The Growth Potential of Startups over the Business Cycle*

Petr Sedláček
University of Bonn †

Vincent Sterk
University College London ‡

September 23, 2016

Abstract

This paper shows that employment in cohorts of U.S. firms is strongly influenced by aggregate conditions at the time of their entry. Employment fluctuations of startups are pro-cyclical, they persist into later years and cohort-level employment variations are largely driven by differences in firm size, rather than the number of firms. An estimated general equilibrium firm dynamics model reveals that aggregate conditions at birth, rather than post-entry choices, drive the majority of cohort-level employment variation, by affecting the share of startups with high growth potential. In the aggregate, changes in startup conditions result in large slow-moving fluctuations in employment.

Keywords: Firm Dynamics, Heterogeneous Agents, DSGE
JEL Codes: E32, D22, L11, M13

---

*We thank Mark Aguiar, Eric Bartelsman, Christian Bayer, Francisco Buera, Mariacristina De Nardi, Mike Elsby, Stephan Fahr, John Haltiwanger, Philip Jung, Leo Kaas, Patrick Kehoe, Javier Miranda, Fabien Postel-Vinay, Benjamin Pugsley, Morten Ravn, Aysegul Sahin, Moritz Schularick, Kjetil Storesletten, Emily and Robert Swift, three anonymous referees as well as participants at many conferences and seminars 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, the decline in entry might create a long-lasting drag on aggregate employment and output. In this paper, we show that the roughly 2 million startups that *did* enter during the downturn are not only less plentiful, but may also be weaker in their potential to create jobs in the future. Specifically, we document that firms born in cohorts with weak job creation upon entry 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 Business Dynamics Statistics (BDS) we follow cohorts of firms, starting from their year of entry. The data span all U.S. non-government sectors and cover the years from 1979 until 2013. We document three new stylized facts: (i) employment created by startups is volatile and procyclical, (ii) these variations persist to a great extent as cohorts age, sharply contrasting with the strong mean-reversion in aggregate employment, and (iii) the majority of variation in employment across cohorts, conditional on age, is driven by changes in average firm size rather than in the number of firms within cohorts.

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 variations in firm size across cohorts are also driven by post-entry decisions made by a given mix of firms. To disentangle the two and to quantify the impact of composition changes, we *estimate* a general equilibrium firm dynamics model with aggregate uncertainty, using both aggregate and cohort-level data. We find that, because of changes in startup composition, the number of jobs created by a cohort is largely determined by the cyclical state of the economy in the year of its entry.

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

In the model, firm heterogeneity stems from differences in the demand for their products. Some firms produce "niche" goods which appeal only to a small subset of consumers, whereas others produce goods that may serve mass markets.[2] The type of good to be produced is chosen during the startup phase. Upon entry, demand is constrained by the size of the firm's consumer base, which can be expanded at the expense of a convex marketing cost. A firm's incentive to do so, however, depends critically on the type of good it has chosen to produce. This generates heterogeneity in growth profiles across startups. A coordination friction among aspiring startups gives rise to an equilibrium with simultaneous entry of firms with high and low growth potential.[3]

The composition of startups fluctuates endogenously over the business cycle in our model. This happens because aggregate shocks affect the profitability of different types of firms asymmetrically. The reason for the latter is that firm types differ in their optimal expenditure shares devoted to various cost components (production, entry and consumer base accumulation). This, in turn, generates heterogeneity in the sensitivity of firms to different shocks affecting these costs. We allow for aggregate shocks to each of the cost categories and estimate their importance from the data.

The estimation reveals that a demand shock, which affects the costs of consumer base accumulation, is quantitatively the most important driver of composition fluctuations. A positive demand shock increases the values of all firm types, but especially the values of the types producing "mass" goods, which optimally devote a large fraction of expenditures to relaxing their demand constraints. This induces a shift in the composition of startups towards types which have the potential to grow large. Simultaneously, aggregate expenditures on marketing increase, whereas firm profits decline. We document external support for this mechanism by showing that, in the

[2]By "niche" firms we mean businesses which are not very scalable as the nature of the good is such that the group of consumers to which they could potentially sell is small. This definition is broad in scope and may include customized or luxury goods, but not exclusively so.

[3]In the data, there are many firms that grow old but never become large. 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).

data, years of high advertising expenditures and low profits give rise to cohorts of startups that grow relatively large.

We also use the estimated model to show that macroeconomic conditions at the startup phase are important in shaping aggregate fluctuations. In particular, the contribution of startup conditions to aggregate employment fluctuations evolves similarly to the *trend* component of the employment rate, often discarded in business cycle analysis. Our results thus help to understand the drivers of macroeconomic fluctuations at a more complete range of frequencies.

An important prerequisite 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 methodological contribution of this paper is to design a computational strategy which allows us to solve the model quickly, and thereby enables us to estimate structural parameters.

The empirical results in this paper 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, which enables us to investigate how their later job creation is affected by aggregate conditions at the time of their birth.[4]

The model builds on a rapidly growing literature studying the importance of demand factors in accounting for firm-level and aggregate outcomes. Foster, Haltiwanger, and Syverson (2016) provide evidence that size differences between young and old plants cannot be well accounted for by differences in technological efficiency. They estimate a model which sug-

[4]Further related studies 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 startups 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. Bartelsman, Haltiwanger, and Scarpetta (2009) use a cross-country data set to study average post-entry behavior of young firms.

gests an important role for demand accumulation over the firm life cycle. Holmes and Stevens (2012) provide empirical evidence for the presence of "niche" and "mass" firms even within narrowly defined industries.[5] Abbring and Campbell (2005) estimate a model with firm-level demand shocks and find that pre-entry scale decisions are important for the variation in sales across existing firms. Other studies in which demand (accumulation) plays an important role include Arkolakis (2010), Drozd and Nosal (2012), Gourio and Rudanko (2014), Perla (2015), and Ravn, Schmitt-Grohé, and Uribe (2006). We integrate a highly tractable notion of consumer demand accumulation into a model with monopolistically competitive firms, endogenous entry and aggregate uncertainty, as in e.g. Bilbiie, Ghironi, and Melitz (2012).

Our model also relates to neoclassical models of firm dynamics, which typically feature heterogeneity in firms' technologies. A workhorse model is presented in Hopenhayn and Rogerson (1993). Models focusing on entry and exit decisions in the propagation of shocks include Campbell (1998), Clementi and Palazzo (2014) and Lee and Mukoyama (2013). In contrast to these 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.

The remainder of this paper is organized 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 Section 6 provides concluding remarks.

2 Empirical evidence

Startups are widely recognized to be important drivers of aggregate job creation on average (see e.g. Haltiwanger, Jarmin, and Miranda, 2013). This section presents three stylized facts regarding cyclical patterns of employment by young U.S. firms, both at the time of their entry and in later

[5] Our notion of "niche" goods is somewhat broader than the one of Holmes and Stevens (2012), who associate the term with goods that require a high degree of customization. We think of "niche" goods as all goods for which the attraction of new customers, beyond a limited initial group, generates only small increases in sales.

years. Our units of analysis are cohorts, that is, aggregates over firms born in the same year. Three stylized facts emerge:

Fact 1. *Employment created by startups is volatile and pro-cyclical.*

Fact 2. *Cyclical variations of startup employment persist into later years.*

Fact 3. *Cyclical variations of cohort-level employment are mainly driven by fluctuations in firm size, with an increasing importance as cohorts age.*

The first stylized fact complements empirical evidence presented in Campbell (1998) and Lee and Mukoyama (2013). These authors find that the number of new manufacturing plants and their job creation is pro-cyclical. Our analysis, by contrast, is not confined to a single industry and it applies to firms rather than establishments.[6] Pugsley and Sahin (2014) use the BDS to study secular changes in firm demographics and cyclical behavior of firms conditional on age. To the best of our knowledge, our second and third stylized facts have no precedent in the empirical literature. The end of this section discusses potential explanations for the stylized facts.

2.1 Data

The BDS database is based on administrative records of U.S. firms covering 98 percent of 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, and for the period between 1979 until 2013.[7]

The BDS is an annual database which allows us to follow cohorts of new firms for up to five years after they enter the economy. Thereafter, the BDS groups firms into age categories spanning five years. Nevertheless,

[6] An establishment is defined as a single physical location where business is conducted. A firm is a business organization consisting of one or more establishments that were specified under common ownership or control.

[7] The data represent a snapshot taken in March of each year. Availability starts in 1976, but we drop the initial three years following Moscarini and Postel-Vinay (2012), who cast doubt on the data quality for the years prior to 1979.

our stylized facts also hold for averages of firms aged 6-10 and 11-15 years. Appendix A.6 presents further evidence that our stylized facts hold beyond the age of five, based on micro data underlying the BDS.

Cyclical indicators from sources outside the BDS are constructed as March-to-March averages, consistent with BDS timing. Throughout the paper, detrending is conducted using the Hodrick-Prescott (HP) filter with a smoothing coefficient of 100 for annual data. The Appendix provides extensive robustness exercises with respect to the detrending method (A.1), construction of the measure of employment (A.2), the exact timing of firm entry (A.3), as well as an analysis of establishments (A.5), rather than firms.

2.2 The cyclicality of startup job creation

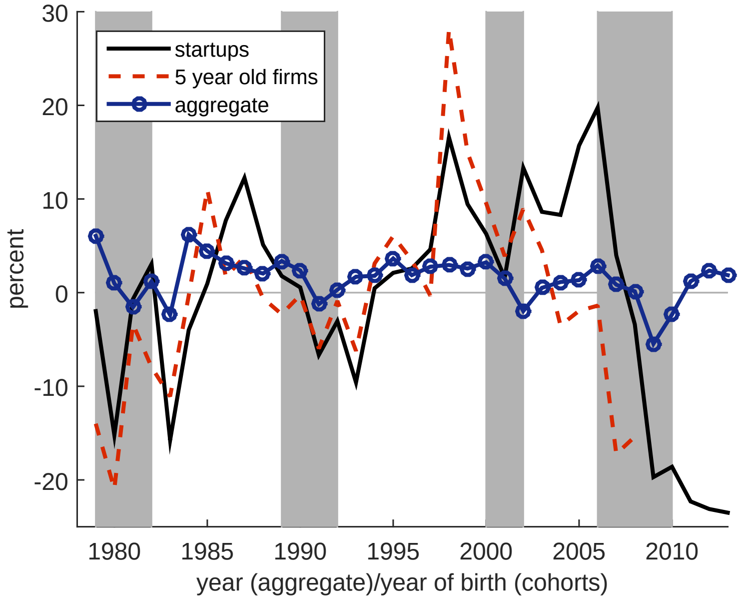
Let $M_{a,t}$ be the number of firms and $N_{a,t}$ total employment in a cohort of firms of age a in year t. Startups enter with age $a = 0$. We measure total employment of a given cohort as the cumulative net job creation since birth, i.e. $N_{a,t} = \sum_{i=0}^{a} NJC_{i,t-a+i}$, where $NJC_{a,t}$ is the net number of jobs created in firms of age a in year t.

To visualize the cyclicality of cohort-level employment, Figure 1 displays employment levels (in deviations from the respective means) 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 employment growth rate. Several patterns stand out. First, fluctuations in cohort-level employment are large, with a volatility exceeding four times the volatility of aggregate employment growth. Also, the cohort-level volatility does not appear to diminish with age. Second, job creation by startups and aggregate employment growth move together and drop during recession years, indicated by shaded areas. The correlation coefficient between entrant employment and aggregate employment growth (GDP growth) is 0.36 (0.45).

2.3 Persistence in cohort-level employment

To quantify the persistence of cohort-level employment, we compute the autocorrelation coefficients of total employment by startups in year t with total employment by the same cohort in year $t + a$. Figure 2 reports

Figure 1: Cohort-level employment by year of birth and aggregate employment growth by year

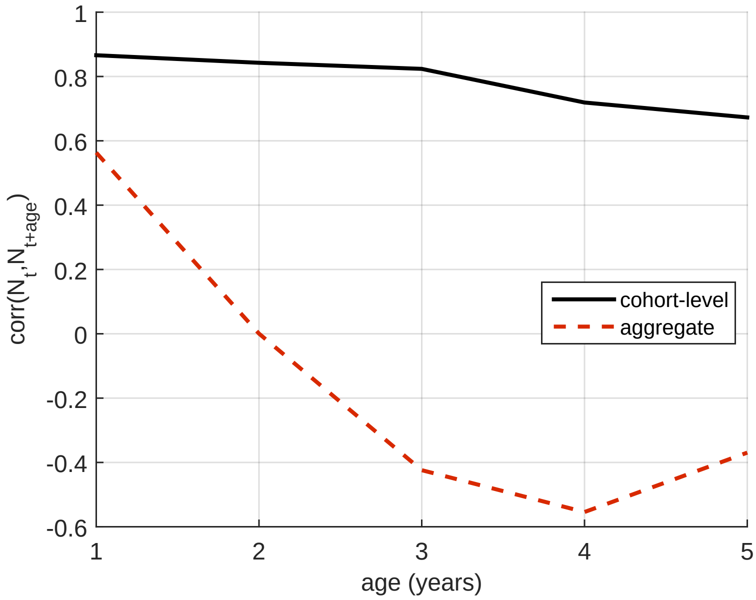


Notes: cohort-level employment in percent deviations from the mean across cohorts of firms of the same age and the year-on-year aggregate employment growth rate. Shaded areas are NBER recessions. Source: BDS, BLS.

these coefficients, as well as the autocorrelation coefficients for aggregate employment. For comparability, we take logs and HP de-trend all variables (i.e. we take out the trend of employment *across* cohorts of the same age and of aggregate employment).

Figure 2 shows that at the cohort level, the autocorrelation with startup employment remains high up to the age of five. Moving beyond the age of five, we find that the correlation of employment in 11-15 year old firms with entrant job creation 15 years earlier is 0.56 (not plotted). Thus, cyclical differences in employment across cohorts persist to a great extent into later years. In other words, we find little evidence that cohorts with initially low levels of employment catch up with other cohorts as they age. This lack of mean reversion contrasts aggregate employment, which displays no positive autocorrelation beyond a two year horizon.

Figure 2: Autocorrelations of cohort-level and aggregate employment



Notes: "cohort-level" refers to autocorrelations of total employment by cohorts of startups with total employment of the same cohort a years in the future, i.e. $\text{corr}(\hat{N}_{0,t}, \hat{N}_{a,t+a})$, where hats indicate log deviations from an HP trend taken across cohorts of the same age. "Aggregate" refers to autocorrelations of aggregate employment in year t and $t+a$, i.e. $\text{corr}(\hat{N}_{agg,t}, \hat{N}_{agg,t+a})$, again for data in log deviations from an HP trend. Source: BDS, BLS.

2.4 Decomposing cohort-level employment variation

Next, we investigate whether the observed variations of cohort-level employment are driven primarily by changes in the number of firms within the cohort (the extensive margin), or by average size (the intensive margin, i.e. the average level of employment per firm).

Toward 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 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:

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

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.[8] 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. 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.[9] Notice that the contribution of average firm size variation is increasing as the cohort ages (accounting for about 50% at birth and 63% at age 5). Extending the analysis to older firms reveals that the average size margin remains very important in determining variations in employment across cohorts, accounting for 70% among 11 to 15 year old firms (not plotted).

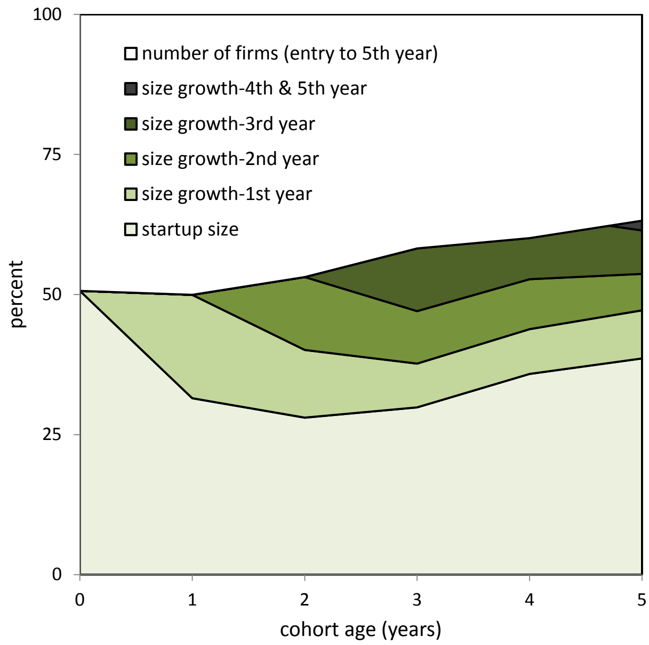
Within the total shaded area in Figure 3, different shades break down the contribution of the intensive margin by age, with the lightest shade denoting startup size. The contribution of the latter to cohort-level employment variation is large, accounting for 38% for five year old firms.

Before presenting the model, we briefly discuss two potential explanations that are outside our model. One possibility is that, during recessions, job creation within newborn cohorts declines because of a reallocation of activity between sectors. Another possibility is that our findings are driven by fluctuations in the entry of "necessity entrepreneurs", who start businesses as a means of escaping unemployment. However, Appendix (A.7) provides evidence that our stylized facts hold true, with a few exceptions, also within sectors and that the vast majority of employment variation of five year old firms is driven by large firms rather than small businesses.

[8] In our case, the residual η is negligible, not exceeding 0.01% of var($\hat{N}_{a,t}$).

[9] The vast majority of the contribution of the extensive margin is due to fluctuations in the number of startups. Changes in firm survival rates account on average for only 1% of employment variation among firms aged 1 to 5 years.

Figure 3: Contributions to variation in cohort-level employment



Notes: contributions (in percent) of changes in the number of firms and in average firm size at different ages to the variation in cohort-level employment. Source: BDS.

3 The Model

The empirical evidence presented in the previous section suggests that fluctuations in cohort-level employment are partly driven by changes in the composition of startups with respect to their growth potential. However, because firms' growth potential is unobserved, the data alone do not allow us to quantify the importance of such composition changes. For the same reason, the empirical facts can provide only limited information about the aggregate implications of decisions made at the entry stage.

To address these issues, we propose and estimate a general equilibrium model of the life cycles of heterogeneous firms which produce differentiated goods. Demand is restricted by the size of a firm's consumer base, which can be expanded by paying a convex marketing cost. Differences in the growth potential of startups stem from heterogeneity in the demand characteristics of goods. For some goods, demand is concentrated among only a small subset of consumers ("niche" goods). Other goods, by contrast, can potentially serve a broad demand base ("mass" goods). Firms which produce "niche" goods optimally invest little into expansion of the consumer base and therefore stay small. The opposite is true for firms which produce "mass" goods, which grow large over time.

Importantly, startups are free to decide which type of good to produce. As a result, the composition of startups with respect to their growth potential fluctuates endogenously over the business cycle. This happens as different firm types optimally allocate their expenditures differently over each of three cost categories: costs of entry, costs of production and costs of marketing. Aggregate shocks, which affect these cost categories differentially, then create type-specific fluctuations in firm profitability. This in turn generates endogenous fluctuations in startup composition.

The model includes several aggregate shocks affecting each of the three cost categories. We then use aggregate and firm-level data to estimate the relative importance of these shocks. As a by-product of the estimation, we back out the entire time-varying distribution of startups with respect to their growth potential. We exploit this to quantify the importance of firm entry in determining fluctuations at the cohort level and the aggregate level, and to understand the drivers of fluctuations in startup composition. The following subsections describe the model. Detailed derivations and a formal definition of the equilibrium can be found in Appendix B.

3.1 Household

There is a representative household which owns all firms, chooses consumption of all the goods varieties and supplies labor on a perfectly competitive market.[10] We first describe household preferences and then move on to optimal household decisions.

3.1.1 Household preferences

The representative household consists of a continuum of members, indexed by k. Household members have heterogeneous preferences over a continuum of available goods varieties, indexed by j. In order to enjoy utility from a particular good, a member has to be made aware that the good exists. This requires a costly marketing effort by the producer of the good.

[10]Firm dynamics models with more detailed descriptions of the labor market include e.g. Elsby and Michaels (2013), Kaas and Kircher (2015), and Sedláček (2015).

Time is discrete and indexed by t. Let Ω_t be the set of available goods and define $C_t = \int C_{k,t} dk$ as the consumption of the representative household. Here, $C_{k,t}$ is the consumption bundle of household member k, which is given by:

$$C_{k,t} = \left(\int_{j \in \Omega_t} \mathbf{1}_{k,j,t} \theta_{k,j}^{\frac{1}{\eta}} c_{k,j,t}^{\frac{\eta-1}{\eta}} dj \right)^{\frac{\eta}{\eta-1}},$$

where $\mathbf{1}_{k,j,t}$ is an indicator function equal to one if member k is aware of good j in period t and zero otherwise. Further, $\theta_{k,j}$ is a utility weight of member k for good j, $c_{k,j,t}$ is the quantity of good j consumed by member k, and $\eta > 1$ is the elasticity of substitution between goods varieties. Without loss of generality, we assume that the distribution of household members' utility weights for a particular good ($\theta_{k,j}$) can be summarized by a cumulative distribution function $F_j(\theta_{k,j})$ with support $[\theta_j^{\min}, \theta_j^{\max}]$, and $\theta_j^{\min} > 0$.

Firms can add household members to their pool of consumers by making costly marketing investments, which can be fully directed. Cost minimization then dictates that firms first attract consumers with the highest valuations for their goods (i.e. highest levels of $\theta_{k,j}$). Let the mass of household members aware of (and thus consuming) good j, be denoted by $s_{j,t}$. We will refer to $s_{j,t}$ as the firm's *consumer base*. Finally, let us define the utility weight for good j at the household level as:

$$\kappa_j(s_{j,t}) = \int_{\overline{\theta}(s_{j,t})}^{\theta_j^{\max}} \theta_{k,j} dF_j\left(\theta_{k,j}\right),$$

where $\overline{\theta}(s_{j,t})$ is the lowest utility weight among all household members who are aware of good j in period t.

The relevant firm-specific demand characteristics of a good are fully summarized by the function $\kappa_j(s_{j,t})$. Note that $\kappa_j(s_{j,t})$ is increasing in the consumer base, $s_{j,t}$, because each additional consumer demands a positive amount of good j. However, the extent to which higher levels of the consumer base increase $\kappa_j(s_{j,t})$ is fully determined by the preference distribution for good j, $F_j(\theta_{k,j})$.

Below we will show that the elasticity of κ_j with respect to s is a crucial determinant of a firm's growth potential. Further, Section 4.1 will clarify that low elasticities are associated with "niche" goods, for which demand is relatively concentrated among a small subset of consumers. For such goods, gains from demand investment are relatively low. On the other side of the spectrum are high-elasticity "mass" goods, for which demand investment generates relatively high returns.

3.1.2 Household decisions

We assume that utility is linear with respect to labor supply, N_t. Following indivisible labor models, we interpret N_t as the employment rate (see e.g. Rogerson, 1988). The household maximizes the expected present value of life-time utility, subject to its budget constraint, taking prices and wages as given:

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

$$\int_k \int_{j \in \Omega_t} p_{j,t} c_{k,j,t} dj dk = P_t W_t N_t + \Pi_t, \tag{1}$$

where $\beta \in (0,1)$ is the household's subjective discount factor, $\sigma > 0$ is the coefficient of risk aversion, $\nu > 0$ is a parameter capturing the disutility of labor, Z_t is a stochastic labor preference shock, W_t is the real wage, Π_t denotes nominal aggregate firm profits, $p_{j,t}$ is the price of good j and P_t is the aggregate price index. The latter can be shown to be $P_t = (\int_{j \in \Omega_t} \kappa_j(s_{j,t}) p_{j,t}^{1-\eta} dj)^{\frac{1}{1-\eta}}$.

The resulting optimal employment choice obeys the familiar first-order condition $W_t C_t^{-\sigma} = Z_t \nu$, with Z_t driving a wedge between the marginal product of labor and the intratemporal marginal rate of substitution. This "labor wedge" is typically thought of as a shock capturing time-varying labor market frictions and as such it directly affects firms' wage costs.

The first-order conditions for consumption lead to the following demand function:

$$c_{j,t} = \kappa_j(s_{j,t}) \left(\frac{p_{j,t}}{P_t} \right)^{-\eta} C_t. \tag{2}$$

The above implies that, as in standard models of monopolistically competitive firms, consumer demand for good j depends on aggregate consumption, C_t, and the relative price, $p_{j,t}/P_t$. The novel feature of our model is that demand is also affected by the firm's consumer base, $s_{j,t}$.

3.2 Firms

There is an endogenous mass of firms which supply differentiated goods varieties on a monopolistically competitive market. We first describe the behavior of incumbent firms and then discuss the startup phase.

3.2.1 Incumbent firms

Firms operate a technology $y_{j,t} = A_t n^G_{j,t}$, where $y_{j,t}$ is the amount of output produced by firm j, $n^G_{j,t}$ is the amount of labor used in goods production and A_t is aggregate total factor productivity (TFP).[11] The sales of the firm are constrained by the demand function for their good, Equation (2). Each firm produces a unique goods variety and hence we index both firms and goods varieties by j. Firms exit with an exogenous, but age-dependent probability ρ_a, where a denotes the firm's age.[12]

Firms can relax their demand constraints by exerting costly marketing efforts. Specifically, the consumer base of firm j evolves as:

$$s_{j,t} = s_{j,t-1} + Q_t g_{j,t}, \tag{3}$$

where $g_{j,t}$ denotes the amount of marketing and Q_t is an aggregate "demand shock".[13] Given $s_{j,t-1}$ and $g_{j,t}$, a decline in Q_t reduces the consumer base $s_{j,t}$ and hence tightens the firms' demand constraint (2). While we take Q_t as an exogenous object to be estimated from the data, one could think of

[11]It is straightforward to extend the model to include firm-specific TFP levels, $\overline{a}_j$, such that firm-level output is given by $y_{j,t} = A_t \overline{a}_j n^G_{j,t}$. However, it can be shown that in our application $\overline{a}_j$ is isomorphic to a scaling factor in the preference distribution F_j and hence we opt to normalize $\overline{a}_j$ to one for all j.

[12]Clearly, this assumption is a simplification as exit rates vary over time and to be related to firm productivity (see e.g. Bartelsman and Doms, 2000). Therefore, Appendix E.1 shows that allowing for stochastic variation in exit rates consistent with the data does not substantially affect our results. This is consistent with the variance decomposition in Subsection 2.4 which implies that variation in exit rates explains on average only 1% of fluctuations in cohort-level employment for firms aged 1-5 years.

[13]In our setup, firms use marketing to make consumers aware that a good exists. We find it natural to assume that consumers do not forget about goods, implying zero depreciation of the consumer base. Nevertheless, Appendix E.5 shows that similar results are obtained when we consider positive depreciation.

it as a shift in consumers' preferences affecting the susceptibility to firms' marketing efforts.

We further assume that firms enter with no consumer base and that a unit of marketing requires an amount of labor given by $n^M_{j,t} = \zeta(g_{j,t})$, where $\zeta(.)$ is an increasing and convex function. The convexity of this cost induces firms to grow only gradually as they age, in line with the positive relation between the age and size of young firms in the BDS. The adjustment cost further makes the consumer base a firm-level state variable, as in Gourio and Rudanko (2014).

Firms maximize the expected present value of real profits:

$$V_j(s_{j,t-1}, \mathcal{F}_t; a) = \max_{n^G_{j,t}, n^M_{j,t}, p_{j,t}, g_{j,t}, s_{j,t}} \begin{bmatrix} y_{j,t}p_{j,t}/P_t - W_t\left(n^G_{j,t} + n^M_{j,t}\right) \\ +(1-\rho_a)\,\mathbb{E}_t\Lambda_t V_j\left(s_{j,t}, \mathcal{F}_{t+1}; a+1\right) \end{bmatrix}$$

subject to (i) their demand constraint $y_{j,t} = \kappa_j(s_{j,t})\left(\frac{p_{j,t}}{P_t}\right)^{-\eta} Y_t$, where Y_t is aggregate demand, (ii) the evolution of their consumer base, Equation (3), and (iii) the evolution of the aggregate state of the economy, denoted by $\mathcal{F}_t$ and described later. In the above equation, V_j is the asset value of firm j and $\Lambda_t = \beta(\frac{C_t}{C_{t+1}})^{\sigma}$ is the stochastic discount factor of the representative household.

The optimal pricing decision takes on the familiar form of a constant markup over the nominal marginal cost of production: $p_{j,t} = \frac{\eta}{\eta-1}P_t W_t/A_t$. This in turn implies that all firms set the same price. Relative prices can then be expressed as $p_{j,t}/P_t = (\int_{j\in\Omega_t}\kappa_j(s_{j,t})dj)^{\frac{1}{\eta-1}}$. This condition stems from households' love of variety and is similar to the "variety effect" in Bilbiie, Ghironi, and Melitz (2012). In our model, however, this effect depends not only on the set of available goods varieties, Ω_t, but also on the distribution of firms' consumer bases.

Finally, the optimal amount of marketing investment satisfies the following first-order condition:

$$\frac{\zeta'(g_{j,t})}{Q_t} = \epsilon^{\kappa,s}_{j,t}\frac{n^G_{j,t}}{s_{j,t}}\frac{1}{\eta-1} + (1-\rho_a)\,\mathbb{E}_t\Lambda_{t,t+1}\frac{\zeta'(g_{j,t+1})}{Q_{t+1}}\frac{W_{t+1}}{W_t}, \tag{4}$$

where a prime denotes the first derivative and $\epsilon^{\kappa,s}_{j,t} \equiv \frac{\kappa'_j(s_{j,t})s_{j,t}}{\kappa_j(s_{j,t})}$ is the elasticity of $\kappa_j(s_{j,t})$ with respect to the consumer base $s_{j,t}$, which we refer to as the *marketing elasticity of demand*. At the optimum, marginal costs of expanding the consumer base are equal to the present value of profits that it generates. Importantly, the latter depends on the marketing elasticity of demand. Firms with higher elasticities choose to invest relatively heavily in expansion of the consumer base and hence grow relatively large.

3.2.2 Entry decisions

Having described the behavior of incumbent firms, we now explain the entry phase and in particular how startups choose the type of good to produce. For tractability, we restrict the number of goods types to be finite, indexed by $i = 1, 2, ..., I$. Underlying this restriction is an assumption that household preferences for individual goods belong to one of a finite number of distributions.

In every period, startups can seize a limited and time-invariant number of business opportunities of each goods type, denoted by ψ_i. Business opportunities are exclusive, allowing for at most one producer each.[14] After paying a stochastic entry cost, labeled X_t, potential startups are free to choose any of the business opportunities.[15] They cannot, however, coordinate among themselves.

Therefore, not all startups will succeed because multiple competitors may attempt to seize a single business opportunity. We assume that the probability of successfully starting up is increasing in the number of business opportunities, but decreasing in the number of startup attempts (similar to models of innovation such as Klette and Kortum, 2004; Saint-Paul, 2002). In particular, the success probability is given by $m_{i,0,t}/e_{i,t}$, where $m_{i,0,t} = \psi_i^{\phi} e_{i,t}^{1-\phi}$ is the number of new firms and $\phi \in (0, 1)$. Free entry then gives rise to the following condition:

$$X_t = \frac{m_{i,0,t}}{e_{i,t}} V_i \left(0, \mathcal{F}_t; 0\right), \text{ for } i = 1, 2, .., I. \tag{5}$$

[14] Exclusivity of business opportunities can arise e.g. from the ownership of patents, or market size limitations coupled with fixed costs in production. For tractability we do not model these factors explicitly.

[15] The entry cost is to be paid before entry and is denominated in units of the household's consumption bundle. Appendix B.1 shows that the firm's demand function then becomes $y_{j,t} = \kappa_j(s_{j,t}) \left(\frac{p_{j,t}}{P_t}\right)^{-\eta} Y_t$.

The above implies that, in equilibrium, aspiring startups are indifferent between goods types. The reason is that goods types with higher firm values also generate more intense competition for available business opportunities, lowering the success probability.

It is straightforward to show that the elasticity of the number of startups within a type, $m_{i,0,t}$, with respect to the startup value of that type, $V_i\left(0, \mathcal{F}_t; 0\right)$, is given by $\frac{1-\phi}{\phi}$. Thus, endogenous fluctuations in the composition of startups arise to the extent that values of firms producing different types of goods fluctuate differently over the business cycle. On the contrary, fluctuations in the entry cost do not generate any direct composition effects because all firms are affected symmetrically.

3.3 Aggregate shocks and market clearing

There are four aggregate shocks in the model, namely shocks to productivity (A), demand (Q), entry costs (X) and labor preferences (Z). We assume that all four aggregate shocks follow AR(1) processes in logarithms:

$$\ln J_t = (1-\rho_J)\ln \overline{J} + \rho_J \ln J_{t-1} + \varepsilon_t^J, \text{ for } J = A, Q, X, Z, \tag{6}$$

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. $\overline{J}$ denotes the mean of the given shock process and it is normalized to one for all shocks except for the entry cost shock. The parametrization of the latter is discussed in the calibration section.

Before describing the market clearing conditions, we exploit that all firms producing the same type of good i and of the same age a make identical decisions. Accordingly, we replace the firm index j by the type and age indices i and a. The labor market clearing condition can then be written as:

$$N_t = \sum_{i=1}^{I}\sum_{a=0}^{\infty} m_{i,a,t}\left(n_{i,a,t}^G + n_{i,a,t}^M\right), \tag{7}$$

where $m_{i,a,t}$ is the mass of firms of type i and age a. Because entry costs are assumed to be paid in terms of the aggregate consumption bundle, the aggregate resource constraint can be written as:

$$C_t + X_t \sum_{i=1}^{I} e_{i,t} = Y_t, \tag{8}$$

where aggregate demand is given by $Y_t = \sum_i \sum_a m_{i,a,t} y_{i,a,t} p_{i,a,t} / P_t$. The law of motion for the mass of firms by age and good type can be written as:

$$m_{i,a,t} = (1 - \rho_{a-1})\, m_{i,a-1,t-1} \text{ for } a = 1, 2, ... \text{ and } i = 1, 2, .., I. \tag{9}$$

Finally, the aggregate state consists of the mass of firms of each age-type combination, the consumer capital levels of these firms in the previous period, as well as the values of the stochastic aggregate shocks, i.e. $\mathcal{F}_t = \left[A_t, Q_t, X_t, Z_t, \{m_{i,a-1,t-1}, s_{i,a-1,t-1}\}_{i=1,..I,a=1,2,...}\right]$.

3.4 Endogenous fluctuations in startup composition

This subsection explains intuitively why the composition of startups fluctuates endogenously in the model. Appendix B.3 provides formal results for a special case of the model which allows for closed-form solutions.

The free entry conditions (5) makes clear that incentives to start up firms producing particular goods types depend on the relative profitability (firm value) of such businesses. Therefore, the composition of startups changes endogenously to the extent that values of firms producing different types of goods are differently sensitive to aggregate shocks.

In the estimated model, a quantitatively important reason why the relative values of different firm types fluctuate over time, is that the profits of "mass" firms are relatively sensitive to demand shocks. This happens because demand shocks shift the effective cost of consumer base expansion and "mass" firms optimally devote relatively large fractions of resources to this cost category. The latter result can be understood from Equation (4) and dates back to Dorfman and Steiner (1954), who show that the optimal marketing (advertising) expenditure share is proportional to the respective elasticity of demand, which is relatively high for "mass" firms.

4 Quantitative Implementation

We parameterize the model using a combination of Maximum Likelihood (ML) estimation and matching of moments in the (BDS). This section describes the calibration and estimation of model parameters and discusses properties of the model along dimensions not directly targeted in the parametrization.

The aggregate state of the model includes the entire firm distribution, creating a challenge in solving the model numerically. Our proposed solution strategy is based on first-order perturbation around the stationary equilibrium (i.e. around the steady-state growth paths of firms) and on imposing a maximum firm age of $K = 50$ years. This makes the aggregate state finite and enables us to solve the model relatively quickly even tough the model consists of more than 900 state variable and all shocks have continuous support. We are also able 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.

Detailed descriptions of the solution and estimation methods, as well as robustness exercises with respect to the calibrated model parameters, are presented in Appendix D.

4.1 Parameters calibrated to match moments

We set the model period to be one year, in line with the frequency of the BDS data. 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 that we target. For clarity, we divide the calibrated parameters into three groups. We start with parameters specific to firm types. Next, we proceed to parameters common to all firms and finally to parameters pertaining to the household. All model parameters are summarized in Table 1.

4.1.1 Firm-type parameters

Heterogeneity across firms derives from differences in consumer demand, summarized by the function $\kappa_i(s)$, which is increasing in firms' consumer base (s) and directly affects firms' demand constraints. The function $\kappa_i(s)$ depends, in turn, on the underlying distribution of household members' preferences for good type i.

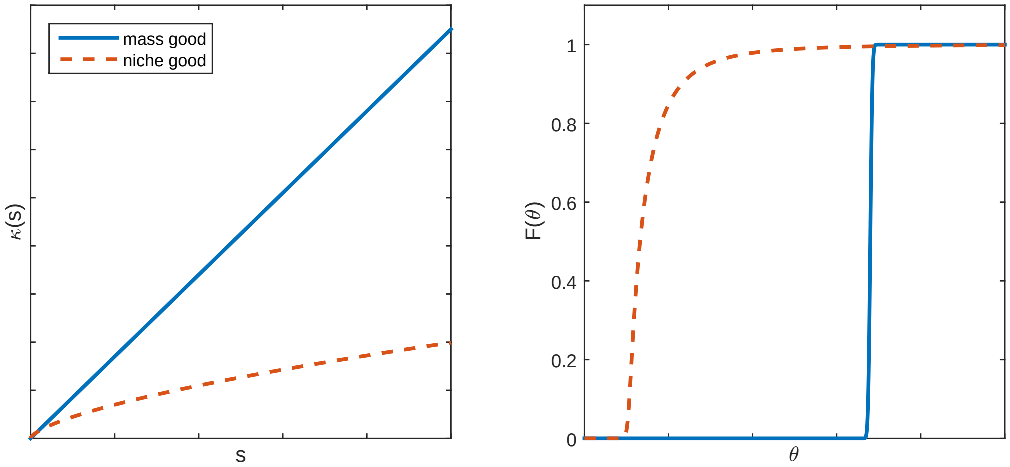
Given that we cannot observe the preference distributions directly, we opt for a parsimonious approach and specify the aggregated preference function as $\kappa_i(s) = \overline{\kappa}_i s^{\mu_i}$, with $\overline{\kappa}_i, \mu_i > 0$. The firm's elasticity of demand with respect to the consumer base is then given by $\epsilon_{i,t}^{\kappa,s} = \mu_i$. The condition for optimal marketing investment (4) makes clear that this elasticity is a crucial determinant of the firm's returns to marketing, and thus of its incentive to grow large as it ages.

To illustrate how the demand function is related to the distribution of preferences over individual goods, Figure 4 plots $\kappa_i(s)$ and F_i for two parameterizations. In the first case, there is no preference dispersion, i.e. F_i is degenerate around a single point. As a result $\kappa_i(s)$ is linearly increasing in the consumer base s, i.e. μ_i equals one. We refer to this type of good as a "mass" good, since the marginal consumer attracted by additional marketing brings in the same amount of demand as existing consumers. As a result, the returns to marketing do not fall as the firm grows larger.

The second case illustrates a parametrization for which μ_i is smaller than one: a "niche" good.[16] In this case, the associated preference distribution F_i features a certain degree of dispersion. Recall that firms first attract consumers with the highest valuations (levels of θ). In the illustration, the valuations of these initial consumers for the "niche" good are similarly high as for the "mass" good. Thus, for low levels of s the total demand for the "niche" and the "mass" good is similar. However, the marginal amount of demand coming from additional consumers falls rapidly in case of the "niche" good, but not in case of the "mass" good. As a result, "niche" firms face relatively low returns to marketing and optimally stay smaller.

[16] The scaling parameter of the "niche" good, $\overline{\kappa}_j$, is set such that low levels of the consumer base generate similar demand as with the "mass" good.

Figure 4: Demand heterogeneity illustration



Notes: illustration of demand heterogeneity between "mass" and "niche" goods. The left panel depicts the demand function $\kappa(s)$. The right panel shows the CDF of the associated preference distribution $F(\theta)$. The "mass good is characterized by $\mu = 1$, while the "niche" good has an elasticity of $\mu = 0.65$. The level of $\overline{\kappa}_i$ of the "niche" good, relative to the "mass" good, is set to 6.

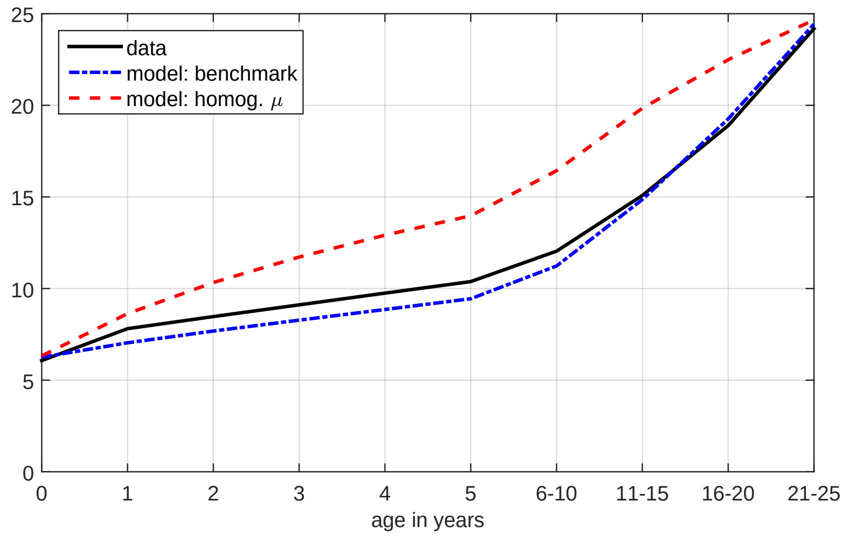
In light of the above discussion, we pin down firm-type parameters by targeting moments of the firm size distribution observed in the BDS data. Towards this end, we consider $I = 9$ firm types, which is the number of size groups available in the BDS database, where we group the three largest size categories into one.

The parameters pertaining to firm (goods) types include $\overline{\kappa}_i$, μ_i and also the mass of business opportunities for each firm type, ψ_i. First, assuming that old firms had enough time to grow towards their optimal size to reveal their type, we use the firm size distribution of 21 to 25 year old firms to pin down $\overline{\kappa}_i$ and ψ_i (up to a scaling factor Ψ). While the former essentially shifts the scale of production, the latter determines the fraction of firms in each size category.

To pin down the levels of μ_i, we exploit the fact that this parameter affects firms' growth profiles. Because we cannot observe growth profiles of individual firm *types* in the BDS, we use information on the average growth profile in the economy, i.e. average firm size by age.[17]

[17]Specifically, we use the BDS information on 0 (startups), 1, 2, 3, 4, 5, 6 to 10, 11 to 15 and 16 to 20 year old firms. Appendix C shows that the calibrated values of marketing elasticities of demand in the benchmark model fall well within the range of empirical estimates found in existing studies. For some firm types, μ_i is larger than

Figure 5: Average size by age: data and model



Notes: average firm size by age in the data and the "benchmark" model and an alternative calibration with "homogeneous μ" (i.e. all firms have identical marketing elasticities of demand), calibrated to match entrant size.

To highlight that average size by age reveals information about heterogeneity in the elasticities μ_i, Figure 5 shows average size by age in the data, the benchmark model and in an alternative model version in which all firm types face a "homogeneous μ" (calibrated to match average entrant size). The figure shows that a model in which all firms grow at the same pace cannot generate the relatively flat average growth profile observed in the data. The success of the benchmark model rests on small firms, which constitute the majority of all businesses, reaching their optimal size relatively quickly. Thereafter, the average growth profile is shaped by (rare) fast-growth firms which gradually gain on importance in the aggregate as they become large employers.

4.1.2 Parameters common to all firms

Parameters that are common across all firm types are the exogenous firm exit rate, ρ_a, the elasticity of substitution between goods varieties, η, the marketing costs function $\zeta(g)$, the mass of potential startups, Ψ, the mean

one. It is straightforward to show that this arises if marketing investments attract not only new consumers, but to some extent also raise the demand coming from existing consumers.

of the entry cost shock, $\overline{X}$, and the elasticity of the number of startups with respect to firm values ϕ.

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 firms below the maximum age ($a < K$), the parameters ξ_0 and ξ_1 are chosen to closely match the empirical exit rates conditional on age in the BDS. The elasticity of substitution η is set to 11, implying a 10% markup over the wage, a common target in the literature. The marketing cost function is assumed to be quadratic with a level normalized to 1, i.e. $\zeta(g) = g^2/2$.[18] The reason for the latter is that the level of adjustment costs is not separately identifiable from the level of demand in our model. The implied average costs of marketing investment amount to 2.7 percent of gross profits. This is similar to the estimated 3 percent costs for (capital) investment in Cooper and Haltiwanger (2006).

The last three parameters in this category pertain to firm entry. The measure of business opportunities is normalized such that Y equals one in the steady state. From the free entry condition (5) it is clear that the level of the entry cost determines the probability of successfully starting up a business of a given type (for a given firm value). Interpreting this probability as the within-year survival rate, $\overline{X}$ is set such that the model matches the average success probability in the data.[19] Finally, we set ϕ, which controls the strength of startup composition effects, such that the model matches the volatility of entrant size observed in the BDS.

4.1.3 Household parameters

Household preference parameters 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

[18] In Appendix E.2 we explore adjustment cost functions with different degrees of curvature and show that similar results are obtained.

[19] Toward this end, we draw on information from the Business Employment Dynamics (BED). Unlike the BDS, the BED has quarterly information (for establishments), starting in 1992Q3, allowing us to calculate the survival rate of establishments younger than one year. We calculate the within-year survival rate assuming that the quarterly survival rates are constant in a given year. Table 1 reports the implied type-specific probabilities of successfully starting up, rather than the type-specific measures of business opportunities (ψ_i) which are difficult to interpret and which depend on the normalization constant Ψ.

household's coefficient of relative risk aversion, σ, is set to one, implying log utility with respect to consumption. Finally, the disutility of labor, ν, is backed out from the household's labor supply condition with the wage normalized such that $p_j/P = \eta/(\eta-1)W = 1$.

4.2 Parameters estimated using Maximum Likelihood

The remaining parameters pertain to the four exogenous aggregate shocks and they are estimated using Maximum Likelihood. We estimate the model using four data series: aggregate real GDP, the aggregate employment rate, the number of startups and the average size of five year old firms. The number of startups and firm size of five year old firms is taken from the BDS. All time series are in logs and linearly detrended. The estimated parameters are reported in Table 1 and they are in line with estimates in the literature. Further discussion of the estimation (results) can be found in Appendix D.4.

While real GDP is primarily informative about aggregate TFP, aggregate employment is closely related to the labor preference shock. Even though all shocks affect the number of startups, the model matches this variable exactly due to the presence of the entry cost shock. Finally, using the average size of five year old firms helps to pin down fluctuations in the demand shock (Q).

To understand the last point, the left panel of Figure 6 shows impulse response functions of average size of firms between 0 and 5 years of age to a positive demand shock. While average size increases at all ages upon impact, there is an "echo" effect, creating sequential upwards spikes in average size for the age categories 1 to 5 years. These spikes reflect the fact that the composition of the cohort born in the initial year of the shock is skewed towards "mass" firms, which grow to be relatively large. While echo effects are also created by other shocks, these effects are quantitatively smaller. Thus, using information on the size of firms several years after birth helps to discipline the relative strength of the demand shock, and implicitly the importance of composition effects.

The impulse responses also make clear that the echo effects gain on strength as the affected cohort ages, because the greater share of "mass"

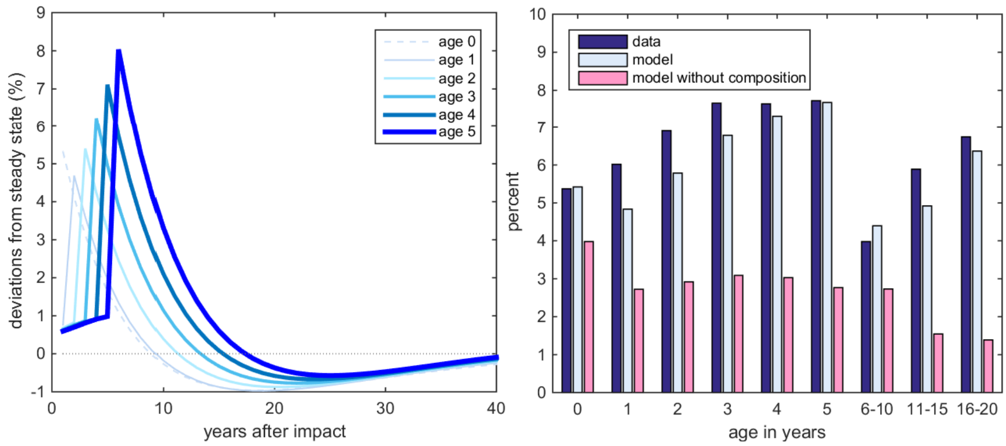
Table 1: Model parameters

	parameter	value	target
β	discount factor	0.96	annual interest rate 4%
σ	relative risk aversion coefficient	1	log-utility
υ	disutility of labor	0.939	wage normalization, $p_j/P = 1$
Ψ	measure of business opportunities	0.006	output= 1, normalization
η	price elasticity of substitution	11	10% markup
ξ_0	exit rate function, level	0.050	exit rates by age, BDS
ξ_1	exit rate function, curvature	0.170	exit rates by age, BDS
$\overline{X}$	entry cost, mean	1.206	entrant survival rate = 0.21%, BED
ϕ	elasticity of entry function	0.156	std(entrant size)/std(output)=1.4, BDS
ρ_A	TFP shock (A), persistence	0.944	
σ_A	TFP shock (A), standard deviation	0.011	
ρ_Q	demand shock (Q), persistence	0.867	
σ_Q	demand shock (Q), standard deviation	0.057	
ρ_X	entry cost shock (X), persistence	0.521	
σ_X	entry cost shock (X), standard deviation	0.014	
ρ_Z	labor preference shock (Z), persistence	0.983	
σ_Z	labor preference shock (Z), standard deviation	0.011	

		goods types i								
	parameter value/target	1	2	3	4	5	6	7	8	9
$\overline{\kappa}_i$	scale of $\kappa_i(s)$	2.21	6.27	12.05	10.47	22.16	2.98	1.53	0.83	0.71
	average size when old	2.2	6.3	13.1	29.8	68.0	150.2	336.6	658.8	3511.2
$\frac{m_{i,0}}{e_i}$	startup probability (%)	83.6	29.3	15.1	11.0	4.6	6.2	4.1	3.0	0.7
	firm share when old (%)	45.2	23.9	15.1	9.8	3.2	1.8	0.5	0.2	0.3
μ_i	curvature of $\kappa_i(s)$	0.009	0.012	0.054	0.364	0.344	0.964	1.189	1.355	1.470
	average size by age				see	Fig. 5				

Notes: model parameters and their respective targets. The bottom part of the table shows type-specific parameters and their targets for $i = 1, ..., 9$. Instead of reporting the mass of business opportunities in each type (ψ_i), which is subject to a normalization, we report the implied startup probabilities ($m_{i,0}/e_i$). The demand elasticities (μ_i) are not targeted by type, but rather they are implicitly pinned down by matching average size by age (see Figure 5).

Figure 6: Impulse response functions and volatility of firm size by age



Notes: left panel shows impulse response functions of average size, by age, to a positive one-standard-deviation demand shock in percent deviations from steady state. The right panel shows average size volatility by age in the "data", "model" and in a model where composition is fixed ("model without composition").

firms steepens the growth profile of the cohort. This means that composition effects leave their mark by creating a positive relation between age and the volatility of average firm size. The right panel of Figure 6 illustrates that the estimated model captures well this empirical pattern of volatility of average firm size by age (only averages over five cohorts are reported for firms older than six years, as in the BDS). To highlight the role of composition effects, the right panel of Figure 6 also depicts the volatility of average size by age when the composition of startups is held fixed. Without fluctuations in the composition of startups, the volatility of average size *declines* with age.[20]

Finally, even for firms older than five years, which were not used in the calibration or estimation procedure, the model correctly predicts an increasing pattern of volatility of average size. The extent of this increase is actually somewhat smaller than in the data. This reassures us that the estimated degree of composition changes, which drives the increasing pattern of average size volatility, is rather conservative.

[20]This is also true if we re-estimate the model. In computing this decomposition, only the startup composition is held fixed at its steady state, but all other variables are left to adjust.

Table 2: Firm dynamics in the data and model

	data	model
A: Employment dynamics of young firms		
Fact 1: $\text{corr}(N_0, \Delta N)$	0.39	0.46
Fact 2: $\text{corr}(N_0, N_5)$	0.59	0.62
Fact 3: $\frac{\text{var}(S_{0-5})}{\text{var}(N_5)}$	70%	72%
B: Employment dynamics of old firms		
$\text{corr}(N_0, N_{11-15})$	0.86	0.73
$\text{corr}(\Delta \log(N_{11-15}), \Delta \log(S_{11-15}))$	0.88	0.82

Notes: untargeted model statistics and their empirical counterparts. corr(.,.) denotes the correlation, var(.) denotes the variance and Δ is the first-difference operator. N_a and S_a denote, respectively, employment and average size in firm cohorts of age a, N denotes the aggregate employment rate and $\frac{\text{var}(S_{0-5})}{\text{var}(N_5)}$ denotes the fraction of total cohort-level employment variation among five year old firms attributed to variations in average size.

.

4.3 Model properties

This subsection assesses the model's performance along several dimensions not directly exploited in the estimation and provides external support for the demand channel in the data.

4.3.1 Firm dynamics

Table 2 displays several model statistics and compares them with their empirical counterparts. Panel A shows that the model is successful in matching the empirical stylized facts described in Section 2.

Panel B of Table 2 shows that the predictions of the model are close to the data also for dynamics of firms older than five years. First, we show that cohort-level employment at entry is correlated with employment of the same cohort even 11-15 years later (summed over the appropriate five-year window).[21] Second, to gauge the extent to which variation in employment of old firms is related to changes in firm size, we correlate *changes* in cohort-level employment (in percent) with changes in the average size within these cohorts. This relation is highly positive both in the data and in the model.

Finally, we compare the model's predictions on real wages to the data.

[21] We choose to report correlations for the age group 11-15 year old firms as a compromise between a higher firm age and a long enough time-series. Correlations for young firm age groups are also close to those in the data.

The correlation between the real wage in the model and the data is 0.53. Also, the volatility of the real wage relative to output is close to that in the data (0.71 in the model versus 0.61 in the data). The pro-cyclicality of wages, however, is too strong in the model relative to the data, a common finding in business cycle models without wage rigidities.[22]

4.3.2 Inspecting the demand channel

This subsection provides external support for the demand channel which, as will be quantified in the next section, is the key driver of endogenous composition changes of startups.

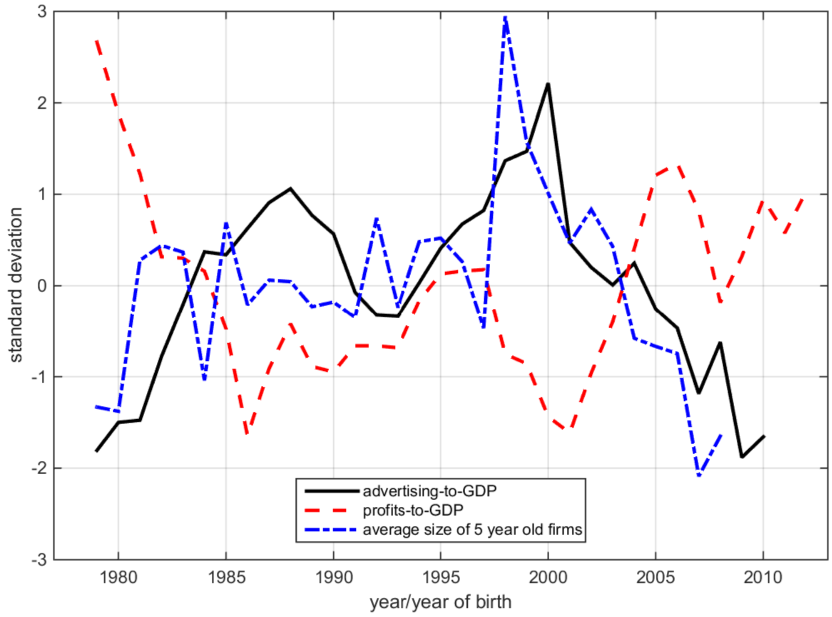
A positive demand shock eases the expansion of firms' sales capacities. Firms types that need a larger consumer base to reach their efficient scale benefit relatively strongly from the positive demand shock. This creates stronger incentives to start up "mass" goods producing firms. At the same time, aggregate profits decline, as firms seize the opportunity to invest in consumer base expansion at a low cost. In the estimated model, the correlation between average firm size of five year old firms in year t and aggregate marketing expenditures (profits) relative to GDP at the *time of birth* of the firm cohort, i.e. year $t-5$, is 0.75 (-0.60).

Figure 7 corroborates this prediction in the data by plotting aggregate advertising-to-GDP and profits-to-GDP together with average firm size of five year old firms, where the latter has been shifted back to the respective year of birth. In the data, the correlation between the size of five year old firms in year t and aggregate advertising expenditures (profits) relative to GDP in year $t-5$ is 0.60 (-0.53). Both correlations are reasonably close to the aforementioned counterparts in the model.

In addition to the presented aggregate evidence, Appendix A.10 provides empirical support for the proposed mechanism using 4-digit industry data from the Quarterly Workforce Indicators linked with the input-output tables of the Bureau of Economic Analysis. Specifically, we show that, as predicted by our model, industries with relatively high marketing expendi-

[22]The correlation between output and real wages is 0.96 in the model and 0.41 in the data. In the data, we measure the real wage as real hourly compensation in the non-farm business sector.

Figure 7: Demand channel in the data



Notes: advertising-to-GDP ratio (ranging from 1979 to 2010 and taken from Hall, 2014), the profit share (ranging from 1979 to 2012 and computed as corporate after tax profits divided by nominal GDP) and average size of five year old firms (BDS) shifted back to the year of startup (and thus ranging from 1979 to 2008).

ture shares tend to display stronger cohort effects.[23]

5 Model results

The purpose of the model is to quantify fluctuations in the composition of startups with respect to their growth potential and to investigate to what extent such changes shape cohort-level and aggregate dynamics. Our first goal is to establish the importance of the year of birth in determining a cohort's success in providing jobs in later years and to understand the underlying sources of variation. Next, we investigate the importance of startup conditions for aggregate outcomes.

[23]While the QWI allows for a fine sectoral disaggregation, its time and spacial coverage are relatively sparse, preventing the construction of aggregate time series as observed in the BDS. We therefore use the BDS for our main analysis.

5.1 The importance of startup conditions for cohort-level fluctuations

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 because of 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.

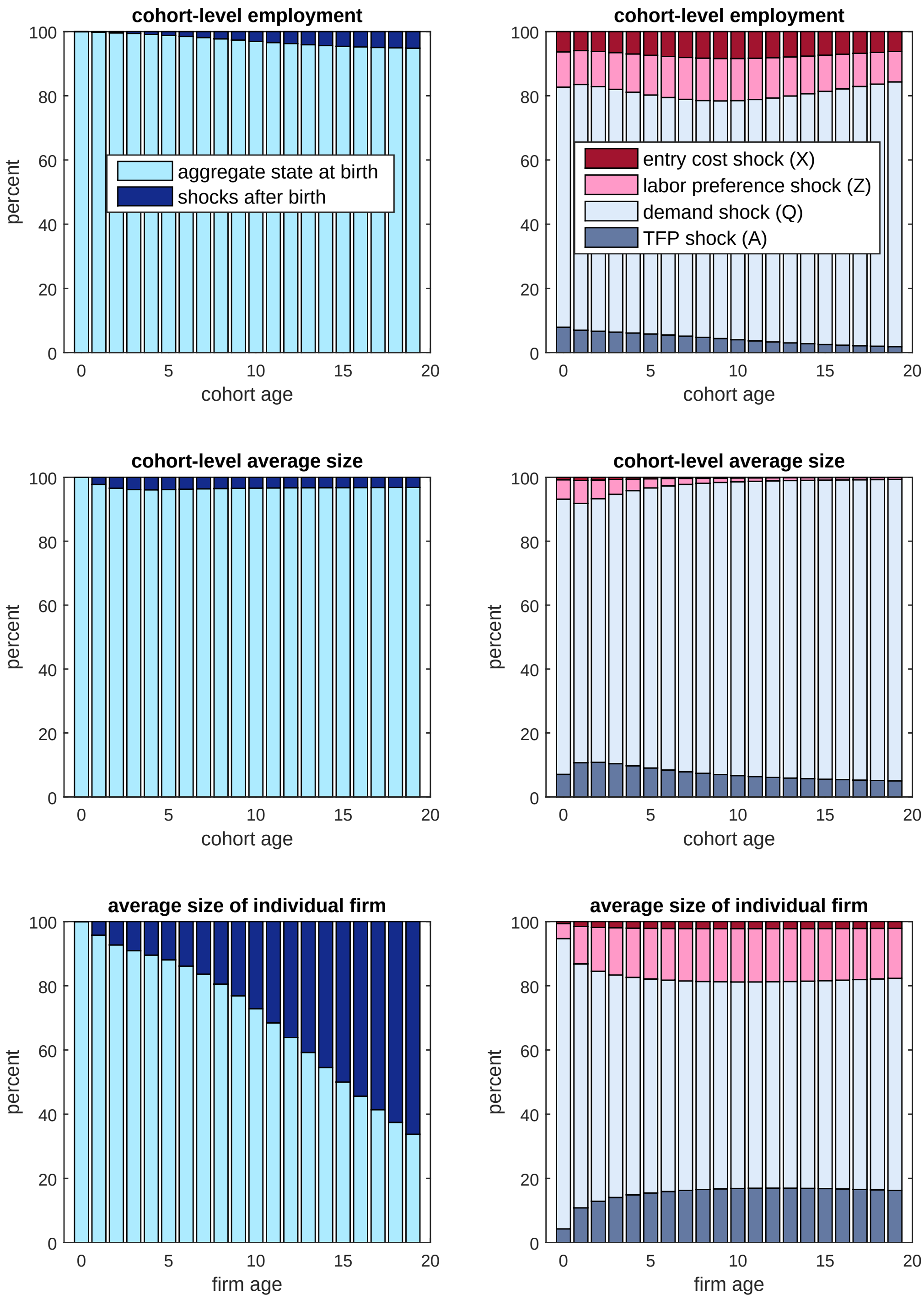
Within our estimated model, however, we can quantify the contribution of the economic state at birth precisely. Let us first define cohort-level employment as $N_{a,t} \equiv \sum_{i=1}^{I} m_{i,a,t} n_{i,a,t}$. We can then decompose cohort-level employment as $N_{a,t} = \mathbb{E}_{t-a}[N_{a,t}] + \widetilde{N}_{a,t}$, where the first term is the expectation of $N_{a,t}$ conditional on information available in the year of birth and $\widetilde{N}_{a,t}$ is the prediction error. The latter is a function of only 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:

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

The top left panel of Figure 8 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 to more than 90 percent of the employment variance, regardless of age. A very similar pattern is found for cohort-level average size (middle left panel), which is consistent with average size being a strong driver of the employment patterns. However, for the average size of an *individual* firm of a certain type, the state at birth loses importance in the years following entry (bottom left panel). This happens as composition effects are not directly relevant. The persistence that remains is driven by the inherent persistence of the shock processes and by the endogenous part of the aggregate state.

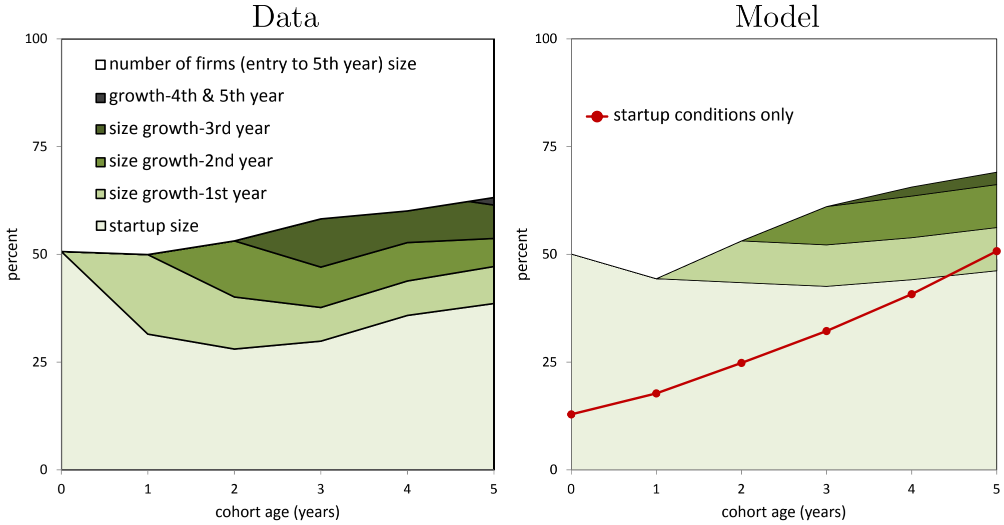
Additional insight into the drivers of cohort-level persistence is obtained by quantifying the contributions of the four aggregate shocks (right panels of Figure 8). The demand shock stands out as the dominant driver of not only cohort-level employment and average size, but also of average size

Figure 8: 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), cohort-level average size (middle row) and individual-firm average size (bottom row).

Figure 9: Contribution of average size to employment variation in model



Notes: contributions of average firm size at different ages to the variation in cohort-level employment as a percentage of its total variance. Data are obtained from the estimated model. The red solid line, "startup conditions only", plots the covariance between cohort-level employment and average firm size obtained by fixing firm-level employment within age/type brackets to its steady-state value, scaled by the total variance of cohort-level employment.

of an individual firm. In other words, it plays a crucial role not only in shaping conditions at entry, but also later in firms' lives. In particular, the shock explains about 90 percent of average size variation of an individual firm at startup and about 60 percent at age 20.

Using our model, we can also shed more light on the variance decomposition presented in Section 2, which quantifies the relative contributions of the intensive and extensive margins to cohort-level employment fluctuations in the data. Specifically, we quantify how much of the observed contribution of the intensive margin is due to changes in startup conditions only. We do so by exploiting that, as a by-product of the estimation procedure, we obtain model-predicted time paths for all model variables. This includes unobservables, such as the entire distribution of firms across age and type bins, and hence enables us to conduct decompositions that cannot be done using data alone.

First, we revisit the variance decomposition of cohort-level employment, as is done in Figure 3, but this time using the model-predicted time paths over the sample period rather than actual data. Figure 9 compares the variance decomposition in the data (left panel) to its counterpart in the model (right panel). Overall, the model-implied decomposition is close to its empirical counterpart, even though it was not directly targeted.[24]

Next, we further decompose the contribution of the intensive margin using the following formula:

$$S_{a,t} = \frac{\sum_i m_{i,a,t} n_{i,a,t}}{\sum_i m_{i,a,t}} = \underbrace{\frac{\sum_i m_{i,a,t} \overline{n}_{i,a}}{\sum_i m_{i,a,t}}}_{\text{startup conditions only}} + \frac{\sum_i m_{i,a,t}(n_{i,a,t} - \overline{n}_{i,a})}{\sum_i m_{i,a,t}},$$

where $S_{a,t}$ is average size of firms of age a in period t, expressed as the weighted average of firm employment levels ($n_{i,a,t}$) across the different firm types i, and where $\overline{n}_{i,a}$ are the associated steady-state values. The component labeled "startup conditions only" represents a time series which isolates fluctuations in average firm size resulting from only startup composition changes, which is achieved by fixing firm-specific employment levels to their steady-state values, conditional on type and age.[25] As before, we then quantify the contribution of such composition changes to cohort-level employment variation by computing the covariance of this time series with cohort-level employment, scaled by the variance of cohort-level employment.

The contribution of changes in startup composition to fluctuations in cohort-level employment is depicted by the solid line in Figure 9 ("startup conditions only"). In the year of entry, only about 13 percent of cohort-level employment fluctuations are due to compositional effects, substantially less than the overall contribution of the intensive margin. The importance of changes in composition, however, grows markedly with age. By the age of five, composition accounts for more than 50 percent of cohort-level employment fluctuations.

[24]In the model, size growth in year 1 covaries slightly negatively with cohort-level employment and it therefore decreases the overall contribution of the intensive margin.

[25]Note that $m_{i,a,t}$ is fully determined in the year of birth, as the exit rate is constant in the model.

5.2 The importance of startup conditions for aggregate fluctuations

We now use the estimated model to better understand how startup decisions affect *aggregate* employment dynamics. First, we isolate aggregate employment fluctuations driven only by cyclical changes in the number of startups in the various firm types, i.e. by aggregate fluctuations in startup conditions. Second, we investigate to what extent the demand shock impacts on aggregate employment dynamics.

To quantify the extent to which employment fluctuations are driven by changes in startup conditions, we again exploit the model-predicted time-varying distribution of firms across types and ages and the associated employment levels. Specifically, we decompose aggregate employment as follows:

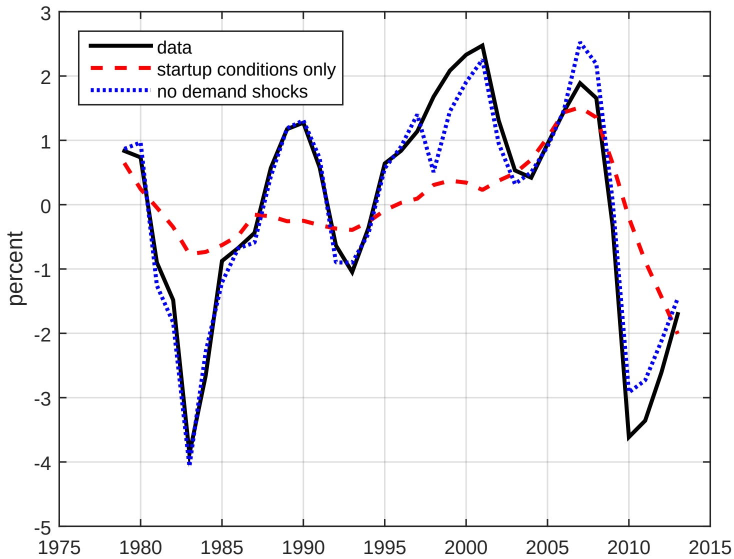
$$N_t = \sum_a \sum_i m_{i,a,t} n_{i,a,t} = \underbrace{\sum_a \sum_i m_{i,a,t} \overline{n}_{i,a}}_{\text{startup conditions only}} + \sum_a \sum_i m_{i,a,t}(n_{i,a,t} - \overline{n}_{i,a}).$$

This formula allows us to construct a time series for a component of aggregate employment which isolates variation purely due to fluctuations in entry into the various type bins. Again, this achieved by setting employment levels, conditional on age and type, to their steady-state values, although this time we aggregate over all firms rather than firms in specific cohorts. We again refer to this time series as "startup conditions only", because entry decisions depend purely on economic conditions in the year of startup.

Figure 10 shows that the contribution of startup conditions to aggregate employment fluctuations is large. Interestingly, the series resembles a slow-moving *trend* in aggregate employment. In fact, the correlation between the component of aggregate employment isolating "startup conditions only" and the HP-trend in aggregate employment is 0.65 for a smoothing coefficient of 100 and 0.73 for a smoothing coefficient of 6.23, the latter following Ravn and Uhlig (2002).[26] Thus, startup decisions appear to be important also for understanding the low-frequency movements of *aggregate* employ-

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

Figure 10: Employment rate: data and estimated contribution of startup conditions



Notes: "startup conditions only" refers to the time series for the employment rate that is constructed by fixing the age/type firm sizes to their respective steady state values. "No demand shocks" is constructed by feeding all the estimated shocks through the model, except for the demand shocks which are set to zero.

ment, often ignored in business cycle analysis.

Finally, given the importance of the demand channel for cohort-level outcomes, we investigate to what extent demand shocks explain dynamics at the aggregate level. Towards this end, we fix the aggregate demand shock to be equal to zero in our model, but leave the remaining estimated three shocks untouched. Figure 10 shows the resulting time path of aggregate employment ("no demand shocks").

Without demand shocks, which are particularly important for changes in the composition of firms with respect to their potential to grow large, the resulting time series for aggregate employment is roughly 10 percent less volatile than the actual employment rate observed in the data. Moreover, in certain periods, demand shocks were particular important. On the one hand, demand shocks, and the associated shift towards high growth potential firms, served to increase aggregate employment by $0.5-1$ percent at the end of the millennium. On the other hand, the opposite happened during and in the aftermath of the Great Recession, where aggregate employment would have fallen by about 1 percentage point less had it not been for the demand shock. Thus, the estimated model predicts that demand shocks have had important effects not only on cohort-level outcomes, but also on the aggregate economy.

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 to the composition of startups with respect to their growth potential. Our results indicate that cohorts of large firms tend to be born during periods of booming consumer demand, when it is relatively easy for firms to acquire new customers. Moreover, the impact of entry decisions not only persists as cohorts mature, but their magnitude increases over time since firms with highly scalable businesses 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.

References

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

Arkolakis, C. (2010): "Market Penetration Costs and the New Customers Margin in International Trade," *Journal of Political Economy*, 118(6), 1151–1199.

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.

Bilbiie, F., F. Ghironi, and M. Melitz (2012): "Endogenous Entry, Product Variety and Business Cycles," *Journal of Political Economy*, 120(2), 304–345.

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., and J. Haltiwanger (2006): "On the Nature of Capital Adjustment Costs," *Review of Economic Studies*, 73(3), 611–633.

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

Dorfman, R., and P. Steiner (1954): "Optimal Advertising and Optimal Quality," *The American Economic Review*, 44(5), 826–836.

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

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*.

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

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

Hall, R. (2014): "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.

Hurst, E., and B. Pugsley (2011): "What Do Small Businesses Do?," NBER working paper no. 17041.

Kaas, L., and P. Kircher (2015): "Efficient Firm Dynamics in a Frictional Labor Market," *American Economic Review*, 105(10), 3030–3060.

Klette, T. J., and S. Kortum (2004): "Innovating Firms and Aggregate Innovation," *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.

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.

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

Pugsley, B., and A. Sahin (2014): "Grown-up Business Cycles," Federal Reserve Bank of New York, Staff Report.

Ravn, M., S. Schmitt-Grohé, and M. Uribe (2006): "Deep Habits," *Review of Economic Studies*, 73(1), 195–218.

Ravn, M., and H. Uhlig (2002): "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations," *Review of Economics and Statistics*, 84(2), 371–380.

Rogerson, R. (1988): "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. (2015): "Lost Generations of Firms and Aggregate Labor Market Dynamics," mimeo.