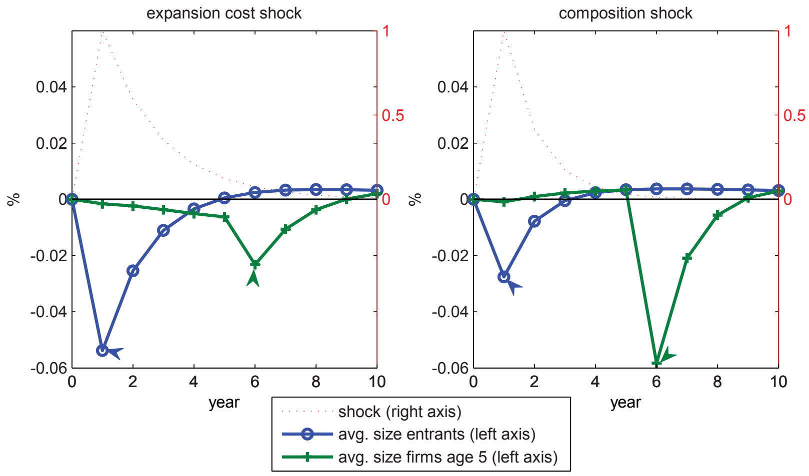
4.2 Parameters estimated using Maximum Likelihood

The remaining parameters pertain to the exogenous aggregate shocks and these are estimated using Maximum Likelihood. In particular, we estimate the persistence, ρ_J, and volatility, σ_J, parameters for $J = A, Q, X, Z$. We use four data series for this purpose: aggregate real GDP, the aggregate employment rate, the average size of entrants and the average size of five year old firms taken from the BDS data.[29] An important by-product of the estimation is that we obtain estimated time-series for all model variables, which we use in counterfactual exercises.

The estimated parameters are reported in Table 2. The parameter values of the aggregate TFP, preference and expansion cost shocks are in

[29] All time series are in logs and linearly detrended. We use a linear trend instead of the HP-filter in order not to introduce artificial serial correlation in the model's state-space representation. Nevertheless, the results are robust to using HP-filtered data.

Figure 4: Shock identification: impulse response functions



Notes: Impulse response functions to positive one-standard-deviation shocks. Dotted lines denote responses of the exogenous shock variables and have been re-scaled to increase by one on impact. Arrows denote the cohort born in the initial period of the shock

line with estimates in the literature.[30] The parameter values pertaining to the composition shock are more difficult to interpret directly, but in Section 5 we extensively discuss the quantitative importance of the composition shock, as implied by the estimation, and its potential interpretations.[31]

4.2.1 Shock identification

Because the results depend crucially on the estimated shocks, we now discuss their identification. While real GDP and the employment rate are directly informative about the aggregate TFP and preference shocks, the average size time-series are the main sources of identification for the the composition and expansion cost shocks.

To help understand the identification of the latter two, Figure 4 depicts the impulse response functions of average entrant size and average size of

[30]Interpreting the expansion cost shock as an investment shock, the reported values are broadly in line with e.g. Justiniano, Primiceri, and Tambalotti (2010), who also find that the investment shock is considerably more volatile than a TFP shock (in their case roughly 7 times).

[31]To get a sense of the magnitude of the composition shock we calculate average firm size in an economy assuming that the composition shock is permanently one standard deviation above (below) its steady state value. The implied average size deviates by about ±2.75% from its benchmark steady state level.

five year old firms to positive one-standard-deviation expansion cost and composition shocks. The figure shows that both shocks reduce average entrant size upon impact. Five years later, this fall is reflected in a reduction in the average size among five year old firms, i.e. within the cohort born in the initial period of the shock. While this is true for both types of shocks, the relative magnitudes differ starkly.

In the case of the expansion cost shock, the decline after five years is substantially smaller than the initial decline, whereas the opposite is true for the composition shock. The underlying reason is that while the expansion shock triggers some composition effect, it mainly affects post-entry employment decisions. The latter gradually die out as the cohort ages and the expansion cost shock reverts to its mean. The opposite is true for the composition shock which mainly affects the distribution of firm types among startups. As the cohort ages, the effects of the composition shock gain in strength as the lower number of high-growth potential firms starts kicking in. The fact that we use information on cohorts at different ages thus enables us to disentangle the various shocks and, by implication, estimate the overall degree of fluctuations in entrant composition.[32]

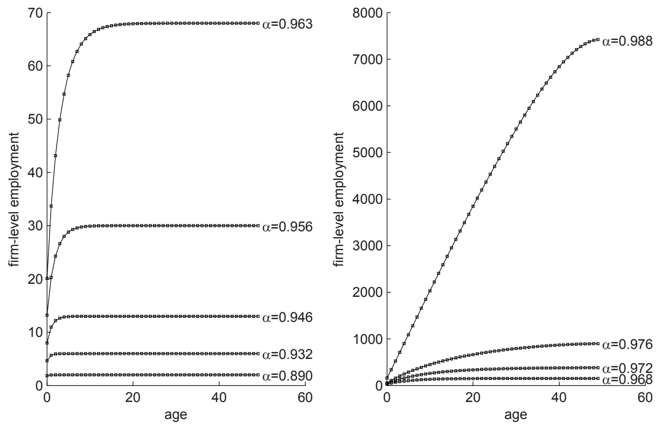
4.3 Properties of the steady-state equilibrium

Before further evaluating the dynamics of the model, we analyze the deterministic steady state equilibrium. Figure 5 plots the steady-state employment of firms by age and technology type. Firms of the lowest returns-to-scale type ($\alpha = 0.890$) start with an employment level of 1.8 which grows to only 2 later on in the firms' lives. On the other extreme, the most scalable firms ($\alpha = 0.988$) have nearly constant returns to scale and grow from 247 employees in the year of startup to a maximum of 7800 employees.

As firms with high returns to scale grow older, they account for an increasingly large share of the cohort's total employment. While the most scalable firms account for only about 7 percent of the cohort's total employment in the year of birth, they account for more than half by the age of fifty. Firms with low returns to scale, by contrast, are relatively more

[32]We have also checked that the estimation procedure does well in identifying the true parameters using Monte Carlo simulations in which the model is repeatedly used as the true data-generating-process.

Figure 5: Steady state: firm size by age and type



Notes: Steady state firm growth paths by type.

important during the early years of a cohort's life.

4.4 Model performance

We now evaluate the model's dynamic performance along several dimensions not directly exploited in the estimation. To this end we calculate several model statistics and compare them with their empirical counterparts (see Table 3). First, panel A shows that the model does a good job in matching our empirical stylized facts described in Section 2.[33]

Panel B shows that the predictions of the model are close to the data also for dynamics of firms older than five years. First, we quantify how cohort-level employment *changes* (in percent) are related to changes in the average size within these cohorts. Second, we correlate employment of older firms with that at entry summed over the appropriate five-year window. Third, we compare the dispersion in (log) average sizes across cohorts of old firms with that of startups. These results are reassuring because we do not use any direct information on firms between 5 and 15 in our calibration or estimation procedure and we only use cohort-level information about average size, not total cohort-level employment. The last statistic in panel B is especially interesting because it shows that in the data (and the benchmark model) the dispersion of firm sizes across

[33]Note that in order to be consistent with the model-generated statistics, the data has been linearly detrended instead of HP-filtered (see footnote 29).

Table 3: Model performance

	data	model
A: Employment dynamics of startups		
$\text{corr}(N_0, N)$	0.59	0.85
$\text{corr}(N_0, N_5)$	0.68	0.71
$\frac{\text{var}(S_{0-5})}{\text{var}(N_5)}$	69%	65%
B: Employment dynamics of older firms		
$\text{corr}(\Delta\log(N_{11-15}), \Delta\log(S_{11-15}))$	0.88	0.73
$\text{corr}(N_0, N_{11-15})$	0.89	0.65
$\text{std}(\log(S_0))/\text{std}(\log(S_{11-15}))$	1.83	1.48
C: Dynamics of large and small firms		
$\text{corr}(\Delta\hat{g}_{L-S}, u)$	−0.52	−0.67
$\text{corr}(\Delta\hat{g}_{OL-YS}, u)$	−0.24	−0.36

Notes: Untargeted model statistics and their empirical counterparts. corr(., .) denotes the correlation, var(.) denotes the variance and $\sigma(.)$ denotes the standard deviation of a given variable. N_a and S_a denote, respectively, employment and average size in firm cohorts of age a, N denotes the aggregate employment rate, $\frac{\text{var}(S_{0-5})}{\text{var}(N_5)}$ denotes the fraction of total cohort-level employment variation among five year old firms attributed to variations in average firm size, Δ denotes the difference operator, $\Delta\hat{g}_{L-S}$ denotes the differential growth rate of large and small firms as defined in Moscarini and Postel-Vinay (2012) and $\Delta\hat{g}_{OL-YS}$ denotes the differential growth rate between old large and young small firms as defined in Fort, Haltiwanger, Jarmin, and Miranda (2013).

.

cohorts *increases* (in percentage terms) as cohorts age.[34]

Panel C of Table 3 shows that the dynamics of the model are also consistent with recent findings by Moscarini and Postel-Vinay (2012), who document that larger firms are more cyclical than smaller ones. Furthermore, consistent with patterns in the data documented by Fort, Haltiwanger, Jarmin, and Miranda (2013), also in the model the aforementioned correlation is driven more by mature firms rather than young businesses.

Finally, we compare the model's predictions on real wages to the data.[35] The correlation between the real wage in the model and the data is encour-

[34]To highlight how the composition shock affects this feature of the model, we estimated an alternative model without changes in startup composition. In particular, instead of the composition shock, startups are hit by an additional adjustment cost shock in this version. The model is able to match the same four time-series as the benchmark model. However, because of missing composition changes the dispersion of average sizes across cohorts *declines* as they age. In particular, $\text{std}(\log(S_0))/\text{std}(\log(S_{11-15})) = 0.74$ in this model. Thus allowing for compositional effects is crucial in accounting for the long-term persistence present in the BDS data.

[35]As a data equivalent we use real hourly compensation in the nonfinancial corporations sector, as reported by the Bureau of Labor Statistics.

agingly high: 0.6. Also, the volatility of the real wage is only moderately higher in the data than in the model (1.33 in the data versus 1.08 in the model). The pro-cyclicality of wages, however, is too strong in the model relative to the data, a common finding in Neoclassical models of the business cycle.

5 Model results

This section first analyzes cohort-level fluctuations implied by the model. Our primary goal is to quantify the importance of the year of birth in determining a cohort's success in providing jobs and to understand the underlying drivers. Next, we investigate whether birth-determined factors are washed out in the aggregate, or whether they can help to understand fluctuations also in aggregate employment. Finally, we discuss the interpretation of the estimated shocks and potential additional drivers behind the estimated fluctuations in entrant composition.

5.1 Cohort-level implications

At any age after birth, a cohort's employment level is to some extent determined by the economic state in the year of birth. The remainder is due to shocks that realized after birth. Disentangling the relative importance of these two contributors empirically is difficult, if only because the aggregate state may include unobservable variables.

Within our estimated model, however, we can quantify the contribution of the economic state at birth precisely. To do so, first define cohort-level employment as $N_{a,t} \equiv \sum_{i=1}^{I} n_{i,a,t}$. We can then decompose cohort-level employment as $N_{a,t} = \mathbb{E}_{t-a}\left[N_{a,t}\right] + \widehat{N}_{a,t}$, where the first term is the expectation of $N_{a,t}$ conditional on information available in the year of birth. It follows that $\widehat{N}_{a,t}$ is the prediction error depending only on the shocks realized in the years after birth which are orthogonal to the state in the year of birth. Using this orthogonality we can decompose the unconditional variance of $N_{a,t}$ as:

$$Var\left(N_{a,t}\right) = \underbrace{Var\left(\mathbb{E}_{t-a}\left[N_{a,t}\right]\right)}_{\text{aggregate state at birth}} + \underbrace{Var\left(\widehat{N}_{a,t}\right)}_{\text{shocks after birth}}$$

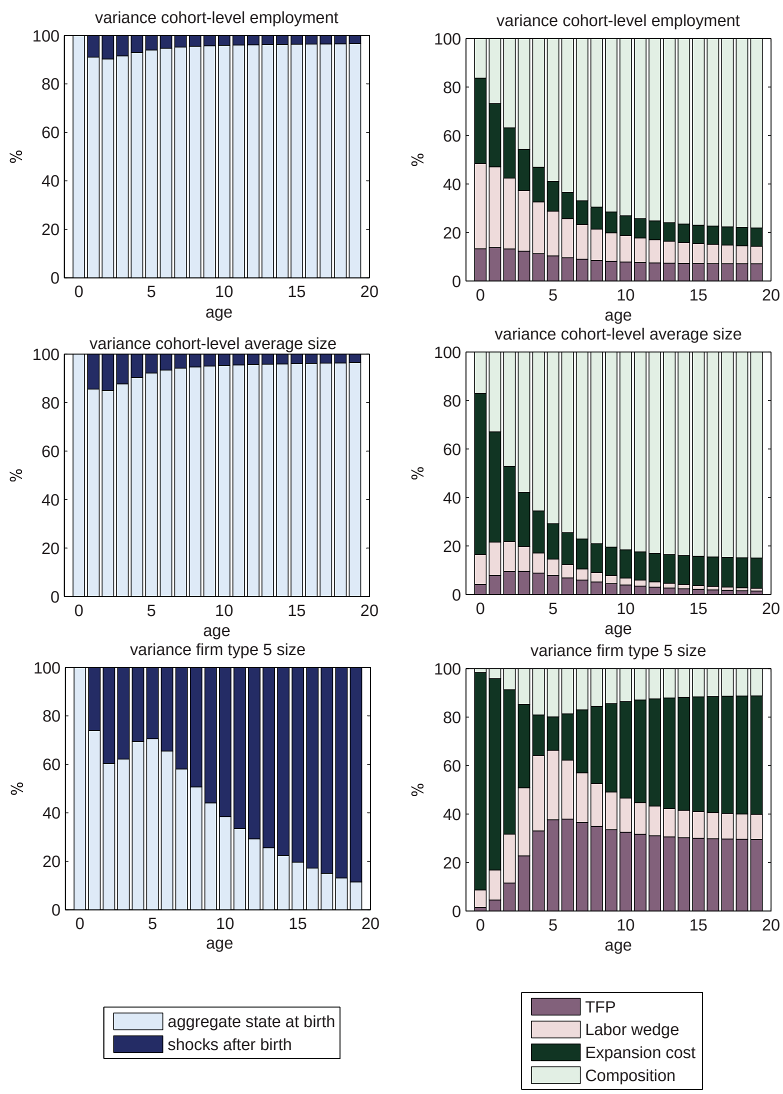
The top left panel of Figure 6 plots the results of the variance decompositions for cohorts up to twenty years after birth. The importance of the aggregate state at birth is overwhelming, contributing over 90 percent to employment variance, regardless of age. A very similar pattern is found for cohort-level average size (middle left panel) which is consistent with it being a strong driver of the employment patterns. However, for the employment of an *individual* firm of a certain type, the state at birth loses importance in the years following entry (bottom left panel),because it does not play any direct role in the firm's evolution. The persistence that remains is driven by that of the shock processes itself and by the endogenous part of the aggregate state.

Additional insight into the drivers of cohort-level persistence is obtained by quantifying the contributions of the four aggregate shocks (right panels of Figure 6). For cohort-level employment, the composition shock is of minor importance in the year of birth. As the cohort ages the composition shock gains prominence, accounting for nearly 80 percent of cohort-level employment twenty years after entry. A similar pattern is observed for average firm size (middle right plot). While the expansion cost shock remains relatively important even for cohorts of twenty year old firms, the impact of TFP and labor wedge shocks almost entirely dies out. For an individual firm of a given type, the composition shock does not account for much of the fluctuations (bottom right panel). However, its contribution does not fall to zero, which is because of general equilibrium responses triggered by the shock.

5.1.1 Revisiting the empirical variance decomposition

We now shed more light on the variance decomposition of Section 2.4, by conducting the same exercise using model-generated data, see Figure 7. Given that we have not targeted this variance decomposition directly, the figure looks reassuringly similar to its empirical counterpart presented in

Figure 6: Model: variance decompositions.

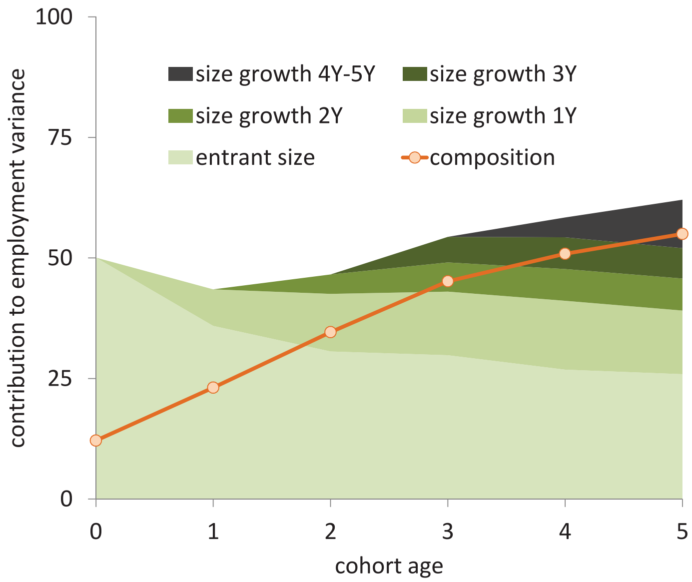


Notes: Contributions of the aggregate state at birth and post-entry shocks (left panels) and the contributions of the four aggregate shocks (right panels) to variation in cohort-level employment (top row), average size (middle row) and employment of an individual firm (bottom row).

Figure 3.

Unlike in the empirical exercise, we can now use the model to quantify precisely how much of the contribution of the intensive margin is due to compositional effects (solid line with circles).This contribution is computed using the counterfactual time series for average size obtained by fixing firm-level employment within each age/type bracket to its steady state value.

Figure 7: Contribution of average size to employment variation: model



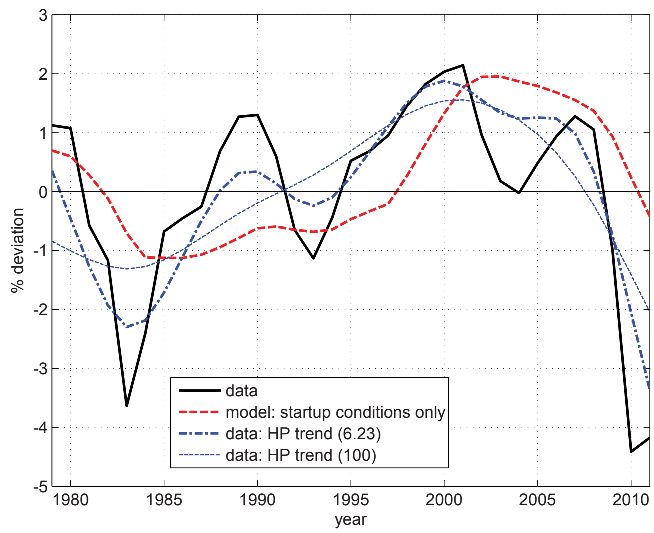
Notes: Contributions of average firm size at different ages to the variation in cohort-level employment of cohorts as a percentage of the total variation. Data are obtained from the estimated model. The orange solid line, "composition", plots the covariance between the counterfactual average size series (obtained by fixing firm-level employment within age/type brackets to its steady-state value) and overall cohort-level employment, scaled by the total variance of cohort-level employment.

The line thus represents the covariance between this counterfactual average size series and the model-predicted time series for cohort-level employment, scaled by its total variance. In the year of entry, only about a quarter of average size fluctuations is due to compositional effects. The importance of composition, however, grows markedly with age and by the age of five it accounts for almost 90 percent of the contribution of the intensive margin. At this age, the total contribution of composition far exceeds the contribution of entrant size (lightest shade), because compositional shifts importantly affect also *post-entry growth*.

5.2 Aggregate implications

This subsection discusses the aggregate implications of changes in startup conditions. As with cohort-level variables, we conduct variance decompositions of aggregate variables into the contribution of startup conditions and post-entry choices. We then move on to analyze the general equilibrium ef-

Figure 8: Employment rate: data, HP-trends and estimated contribution of startup decisions



Notes: Employment rate data, its HP trend, and a model-based counterfactual employment rate based on the fixing the age/type size to their respective steady state values.

fects present in the model and to discuss potential drivers of the estimated changes in startup composition.

5.2.1 Aggregate employment fluctuations

To analyze the importance of startup conditions for aggregate fluctuations, we first isolate (in an accounting sense) the contribution of startup conditions to fluctuations in aggregate employment. Figure 8 plots the fluctuations in *aggregate* employment accounted for by changes in startup conditions alone together with the actual aggregate employment rate.[36] The figure also plots two time series for the trend component of the actual employment rate, constructed using the HP filter with smoothing coefficients 6.23 and 100.

Figure 8 shows that the magnitude of aggregate employment fluctuations due to startup conditions alone is large, with a volatility of nearly two thirds of the actual employment rate series. More interestingly, the

[36]The employment rate due to fluctuations in startup conditions is calculated by fixing average size in each age/type bracket to its steady state value, but letting the distribution of firms (their number and composition) vary as predicted by the estimated model, and aggregating over all firms.

employment rate implied by changes in startup conditions closely resembles the empirical trend components in aggregate employment. The correlation coefficient between the counterfactual employment rate and the HP-filter trend is 0.61 and 0.56 for smoothing coefficient 6.23 and 100, respectively.[37] Thus, startup decisions appear important for understanding the low-frequency movements of *aggregate* employment, often ignored in business cycle analysis.

Decomposing the counterfactual further into the contribution of the number of firms and the contribution of the composition of firms reveals a roughly equal importance. The volatility of the counterfactual employment rate resulting from changes in the composition of firms only is 45 percent that of the counterfactual which allows for changes in both the composition and the number of firms.[38]

5.2.2 General equilibrium effects

Next, we investigate the importance of startup conditions further, by analyzing the aggregate effects of fully stabilizing the number of entrants and their composition. In particular, we run the estimated shocks through a version of the model which does not allow for changes in startup conditions. This counterfactual exercise differs from the one above because all variables in the economy are free to adjust in equilibrium.

Table 4 displays the effect on various aggregate variables between 1979 and 2011. The effects of stabilizing startup conditions are sizeable. Changes in aggregate employment and output can be as large as one third to one half of a percentage point on average in the five year windows.[39] However, the table also makes clear that stabilizing startup conditions does not fully eliminate the low-frequency components of aggregate employment. In general equilibrium, post-entry employment decisions adjust in response to the fixed startup conditions, offsetting part of its effects.

[37]The estimation uses linearly detrended employment rate data. However, the linear trend is very modest and therefore comparing the counterfactual with the HP-filter trend of the data used for estimation delivers very similar results.

[38]The extensive margin appears especially important at medium-term horizons whereas the intensive margin (average size) accounts for lower-frequency movements.

[39]Further splitting up the period into annual intervals reveals a maximum effect of 0.7 percentage points.

Table 4: Counterfactual scenarios: aggregate implications

	1979-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2011
average firm size	0.63	−0.21	0.36	1.32	1.71	0.85
employment	0.34	−0.02	−0.02	−0.31	0.11	0.19
output	0.46	0.09	0.04	−0.35	0.01	0.13
labor productivity	0.12	0.11	0.06	−0.04	−0.10	−0.06
real wage	0.07	0.15	0.07	−0.04	−0.15	−0.09

Notes: Percentage deviations from the benchmark model.

5.3 Potential drivers behind compositional variations

The main purpose of this paper is to identify changes in the composition of startups over the business cycle and to quantify their effects. Part of these fluctuations are driven by a shock to the distribution of startup technologies, but the reduced form of our model nests other microfounded frictions that form alternatives or complements to this shock. The benefit of our approach is that we do not rule out that all or a subset of these frictions operate at the same time, avoiding potential underestimation of overall composition fluctuations. Nonetheless, it is interesting to explore which frictions appear quantitatively most relevant.

In this section, we investigate the link between our estimated composition changes and the three frictions proposed in Appendix I: financial frictions, uncertainty and product market frictions. We do so by correlating measures of estimated variation in entrant composition with proxies for the above frictions proposed in the literature (see Table 5).[40] For financial frictions we consider the "GZ credit spread" developed in Gilchrist and Zakrajšek (2012) and the net percentage of domestic banks tightening lending standards for small firms from the Senior Loan Officer Opinion Survey on Bank Lending Practices. As uncertainty measures we take the uncertainty index computed by Bloom (2009) and a measure of firm-level uncertainty constructed in Jurado, Ludvigson, and Ng (2013). Finally, product market frictions are proxied by the ratio of advertising expenditures to GDP constructed by Hall (2013) and the consumer sentiment index taken from the

[40] We exclude the final six years of data from the analysis because for entrants born during this period we have no observation when they are aged 5. Following the reasoning in Section 4.2.1 the composition of entrants is not likely to be pinned down very precisely over this period. Including the final six years gives similar results, however.

Table 5: Correlations of startup composition with various indicators

	exogenous part	*total*
Financial frictions		
credit spread (Gilchrist and Egon Zakrajšek, 2012)	−0.27 (0.18)	−0.09 (0.67)
lending standards (Senior Loan Officer Opinion Survey)	−0.07 (0.81)	0.02 (0.93)
Uncertainty		
uncertainty index (Bloom, 2009)	0.05 (0.81)	0.22 (0.25)
firm-level uncertainty (Jurado, Ludvigson, Ng, 2013)	−0.32 (0.10)	−0.13 (0.50)
Product market frictions		
advertising-to-GDP (Hall, 2013)	0.43 (0.02)	0.42 (0.03)
consumer sentiment (University of Michigan Survey)	0.29 (0.13)	0.25 (0.19)

Notes: "exogenous part" refers to entrant size due to composition variation driven by only the composition shock, while "total" refers to that driven by all shocks. Numbers between brackets denote p-values. All proxies are logged (except for the lending standards which are expressed in percent) and detrended with a linear trend.

University of Michigan Survey.

The only statistically significant relationship is found in the case of the advertising-to-GDP ratio. Hall (2013) suggests a model in which an increase in product market frictions reduces advertising-to-GDP. The reported correlation in Table 5 is consistent with the interpretation that times during which acquiring new customers is difficult may be unattractive for the entry of highly scalable firms which need to build a large customer base.

6 Conclusion

This paper exploits the recent opportunity to break down aggregate employment data into cohort-level observations, in order to improve our understanding of fluctuations in macroeconomic aggregates. New stylized facts direct our attention to the birth stage of entering firms and in particular decisions affecting their scalability. Our results indicate that the impact of these decisions not only persists as cohorts mature, but actually grows over time since highly scalable firms need time to reach their full potential. Hence, compositional differences across cohorts become increasingly pronounced with age, accounting for slow-moving but large fluctuations in aggregate employment. Our estimates open up a promising avenue for further research, suggesting that product market frictions and vintage technology shocks are at the origin of these low-frequency movements.

References

Abbring, J., and J. Campbell (2004): "A Firm's First Year," Tinbergen Institute Discussion Paper 05-046/3.

Bartelsman, E., and M. Doms (2000): "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, 38(3), 569–594.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2009): "Measuring and Analyzing Crosscountry Differences in Firm Dynamics," In *Producer Dynamics: New Evidence from Micro Data*,NBER Chapters, pp. 15–76. National Bureau of Economic Research, Inc.

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

Bloom, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77(3), 623–685.

Campbell, J. (1998): "Entry, Exit, Embodied Technology, and Business Cycles," *Review of Economic Dynamics*, 1(2), 371–408.

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

Clementi, G. L., and D. Palazzo (2014): "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations," mimeo.

Cooper, R., J. Haltiwanger, and J. Willis (2007): "Search Frictions: Matching Aggregate and Establishment Observations," *Journal of Monetary Economics*, 54, 56–78.

Decker, R., J. Haltiwanger, R. Jarmin, and J. Miranda (2013): "The Secular Decline in Business Dynamism in the U.S.," mimeo.

Elsby, M., and R. Michaels (2013): "Marginal Jobs, Heterogeneous Firms and Unemployment Flows," *American Economic Journal: Macroeconomics*, 5(1), 1–48.

Fort, T., J. Haltiwanger, R. Jarmin, and J. Miranda (2013): "How Firms Respond to Business Cycles: The Role of Firm Age and Firm Size," *IMF Economic Review*.

GILCHRIST, S., AND J. WILLIAMS (2000): "Putty-Clay and Investment: A Business Cycle Analysis," *Journal of Political Economy*, 108(5), 928–960.

GILCHRIST, S., AND E. ZAKRAJŠEK (2012): "Credit Spreads and Business Cycle Fluctuations," *American Economic Review*, 102(4), 1692–1720.

HALL, R. (2013): "What the Cyclical Response of Advertising Reveals about Markups and other Macroeconomic Wedges," mimeo.

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.

HOLMES, T., AND J. STEVENS (2012): "An Alternative Theory of the Plant Size Distribution with Geography and Intra- and International Trade," mimeo.

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

HORNSTEIN, A., P. KRUSELL, AND G. VIOLANTE (2005): "Unemployment and Vacancy Fluctuations in the Matching Model: Inspecting the Mechanism," *Federal Reserve Bank of Richmond Economic Quarterly*, 91, 19–51.

HURST, E., AND B. PUGSLEY (2011): "What Do Small Businesses Do?," NBER working paper no. 17041.

JURADO, K., S. LUDVIGSON, AND S. NG (2013): "Measuring Uncertainty," mimeo.

JUSTINIANO, A., G. PRIMICERI, AND A. TAMBALOTTI (2010): "Investment Shocks and Business Cycles," *Journal of Monetary Economics*, 57(2), 132–145.

KAAS, L., AND P. KIRCHER (2011): "Efficient Firm Dynamics in a Frictional Labor Market," IZA discussion paper no. 5452.

KLETTE, T. J., AND S. KORTUM (2004): "Innovating Firms and Aggregate Innovatino," *Journal of Political Economy*, 112.

KRUSELL, P., AND A. SMITH (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106(51), 867–896.

LEE, Y., AND T. MUKOYAMA (2013): "Entry, Exit, and Plant-level Dynamics over the Business Cycle," mimeo.

Moen, E. (1997): "Competitive Search Equilibrium," *Journal of Political Economy*, 105(2), 385–411.

Moscarini, G., and F. Postel-Vinay (2012): "The Contribution of Large and Small Employers to Job Creation in Times of High and Low Unemployment," *American Economic Review*, 102(6), 2509–2539.

Poschke, M. (2012): "Who Becomes an Entrepreneur? Labor Market Prospects and Occupational Choice," *Journal of Economic Dynamics and Control*, forthcoming.

Rogerson, R. (1985): "Indivisible Labor, Lotteries, and Equilibrium," *Journal of Monetary Economics*, 21(1), 3–16.

Saint-Paul, G. (2002): "Employment Protection, International Specialization, and Innovation," *European Economic Review*, 46(2), 375–395.

Sedláček, P. (2014): "Lost Generations of Firms and Aggregate Labor Market Dynamics," mimeo.

Syverson, C. (2011): "What Determines Productivity?," *Journal of Economic Literature*, 49(2), 326–265.