Recession Scars and the Growth of Newborn Firms in General Equilibrium*

Petr Sedláček† and Vincent Sterk‡

October 4, 2013

Abstract

This paper shows that episodes of subdued firm entry can have persistent negative effects on the aggregate economy through a reduction in the average firm size within incoming cohorts of young firms. Using data from the Business Dynamics Statistics (BDS) we follow cohorts of newborn firms up to the age of five. We find that a cohort's success in creating jobs is largely determined around the time of its birth and that most of the variation in cohort-level employment is driven by the intensive margin (average firm size) rather than the extensive margin (the number of firms). The cyclical pattern that emerges is that cohorts born during recessions consist of firms that are smaller and remain smaller even after aggregate conditions have recovered. To assess the implications of this pattern for aggregate outcomes, we develop and estimate a general equilibrium model designed to speak to the BDS data. Accordingly, we model heterogeneity in returns to scale, endogenous entry and firm-level growth subject to convex adjustment costs. We estimate the underlying aggregate shocks and perform counterfactual simulations, which reveal that the effects of fluctuations in job creation by young firms on aggregate output are substantial and long-lasting despite the presence of general equilibrium forces.

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

*We thank Christian Bayer, Francisco Buera, Fabien Postel-Vinay, Morten Ravn and Kjetil Storesletten as well as seminar participants at UCL, the University of Konstanz and the University of Oxford for helpful comments. This paper is preliminary. All errors are our own.

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

‡University College London, Department of Economics, and Centre For Macroeconomics, e-mail: v.sterk@ucl.ac.uk.

1 Introduction

Following the financial crisis of 2008, the United States and many other countries experienced an unusually deep and prolonged economic downturn, raising concerns about a long-lasting drag on aggregate employment and output. These concerns are fueled by the observation of a particularly large drop in job creation by young firms in the U.S. in recent years. According to data from the Business Dynamics Statistics (BDS), aggregate private employment in the US fell by 7.9 million jobs between March 2006 and March 2010. During the same period, aggregate employment by firms up to the age of five fell by 3.8 million jobs, accounting for nearly half of the decline in aggregate employment.[1]

While the quantitative importance of young firms for average aggregate job creation has been recognized at least since Fort, Haltiwanger, Jarmin, and Miranda (2012), this paper focuses on the role of young firms in aggregate fluctuations. In particular, we investigate whether weak performance of young firms in recessions can have persistent negative effects, possibly far exceeding the horizon of the downturn. Specifically, we follow the cyclical behavior of newborn cohorts using BDS data from 1979-2010.[2] The narrative that arises is that economic downturns produce cohorts of firms that are relatively small and, more importantly, which remain small even after aggregate fundamentals recover. That said, such developments are relevant for aggregate outcomes only to the extent that other firms do not compensate in equilibrium. To this end, we develop a general equilibrium model with heterogeneous firms allowing us to quantify the effects on aggregate outcomes. The results show that, even in general equilibrium, young firm cohorts have substantial and very persistent effects on the aggregate economy.

In the empirical section of this paper we use BDS data, which allows one to follow cohorts of newborn firms up until the age of five, to establish three stylized facts. First, cyclical deviations in the level of employment created by newborn cohorts persist as the cohort ages, sharply contrasting the strong mean-reversion in aggregate employment. Second, at age five, three quarters of the variation in cohort-level employment is driven by the

[1] By contrast, the share of firms up to five year of age in the *level* of aggregate employment is much smaller: 14 percent in 2006 and 12 percent in 2010.

[2] The BDS data are based on administrative records covering nearly all private employers in the United States. Fort, Haltiwanger, Jarmin, and Miranda (2012) and Moscarini and Postel-Vinay (2012) use BDS data to document cyclical patterns in employment by firm size. The latter document that large firms cut back more on employment during times of high aggregate unemployment. Here, we zoom in on newborn cohorts.

intensive margin (average firm size) rather than the extensive margin (number of firms). Third, even after five years recession-born cohorts consist of smaller firms (and create less jobs) on average.

Next, we develop a general equilibrium model with heterogeneous firms, endogenous firm entry, growth subject to convex adjustment costs and aggregate uncertainty. By conducting counterfactual simulations we can investigate the importance of youn firms while preserving general equilibrium effects. The model is designed to speak to BDS data which includes almost the entire cross-section of U.S. firms, and to facilitate this we allow for a set of production technologies which differ in the degrees of returns to scale. Although the model has a large-dimensional aggregate state, giving rise to rich endogenous propagation of shocks, our computational approach nevertheless allows for a fast solution using standard techniques.[3] This computational advantage allows us to bring the model close to the data by estimating parameters as well as the realizations of stochastic shocks with continuous support. We use the latter as inputs for counterfactual model simulations which uncover the role of firm entry in shaping aggregate dynamics, while preserving general equilibrium effects.

The model simulations reveal that sizeable aggregate effects associated with fluctuations in entrant size prevail in general equilibrium. Two main effects are behind this result. First, during an economic downturns the *composition* of startups shifts towards firms with lower returns to scale and hence smaller firms. Second, labor is *reallocated* during recessions as young firms with high returns to scale are held back on their way to reaching their optimal size. Consequently, a smaller fraction of output is produced by firms with high returns to scale persistently reducing average labor productivity and output. Interestingly, the effect on employment is similarly strong upon impact, but much less persistent.

Our model builds on a rich literature studying the dynamics of firms and variations in firm size. Lucas (1978) explains observed variations in firm size using a model with managers who differ in their "spans of control", i.e. the ability in running run large firms. Firm dynamics are introduced in Jovanovic (1982) who develops a model in which new firms grow faster and are more likely to fail compared to older firms as they learn about their efficient scales of operation. A workhorse firm dynamic model is presented

[3]Our model can be solved in one step as opposed to most models with a large number of agents, which are typically solved using the computationally demanding algorithm of Krusell and Smith (1998). Moreover, in contrast to these models we preserve exact aggregation.

in Hopenhayn and Rogerson (1993) who analyze the welfare effects of firing taxes in a general equilibrium model of firm dynamics (without aggregate uncertainty). More recently, Clementi and Palazzo (2010) and Lee and Mukoyama (2012) have extended the Hopenhayn-Rogerson framework to study how entry and exit propagates shocks and apply their models to the manufacturing industry. The empirical focus of the model in our paper is on aggregate data from BDS and we adapt our model accordingly. Moreover, we incorporate multiple sources of aggregate uncertainty and estimate the realizations of the shocks as well as parameters pertaining to their laws of motion.

Firm heterogeneity is commonly modeled exclusively through variations in the degree in Total Factor Productivity (TFP) across firms. This appears a plausible assumption in many applications, which are often focused on specific sectors like manufacturing plants, in which there is arguably little heterogeneity otherwise. Nonetheless, Holmes and Stevens (2012) provide evidence that the usual approach to modeling firm heterogeneity overstates the importance of TFP differences, even within narrowly defined industries, and build a model with large-scale standardized plants and small-scale specialty plants. In this paper, we are primarily interested in aggregate outcomes and apply our model to the entire cross-section of private employers in the economy. Given the vast amount of heterogeneity in the type of activities performed by firms, there is arguably large variation in the degree to which various types of businesses are scalable.[4] For example, an insurance company and a dental practice may have similar degrees of TFP, but the former type of business enjoys much greater benefits from being scaled up than the latter. In the quantitative implementation of our model, we find that a small degree of heterogeneity in returns to scale provides a surprisingly good fit of the firm size distribution in the BDS, conditional on firm age. Moreover, the implied parameter values have reasonable implications for firm profits.[5]

Finally, this paper proposes a novel modeling of the firm entry phase. After paying an entry cost potential entrants can choose the technology type they wish to startup. However a coordination friction prevents all business opportunities from realizing. The probabilities of successfully starting up a business of a certain type endogenously adjust such that potential entrants are indifferent between technology types. As a result, firms

[4] Basu and Fernald (1997) provide evidence in for heterogeneity in returns to scale across sectors.

[5] That said, our framework does allow for simultaneous heterogeneity in TFP and returns to scale. In typical firm dynamics models introducing more than one dimension of heterogeneity would be challenging, since solutions in these models are described by cutoff rules.

with attractive and less attractive technologies enter simultaneously in equilibrium, which enhances the model's ability to speak to the BDS data in which a vast number of small firms is observed, even among very cohorts of old firms.[6] This feature of our model is also consistent with empirical evidence that many starting entrepreneurs have low growth expectations.[7] Furthermore, our model can match the cyclicality of firm entry without relying on exogenous shocks to the entry cost.

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. Concluding remarks are made in Section 6.

2 Empirical evidence

The first step is to investigate whether the somber hypothesis that weak firm entry during a recession holds back aggregate employment in subsequent years is consistent with empirical patterns. Using publicly available data from the Business Dynamic Statistics (BDS), described in Subsection 2.1, enables one to single out cohorts of newborn firms and track their job creation in the year of birth as well as during the five years after. The BDS data also allow for a break down of cohort-level employment into an extensive (the number of firms) and an intensive margin (average employment per firm) and analyze their relative importance. Furthermore, we pay special attention to whether business cycle conditions in the year of a cohort's birth are related to the cohort's job creation performance throughout its life.

The empirical results are presented and discussed in Subsections 2.2 and 2.3. Subsection 2.4 explores several channels through which the characteristics of startup firms may be affected by the business cycle. Finally, Subsection 2.5 presents a first step in assessing whether firm entry effects can be quantitatively important drivers of aggregate employment fluctuations.

[6] In 2007, the fraction of firms with 10 or less employees among firms between 21 and 25 years of age is was about two thirds.

[7] See for example Campbell and De Nardi (2009) and Hurst and Pugsley (2011).

2.1 Data and definitions

The Business Dynamic Statistics (BDS) database covers a very large fraction of US private employment (98 percent), which is an important advantage over alternative data sources, especially given our objective to study implications for aggregate outcomes. We use annual information on the number of firms and their associated job flows broken down into age categories, for the period 1979-2010.[8]. The available age breakdown in the BDS allows one to follow cohorts of new firms for up to five years after they enter the economy. The BDS groups older firms into age categories of $5-9$, $10-15$ and 16 years and older, preventing further following of an individual cohort. Nevertheless, the five year cutoff strikes a reasonable compromise between the length of a cohort and the number of cohorts available.

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 $a=0$.[9] 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 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.[10]

2.2 Job creation by newborn firms over time

Startups are known to be important drivers of job creation. Figure 1 displays the employment levels for startup cohorts ($N_{0,t}$) born between 1979 and 2010. On average, newborn cohorts created 2.96 million jobs, without any clear trend and accounted for about 14 percent of aggregate gross job creation and 156 percent of aggregate net job creation. There

[8] The BDS data start in 1977 but we drop the initial two years following the recommendations of Fort, Haltiwanger, Jarmin, and Miranda (2012) and Moscarini and Postel-Vinay (2012), who cast doubt on the quality of the initial two years of data

[9] A new firm is defined as a firm having a positive employment entry in March of year t, while not having an employment entry in March of $t-1$.

[10] Alternatively, one could use the employment stock data presented in the BDS. These employment numbers do not equal to the sum of net job creation, which is because the net job creation data is cleaned from observed entrants that are not believed to be true startups, whereas the employment data are not cleaned from this noise. 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" (http://www.census.gov/ces/dataproducts/bds). Thus, the net job creation data are superior, at least for our purpose.

is large variation between cohorts, ranging from 2.24 million jobs to 3.54 million jobs.

Figure 1 displays a second snapshot of the cohorts' employment levels, but now five year after birth ($N_{5,t}$).[11] Large variation remains present five years after birth, with employment levels ranging from 1.94 to 3.19 million jobs.[12] Moreover, Figure 1 reveals that cohort-level employment is persistent, as there is a clear positive comovement between the employment levels of cohorts at birth and at their five year anniversaries.

2.2.1 Persistence of cohort-level job creation

The observed persistence of cohort-level job creation becomes especially striking once one makes a comparison with employment at the aggregate level. We correlate employment in year t and year $t + a$, both at the level of an individual cohort born in period t and at the aggregate level. Figure 2 plots the correlation coefficients for $a = 1$ up to $a = 5$.[13] While cohort-level employment at birth and 5 years into existence is highly correlated (the correlation coefficient of 0.64), its aggregate counterpart displays no persistence after a three year horizon. Thus, deviations in job creation by individual startup cohorts persist as the cohorts age to a degree that far exceeds the persistence inherent to the aggregate business cycle.

2.2.2 Intensive versus extensive margin

How important is cohort-level persistence quantitatively and what are the driving factors? The observed variation of employment of individual cohorts can be decomposed along two dimensions. The first is the age dimension, separating out the contributions of initial employment levels and growth rates in subsequent years. Secondly, one can dissect cohort-level employment into an extensive margin, measured by the number of firms, and an intensive margin, measured by the average level of employment per firm within the cohort.

[11] As our data sample ends in 2010, the last cohort we observe is the one born in 2005.

[12] On average, cohorts at this age provide 2.41 million jobs, a decline of on average 0.55 million jobs relative to the year of birth. An important driver behind this average decline is exit by firms in the during the five years after birth: out of the initially observed firms only 54 percent survives the five years after birth.

[13] While cohort-level employment does not display a trend, aggregate employment does and therefore we choose to detrend both time series with a linear trend. Using alternative detrending methods gives similar results. For the aggregate we simply correlate period t employment with employment in years $t, t+1, ..., t+5$.

To quantify the relative importance of the various margins, we decompose the natural logarithm of cohort-level employment as:

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

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

$$\begin{aligned} Var\left(\hat{N}_{a,t}\right) = \; & Cov(\hat{N}_{a,t}, \hat{S}_{0,t-a}) + Cov(\hat{N}_{a,t}, \hat{M}_{0,t-a}) + \sum_{j=1}^{a} Cov(\hat{N}_{a,t}, \hat{\gamma}_{j,t-a+j}) \\ & + \sum_{j=1}^{a} Cov(\hat{N}_{a,t}, \hat{\delta}_{j,t-a+j}), \end{aligned}$$

where a hat indicates deviations from a linear trend of a logged variable.[14]

Figure 3 plots the contributions of average firm size and the number of firms to employment levels of cohorts at age five, expressed as a percentage of the total variance. Average firm size accounts for the lion's share of the variation in employment of five year old cohort: about 75 percent. Within the part accounted for by average size, nearly half is due to entrant size. Moreover, entrant size and firm growth in the year after birth jointly account for over half of the overall variation in employment. Thus, average firm size at the early stage of a cohort's existence emerges as a key determinant of a cohort's success in providing jobs later in life.

2.3 Startups and the aggregate business cycle

2.3.1 Cohort-level employment

The following paragraphs analyze the relation between cohort-level employment and the aggregate business cycle. First, to visualize the link we focus on episodes of particularly strong (weak) job creation by entrants. Second, we conduct a more formal business cycle analysis using the entire sample at hand.

Specifically, we select the five weakest cohorts in terms of initial job creation over the period 1979-2005, as well as the five strongest cohorts over this period.[15] For each age

[14] Detrending with other methods gives similar results.

[15] The weakest cohorts are those born in the twelve months preceding March 1980, 1983, 1984, 1991, and 1993. For the strongest cohorts, the corresponding years are 1987, 1998, 1999, 2002 and 2005.

from zero up to five, we compute the average *cohort-level* employment by age over the two subsets of vintages. We then compare the evolution of cohort-level employment from age zero until five to its *aggregate* counterpart. We linearly detrend both series and for the sake of comparability we scale the deviations from the trend by the standard deviation in the year corresponding to age zero.

The top panel of Figure 4 plots the results. Two patterns emerge. First, there is a positive relation between cohort-level employment and aggregate employment: years of exceptionally weak (strong) aggregate employment are also years of low (high) job creation by firm entrants. Thus, cohort-level employment of entrants indeed appears to behave procyclically. Second and most strikingly, cohort-level employment continues to deviate markedly from the trend in the years after birth, sharply contrasting with aggregate employment which quickly reverts back to the trend after two years and has even crossed the trend line after five years.

To verify the robustness of the patterns suggested by Figure 4, we correlate cohort-level employment over the entire sample, as well as its aggregate counterpart, with measures of the aggregate business cycle in the year of the cohort's birth. We use the employment rate and real GDP as our business cycle measures.[16] We analyze raw data in levels, linearly detrended data, and data detrended using the Christiano and Fitzgerald (1999) filter.[17] The band-pass filter is used to isolate medium-term business cycles with frequencies between 6 and 12 years. By its very nature, firm entry is a forward-looking decision that may respond more to persistent swings in aggregate conditions rather than to short-lived fluctuations.

The correlations are reported in Table 1 and confirm the patterns visible in Figure 4. Cohort-level job creation at birth is highly positively correlated with the business cycle. This relation with the business cycle at birth remains positive as the cohort has grown to age five. Aggregate employment, by contrast, displays highly negative correlations at a five year horizon. Finally, the correlations computed using band-pass filtered data are considerably stronger, implying that the link between cohort-level performance and aggregate conditions at birth is particularly close at medium-run frequencies.

[16]The employment rate is defined as 1 minus the unemployment rate, taken from the BLS. Both series are averages over March-to-March periods, corresponding to the timing of the BDS.

[17]The exception is GDP, which exhibits a strong upward trend and is thus is not suitable for use in levels.

2.3.2 Average size

Given that the variance decomposition revealed a prominent role for average size in accounting for cohort-level employment fluctuations, we analyze the cyclical behavior of the intensive margin individually. The bottom panel of Figure 4 displays average size (employment per firm) within the five weak and strong cohorts identified above. In the five years of weak entry, startup size is on average more than one standard deviation below its mean. Moreover, the deviations from the mean are persistent as the cohort ages and even grows substantially larger during the first two years after birth. Again, the cohort-level pattern differs markedly from its aggregate counterpart, also plotted in Figure 4. After five years, the deviation of average size in the aggregate data is close to the trend level and has even switched sign.

Correlations between average size and business cycle conditions in the year of birth are displayed in Table 1. In accordance with Figure 4, average firm size moves procyclically, both at the level of newborn cohort and at the aggregate level. Moreover, at the cohort level firm size at age five remains positively correlated with the state of the business cycle at birth. In the raw data, the correlation between average size within a cohort of age five and the aggregate employment rate in the year of birth is 0.44. Although the sign of the correlation is robust, its magnitude varies with the detrending method and business cycle indicator used, with the correlations at medium-term frequencies being particularly high. Nonetheless, in all cases there is a stark difference with the behavior of average size in the aggregated data, which switches from a very high contemporaneous correlation to a negative correlation at horizons five years ahead.

2.3.3 Composition changes vs. employment choices

The observed patterns suggest that firm cohorts born in recessions create both less jobs initially and in later years. This can be either due to a shift in the composition of heterogeneous firms towards smaller businesses or due to firms choosing lower employment levels for in a given mix of firm types.

To get a glimpse of these two effects we compute the average size and the firm share of startups within each of the size categories reported in the BDS data. The questions is whether variation in entrant size is predominantly driven by variation in average sizes of the given size categories, or rather due to variation in the composition of firms across these categories.

Figure (5) shows the data on entrant size and two counterfactuals. The first is calculated by fixing the composition of firms and letting only average firm size within each size category to vary. The second counterfactual does the opposite by fixing average firm size and letting the share of firms across the size categories to vary. The figure suggests that the main driver of entrant size variation are changes in the composition of firms.[18]

2.4 Summary and possible explanations

We can condense our results into four stylized facts applying to cohorts of young firms:

Fact 1. *Cohort-level employment is largely determined in the year of birth.*

Fact 2. *The intensive margin (average firm size) is the main driver of variations in cohort-level employment.*

Fact 3. *Cohorts of small firms are born in times of low economic activity.*

These stylized facts are difficult to reconcile with the view that cohort-level employment at a given point in time is primarily driven by the *current* state of the business cycle. Instead, the observed patterns lend support to a view in which firm characteristics *at the entry stage* are important in determining a cohort's potency to create jobs, both initially and later in its life. Moreover, our results show that job creation by recession-born cohorts is primarily weak because these cohorts consist of smaller firms on average. While a subsequent economic recovery may contribute to average size growth at the cohort level, the upward effect appears by no means large enough to offset the lower average size levels that recession-born cohorts are born with.

Our results also indicates that the composition of firm entrants varies over the business cycle, suggesting that recessions are especially bad times for large firms to startup. Inspecting the BDS data reveals a tremendous amount of size heterogeneity and it therefore appears unlikely that the composition of entrants is constant over the business cycle.[19]

[18]The figure could be overstating the effect of composition changes if firms' choices were so volatile that they fall into different size brackets. However, redoing the counterfactuals with wider size brackets yields similar results.

[19]Large heterogeneity is present even among older firms. According to the BDS data, 40 percent of firms that were older than 25 years in 2005 had less than five employees.

However, one can think of several plausible explanations for why the composition of entrants may fluctuate over the business cycle. In this subsection we discuss several candidate explanations for our stylized facts.

The first possibility we consider is that during recessions average firm size within newborn cohorts declines because of reallocations of activity between sectors. For example, if average firm size in manufacturing is relatively large and if activity in this sector declines relatively strongly during a recession, then a composition shift away from manufacturing may create a decline in average entrant size. We conduct a simple exercise, using the fact that the BDS data can be decomposed to the level of nine sectors.[20] We compute a counterfactual time series for the average size within cohorts of entrants under the assumption that the distribution of the number of entrants over the nine sectors remains fixed over time, setting the fractions equal to their sample averages. This series captures variation that is due within sector variations in average size only. Similarly, we compute a counterfactual series that captures only between-sector shifts, by setting the average entrant size within each sector equal to the sample average, but fractions of entrants in the nine sectors to vary over time as in the data. The top panel of Figure 6 displays the two counterfactual time series, as well as the actual series for average size within newborn cohorts. It is immediately clear that within-sector variations account for almost all of the variation in average size; between-sector shifts appear to play an extremely limited role.[21]

Secondly, we investigate the importance of very small firms in driving our results. Several studies emphasize the role of entrepreneurship as a way to escape unemployment ("necessity entrepreneurs"), e.g. Hurst and Pugsley (2011) and Poschke (2012). Businesses created out of a necessity motive are likely to remain very small. Given that unemployment is high during recessions, one may expect necessity entrepreneurship to have a negative effect on the average size of firms born in recessions. We again carry out a simple exercise this time exploiting the fact that the BDS provide data decomposed into size categories. We construct a counterfactual time series for the total employment of cohorts at age five, but set the employment levels of firms with less than 10 employees equal to the sample average. Similarly, we construct a second counterfactual time series by setting the employment of firms with 10 employees or more equal to the sample average. The

[20] These are: (i) Agriculture, Forestry, and Fishing, (ii) Mining, (iii) Construction, (iv) Manufacturing, (v) Transportation, Communication, and Public Utilities (vi) Wholesale Trade, (vii) Retail Trade, (viii) Finance, Insurance, and Real Estate, (iv) Services.

[21] The Appendix provides further details on our findings within sectors.

bottom panel of Figure 5 shows that the vast majority of fluctuations in employment of five year old firms is in fact driven by firms with 10 or more employees.[22,23] Of course, this observation does not refute the existence of necessity entrepreneurship, but it appears unlikely that cyclical variations in this entrepreneurship motive are an important driver the documented fluctuations in cohort-level employment.

A third explanation for the observed patterns is that certain businesses are simply more suitable to be scaled than others. Recessions may be relatively unattractive times to start up scalable businesses, for example because achieving high firm growth is difficult during a recession period.[24] We evaluate this possibility using a structural model, to be presented in Section 3.

2.5 Aggregate implications

Are variations in firm size within young cohorts large enough to have substantial effects on *aggregate* employment? To answer this question, we compute time series for aggregate employment under two counterfactual assumptions. First, we assume that the numbers of firms aged 0 to 5 years are fixed at their sample averages, but average firm sizes are as in the data. Contrasting this counterfactual series to the actual one reveals the effect of the "extensive margin" on aggregate employment, as all differences are due to the number of young firms. Second, we assume that both the number and the average size of firms aged 0 to 5 years are fixed at their sample averages, capturing the effects of both the extensive and the intensive margins on aggregate employment.

Figure 6 depicts the *differentials* between aggregate employment and the two counterfactual measures. A negative value means that actual employment was lower than what would be predicted if firm entry (and average firm sizes) were at their average values. The figure clearly shows pro-cyclical movements in both series as employment drops in recessions due to lower than average firm entry (and lower average firm sizes).[25]

[22] Alternative size cutoffs give similar results.

[23] As a second check, we recompute the correlation between the average size of five year old firms and the employment rate at birth, but excluding firms in the smallest size categories. If the observed cyclicality of average size would be driven by necessity entrepreneurs creating only small firms during recessions, the positive correlation between average firm size and the business cycle conditions at birth is likely to disappear. However, it turns out that correlation coefficient actually *increases* from 0.44 to 0.66.

[24] Alternatively, selection effects may occur because some potential startups are more productive than others. We consider this possibility as well. Note, however, that small firms in the BDS data are not necessarily unproductive, as they for example include lawyers and notaries.

[25] The decline in the differential due to only the extensive margin may seem surprisingly low in the latest

We can draw three conclusions from this exercise. First, variations in job-creation by young firm cohorts can have large implications for aggregate employment. Even the effect of the extensive margin alone can be as large as 2% of aggregate employment.

Second, the effect of the intensive margin is strong. This is particularly visible during the latest recession and the one in the early 80's. In both episodes, lower average firm size among young firms chipped off an additional 2% of aggregate employment. Similarly, the boom years around the turn of the millennium which were characterized by firm entrants with high job creation potential boosting aggregate employment.

Third, the intensive margin holds back aggregate employment even after the number of entrants picks up. This is most apparent in the late 80's where while the number of entrants started to recover, the low average size within young cohorts was still pulling aggregate employment down.

An important caveat in these exercises is that it ignores any responses by older firms, which in equilibrium possibly fill the gaps in job creation left by young cohorts. This is precisely one of the motivating factors to build a structural general equilibrium model within which it is possible to address this issue. Such a model is described in the next section.

3 The model

This section presents our model. The model economy is populated by a representative household and a continuum of heterogeneous firms. Firms and households trade on a goods market and a labor market, both of which are perfectly competitive. [26]

The following subsections describe the decision problems of the agents in turn and define the equilibrium. We discuss the quantitative implementation of the model Section (4). Results are presented in Section (5).

---

recession. Firm entry was indeed very weak in the last few years of the sample (roughly 20% below average). This is, however, largely compensated for by the particularly strong cohorts prior to the recession.

[26] Firm dynamics models with more detailed descriptions of the labor market include Acemoglu and Hawkins (2010), Elsby and Michaels (2010), Kaas and Kircher (2011), Moscarini and Postel-Vinay (2010), Schaal (2010) and Sedláček (2011) who extend the Mortensen-Pissarides model to include multiworker firms.

3.1 Firms

An endogenous measure of heterogeneous firms operates in the economy, producing a homogeneous good. Firms are owned by a representative household and use the household's stochastic discount factor to compute the expected present value of future profit flows. Firms exit the economy with an exogenous probability ρ_a, which we allow to depend on the firm's age.

A common feature of all firms is that they use only labor as an input in production, but they may differ in their precise production technologies. There is a finite number of production technology types, characterized by a certain level of technology-specific total factor productivity and a certain degree in returns to scale. Let the technology types be indexed by $i = 1, 2, .., I$. Technology type i is associated with the production function

$$y(n_t; i) = z_i A_t n_t^{\alpha_i},$$

where n_t is the firm's level of employment, $z_i > 0$ is a technology-specific total factor productivity parameter, A_t is an exogenous and stochastic aggregate TFP variable with mean one, and α_i is a technology-specific returns to scale parameter.[27] In our quantitative simulations $\alpha_i \in (0, 1)$ 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. In the presence of technological heterogeneity, some firms grow up to become very large, whereas others remain small.

Firm growth is limited by a quadratic adjustment costs on the change in employment, given by $\frac{\zeta_a Q_t}{2}(n_t - n_{t-1})^2$, where $\zeta_a > 0$ is a parameter which we allow to depend on the firm's age a.[28] Due to the adjustment cost, newborn firms grow to their optimal size gradually over time, in line with the positive relation between firm age and size present in the BDS data. Q_t is a stochastic shock with mean one which shifts the level of the adjustment cost. A rise in Q_t increases the cost of firm expansion and we therefore label Q_t an " expansion cost shock".[29] Given that firm expansion is a form of investment in our model, Q_t resembles the investment-specific technology shock that features prominently in the DSGE literature and is sometimes thought of as a stand-in for time-varying financial

[27] For computational feasibility we assume firm-level employment is a continuous variable.

[28] The initial employment levels for newborn firms are treated as parameters. The calibration is detailed in Section 4. In the period of exit no adjustment cost needs to be paid.

[29] A rise in Q_t also increases the cost of firm contraction, but this is less relevant in our context given that firms in our model are typically on upward-sloping growth paths

frictions. Alternative interpretations are possible, however, and we come back to this issue in the Conclusion.

Firms optimally choose their levels of employment in any period, taking as given the evolution of the aggregate state $\mathcal{F}_t$ which pins down the wage W_t. Denoting the stochastic discount factor between period t and $t+1$ by $\Lambda_{t,t+1}$, we can express the firm's maximization problem recursively as

$$V_{i,a}(n_{i,a-1,t-1}, \mathcal{F}_t) = \max_{n_{i,a,t}} \left[\begin{array}{c} y_{i,a,t} - W_t n_{i,a,t} - \frac{\zeta_a Q_t}{2}(n_{i,a,t} - n_{i,a-1,t-1})^2 \\ + (1-\rho_a)\, \mathbb{E}_t \Lambda_{t,t+1} V_{i,a+1}\left(n_{i,a,t}, \mathcal{F}_{t+1}\right) \end{array} \right],$$

where $V_{i,a}(n_{i,a-1,t-1}, \mathcal{F}_t)$ is the value of a firm of type i and age a, $\mathbb{E}_t$ is the conditional expectations operator. By symmetry, all firms of the same age and technology type make the same decisions and we therefore label firms only by technology type and age. The first-order necessary condition for the firm's optimal choice of labor can be written as

$$W_t + \zeta_a Q_t (n_{i,a,t} - n_{i,a-1,t-1}) = \alpha_i z_i A_t n_{i,a,t}^{\alpha_i - 1} + \beta \Lambda_{t,t+1}(1-\rho_a)\zeta_{a+1} Q_{t+1} \mathbb{E}_t (n_{i,a+1,t+1} - n_{i,a,t}).$$

This condition equates the marginal costs of firm expansion to the marginal benefits. Marginal costs consist of the wage and the marginal adjustment cost. Marginal benefits equal the sum of the marginal product of labor and the expected discounted marginal reduction in adjustment costs to be paid next period.

3.2 Entry decisions

Each period, there is an endogenous number of firm start-up attempts, requiring the sacrifice of a cost $\chi > 0$ per attempt. After paying this cost, a potential entrant chooses one business idea from a given measure of opportunities, denoted by $\Psi > 0$, which is known by all agents. Each individual business opportunity is associated with only one of the technology types, but there are multiple opportunities per technology type. Let $\psi_i > 0$ denote the measure of business ideas of type i. For simplicity, we assume that the business opportunities are renewed in each period, so ψ_i is constant. This implies that the measure of total business opportunities, $\Psi = \sum_{i=1}^{I} \psi_i$, is time-invariant.

Startups attempts are subject to a *coordination friction* in choosing business opportunities. This friction gives rise to an aggregate matching function between potential startups and business opportunities. The underlying idea is that without perfect coordination, some potential entrants select the same business opportunity, forcing all but one to exit directly. At the same time, some business opportunities are not selected. Our modeling approach follows the matching models pioneered by Diamond, Mortensen and Pissarides, which are routinely applied to the labor market.

Let $P_{i,t}$ be the probability of successfully starting up a firm, conditional upon paying the entry cost χ and choosing a business opportunity of technology type i. Free entry implies the following condition for each technology type

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

where $V_{i,0,t}\left(0, \mathcal{F}_t\right)$ is the value of a newborn firm of type i. The above equation makes clear that if the value of a new firm of type i increases, the probability of a successful startup, $P_{i,t}$, adjusts downward to restore equilibrium. Note that there is no entry when $V_{i,0,t}\left(0, \mathcal{F}_t\right) < \chi$. Let $x_{i,t}$ denote the measure of potential entrants selecting a business opportunity of type i. The total number of successful startups of type i is determined by a Cobb-Douglas matching function

$$m_{i,0,t} = x_{i,t}^{\phi} \psi_i^{1-\phi}, \text{ for } i = 1, 2, .., I, \tag{2}$$

where $m_{i,a,t}$ denotes the measure of firms of technology type i and age a. The startup success probability satisfies $P_{i,t} = m_{i,0,t}/x_{i,t}$. Hence, the value of newborn firms within each technology type directly pins down the number of startups of each type. To avoid any a priori heterogeneity in entry sensitivities, we assume that ϕ is homogeneous across technology types.

Relative to the firm dynamics literature along the lines of Hopenhayn and Rogerson (1993), our modeling of firm entry has the advantage that it naturally accommodates the presence of multiple dimensions of heterogeneity among firms. In addition to this technical advantage, we believe our approach is conceptually appealing for several reasons. First, agents have a *choice* of what type of firm to start up, rather than being exogenously confronted with a particular technology. Second, agents who aim to start up more ambitious firm types - associated with larger firm values - face tougher competition in starting up their business. Third, technologies of successful entrants are no longer strictly superior to those of failed attempts. This implication is attractive in the light of the empirical evidence that many entrepreneurs have low growth ambitions and are not very skilled ((Hurst and Pugsley, 2011) and Campbell and De Nardi (2009)), while at the same time a substantial fraction of highly skilled and experienced entrepreneurs fail to get a new business ambition off the ground. Finally, as will become clear in the calibration, our framework allows the model to match the volatility of firm entry in the data without introducing an extra shock.

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. Consumption and labor duties are shared equally among the household members. The household's utility function is given by:

$$U\left(C_t, N_t\right) = \frac{C_t^{1-\sigma}}{1-\sigma} - \frac{\nu Z_t N_t^{1+\kappa}}{1+\kappa}.$$

where C_t is the total amount of goods purchased by the household, N_t denotes total employment within the household, σ is the coefficient of risk aversion, κ is the Frisch elasticity of labor supply, ν is a parameter capturing the disutility of labor and Z_t is a stochastic preference shock. The household maximizes the expected present value of life-time utility, subject to its budget constraint:

$$\begin{aligned}
&\max_{\{C_t,N_t\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\sigma}}{1-\sigma} - \nu Z_t \frac{N_t^{1+\kappa}}{1+\kappa} \\
&s.t. \\
&C_t = W_t N_t + \Pi_t
\end{aligned}$$

where Π_t denotes firm profits and $\beta \in (0,1)$ is the household's subjective discount factor. Profits and the wage are taken as given by the household. The optimal employment choice takes on the familiar form:

$$W_t = -Z_t \frac{\nu N_t^{\kappa}}{C_t^{\sigma}}. \tag{3}$$

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 captures time-varying labor market frictions.

3.3.1 Shock processes

The three shocks are each assumed to follow an AR(1) process in logs:

$$\ln J_t = \rho_J \ln J_{t-1} + \epsilon_t^J, \tag{4}$$

where $J = A, Z, Q$ and where ϵ_t^J are i.i.d. innovations distributed normally with mean zero and standard deviations σ_J. ρ_J are the respective persistence parameters.

3.4 Equilibrium

For reasons to be explained in Section (4) we impose a maximum firm age K, that is, we set the exit probability at age K, ρ_K, equal to one. 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 of each technology type i can be written, respectively, as:

$$\sum_{i=1}^{I}\sum_{a=0}^{K} m_{i,j,t}\left(y_{i,a,t} - \frac{\zeta_a}{2}(n_{i,a,t} - n_{i,a-1,t-1})^2\right) - \sum_{i=1}^{I} x_{i,t}\chi = C_t, \tag{5}$$

$$\sum_{i=1}^{I}\sum_{a=0}^{K} m_{i,a,t}n_{i,a,t} = N_t \tag{6}$$

$$m_{i,a,t} = (1-\rho_{a-1})\, m_{i,a-1,t-1}. \tag{7}$$

Let $\mathcal{F}_t = \{m_{i,a-1,t-1}, n_{i,a-1,t-1}\}_{i=1,..I,\ a=0,..,K-1}$ be the aggregate state, consisting of the measure of firms of each age-technology combination up to age $K-1$, the employment levels these firms in the previous period, as well as the values of the stochastic variables. We are now ready to define a recursive equilibrium.

Definition.
A recursive competitive equilibrium is defined by laws of motion for:
- the representative household's labor supply, $N(\mathcal{F}_t)$, and consumption $C(\mathcal{F}_t)$,
- the wage $W(\mathcal{F}_t)$,
- firm-type level value functions $V_{i,j}(n_{i,a-1,t-1}, \mathcal{F}_t)$ and *employment choices $n_{i,a}(n_{i,a-1,t-1}, \mathcal{F}_t)$, for $i = 1, 2, ..I$ and $a = 0, 1, ..K$,*
- the measure of potential entrants $x(\mathcal{F}_t)$and startup probabilities $P_i(\mathcal{F}_t)$ for $i = 1, ..., I$,
- the measure of operating firms of type i and age a, $m_{i,a}(\mathcal{F}_t)$,
that solve the household's problem, solve the firm's problem, satisfy the free entry condition for each technology type $i = 1, ..., I$, satisfy the aggregate resource constraint, clear the labor market, and obey the laws of motion for the elements of the aggregate state $\mathcal{F}_t$.

4 Quantitative Implementation

We parameterize the model using a hybrid method of matching long-run targets, the Simulated Method of Moments (SMM) and Maximum Likelihood (ML) estimation. We solve the dynamic model using a first-order perturbation method around the stationary equilibrium without aggregate shocks. The state variables of the model include the employment levels and the measure of firms for each age/technology type. Given that firms maximally become 50 years and that there are nine technology types, we have more than 900 state variables in our model. In addition, our model features aggregate shocks with continuous support. Nonetheless, we can solve the model in several minutes on a desktop thanks to the use of perturbation methods. Note that the solution procedure does not rely on an approximation of the aggregate state as in Krusell and Smith (1998). The relatively fast computation of the equilibrium makes it possible to estimate parameters, which requires the model to be solved many times.

To facilitate the exposition, we first discuss calibrated parameters used to match long-run targets. Next, we discuss the estimated parameters. Parameters are summarized in Table (4). The model period is one year, corresponding to the frequency of BDS data.

4.1 Parameters calibrated to match long-run targets

We divide the parameters calibrated to match long-run targets into three groups. First, parameters pertaining to the household, second firm-level parameters that are common to all firm types, and third parameters that are specific to technology types.

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 households' coefficient of relative risk aversion, σ, is set to one which implies log utility with respect to consumption. The Frisch elasticity of labor supply, κ, is set to one as well which implies a unit Frisch elasticity of labor supply. The preference parameter ν is chosen to normalize steady-state labor supply to one.

4.1.2 Parameters common to all firm types

Parameters that are common across all technology types are the adjustment cost (ζ_a), the exogenous firm exit rate (ρ_a), the entry cost (χ), the elasticity of startups with respect to firm value (ϕ) and the mass of potential entrants (Ψ). We assume that $\zeta_a = \zeta$ for $a > 0$ such that the model matches the average firm size of 1 year old firms, which is equal to 7.6 in the BDS data. The adjustment cost parameters for startups (a=0) are estimated and discussed in the next subsection. To capture age-dependency of exit rates observed in the data, we introduce the following parametric relation between age and the exit probability

$$\rho_a = \xi_0 + \frac{\xi_1}{a}, \ \varphi_0, \varphi_1 > 0, \ a < K,$$

where a is the firm's age and K is a maximum age. We impose a maximum age for computational reasons. As a result, our model has a finite number of state variables, allowing for the use of perturbation methods to solve for the dynamic equilibrium. The top row of Table (3) contains average exit rates for firms aged one to five, as well as the average exit rate for older firms. We target these numbers, setting $\xi_0 = 0.05$ and $\xi_1 = 0.17$. The bottom row of Table (3) shows that the implied exit rates match their data equivalents very closely. We impose a maximum firm age of a fifty years, setting $\rho_{50} = 1$.

The final three 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. Finally, 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.

4.1.3 Firm-type parameters

Model parameters that describe firm technology types are the returns to scale parameters and firm-specific TFP levels (α_i, z_i), together with the measure of business opportunities in each technology type (ψ_i). 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. This gives us 9 technology types. Our benchmark specification includes heterogeneity only in returns to scale, setting $z_i = 1$ for all technology types i.[30]

We exclude production functions with increasing returns to scale, that is, $\alpha(i) \leq 1$ for $i = 1, .., I$. We normalize one of the technology types (without loss of generality let it be type $i = 1$) such that the model matches a profit rate of 3% taken from Hornstein, Krusell, and Violante (2005). To pin down the remaining 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

[30] The Appendix provides results for cases when there is also heterogeneity in firm-specific TFP levels.

the returns to scale parameters are shown in the bottom part of Table (2). They range between 0.9305 and 0.9993, which is within the range of estimates of Basu and Fernald (1997).

To pin down the measure of business opportunities in each technology type (ψ_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 returns to scale parameters and the firm shares are reported in Table (2).

4.2 Estimated parameters

The remaining parameters are estimated using either the Simulated Method of Moments (SMM) or Maximum Likelihood (ML) estimation. The reason we use both methods is that certain parameters are closely associated with key second moments that seem important for our model to match. For other parameters, in particular the shock process parameters, we have no obvious associated moments and a likelihood approach seems more appropriate. An important by-product of the maximum likelihood estimation is that we obtain estimated time series for the stochastic shocks, which we will later use in counterfactual exercises.

Our estimation strategy has the following steps. First we guess values for the parameters estimated using SMM. Then we estimate the remaining parameters using ML and simulate the model. Next, we evaluate the second moments of interest over the sample period and go back to the first step, updating the guess in the first step. We repeat this procedure until we match the second moments with a reasonable degree of accuracy.

4.2.1 Parameters estimated using the Simulated Method of Moments

We match two key second moments. One second moment we require our model to match is the relative volatility of (the log of) entry with respect to (the log of) GDP, which is 2.68 in the data. Closely associated with this moment is the elasticity parameter in the matching function for entrants, ϕ.

The other key second moment is the volatility of the average size within five year old cohorts, relative to average size volatility among cohorts in the year of entry. To match this moment, we use the adjustment cost parameter for the initial year (ζ_0) and employment endowments before startup ($n_{i,-1}$), subject to restrictions imposed by our calibration described in the previous subsection.

A natural lower bound for the employment endowments is $n_{i,-1} = 0$ for each type i. Upper bounds are obtained by exploiting the fact that given the size distribution between age 16 and 20 and parameter values used to match long-run targets, we can solve backward for the size of each type in the first year of existence, denoted by $n^*_{i,0}$ for type i. [31] Next, we introduce and auxiliary scaling's parameter, $\theta \in [0, 1]$ and set endowments equal to $n_{i,-1} = \theta n^*_{i,0}$. That is, endowments are at the lower (upper) bound if $\theta = 0$ ($\theta = 1$). It is this parameter θ that we estimate. Given θ, the employment endowments follow directly and we can solve for the adjustment cost parameter ζ_0 by matching the average size of entrants in the steady state to average entrant size in the BDS over the period 2000-2010.

It turns out that θ has a strong effect on both the strength of the composition effects and the second moment we target, the volatility of the average size within five year old cohorts. We show this in subsection 5.4 and discuss the underlying reasons and argue that the parameter is well-identified by our SMM procedure. Nonetheless, one can construct bounds on the results by considering the extreme cases $\theta = 0$ and $\theta = 1$.

4.2.2 Parameters estimated using Maximum Likelihood

The shock process parameters are estimated using Maximum Likelihood. In particular, we estimate the persistence parameters ρ_J and volatility parameters σ_J, both with $J = A, Z, Q$. We use three data series for this purpose: aggregate GDP, the aggregate employment rate, computed as one minus the unemployment rate, and the average size of entrants computed from BDS data. To be consistent with the timing in the BDS, we construct annual time series for GDP and employment over March-to-March time intervals.

The maximum likelihood procedure provides us with estimated realizations of the shocks over the sample, as well as an estimate of the initial state, estimated using the Kalman Filter. When we feed these estimates to the model, the model reproduces precisely the time series used in the estimation: aggregate output, aggregate employment and average entrant size. To investigate the importance of the three shocks for the observed fluctuations, one can conduct counterfactuals in which one or more of the shocks is shut off.

Figure 10 plots three model simulations, starting from the estimated initial state. The black line shows a simulation in which all three shocks are fed to the model. By

[31] To see why these are reasonable upper bounds, note that if employment endowments would exceed these levels, firms would shrink in the initial year which is incompatible with a nonnegative adjustment cost.

construction, this simulation reproduces observed time series. The red line represents a simulation in which only the TFP shock is fed to the model and makes clear that this shock is important for fluctuations in output, but not for fluctuations in employment and entrant size. The blue line represents a simulation with both the TFP and the labor wedge, but without the expansion cost shock. A comparison between the blue and the red line reveals that the labor wedge is important for employment, but not for entrant size. A comparison between the blue and the black line reveals that fluctuations in entrant size are largely driven by the expansion cost shock. Moreover, the expansion cost shock also has substantial effects on aggregate output and employment during various episodes.

5 Model results

This section presents the model results. We first discuss properties of the steady-state equilibrium and the present the main results, investigating the importance of changes in entrant composition for aggregate fluctuations.

5.1 Properties of the steady-state equilibrium

Figure 9 plots the steady-state employment patterns of firms by age and technology type. Firms of the lowest returns-to-scale type ($\alpha = 0.9305$) are born small and stay small during their entire life, starting off 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 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 more in the years after birth, they account for an increasingly large share of the cohort's total employment. The shares of firm types in total cohort-level employment are displayed by age in Figure 9. While the most scalable firms account for only about 7 percent of the cohort's total employment in the year of birth, they provide more than half of the cohort's employment by the age of fifty. Firms with low returns to scale, on the other hand, are relatively important during the early years of a cohort's life, with the firms of the three lowest returns-to-scale types creating more than 50 percent of the cohort's jobs in the year of birth. By the age of fifty, however, their share in the cohort's employment level has declined to about 16 percent.

The steady-state patterns suggest that large employment effects may grow out of initially small deviations in entrant size due to composition. Consider for example a counterfactual in which, ceteris paribus, firms of the largest returns-to-scale type do not enter. In this counterfactual entrant size is 7 percent lower, but as the cohort ages this difference in average size increases, growing to a gap of 50 percent after fifty years. [32]

5.2 Entrant composition effects and aggregate fluctuations

We now investigate the importance of entrant composition effects for aggregate fluctuations. We do so using a counterfactual exercise in which we simulate the model using the estimated shocks, but we fix the composition of startups to its steady state. Comparing this simulation to the benchmark model reveals the importance of cyclical changes in entrant composition, while fully allowing for general responses. One can compute the difference between paths for aggregate output and employment predicted by the benchmark and the counterfactual model over the sample period.[33]. The results reveal that the effects of fluctuations in entrant composition on output are substantial with a maximum absolute effect of 0.28 percentage points. The effects of fluctuations in entrant composition on the employment rate are somewhat smaller, but still substantial with a maximum of 0.22 percentage points. Interestingly, we find entrant composition not to be particularly important during the recent Great Recession.

5.3 Recession scars?

To investigate the persistence of the aggregate effects associated with fluctuations in the size of young firms, let us inspect the model's impulse response functions (IRFs). As shown in the previous subsection, the effects of the total factor productivity and labor shocks on startup size are negligible leaving the expansion cost shock to be the main driver of the composition of entrants. Figure (11) plots IRFs of various variables to a one-time expansion cost shock. The size of the shock is equal to its value at the onset of the recent recession in 2007. The other two shocks are fixed at zero. The figure plots the response of the benchmark model and that of an economy in which shifts in the composition of startups are ruled out.

[32] These percentages are very close to the contributions of the largest firm size in cohort-level employment at age zero and fifty. The reason is that high returns-to-scale types comprise only a very small fraction of the total number of firms (0.175 percent) and hence the number of firms remains largely unaffected in the counterfactual.

[33] By construction, the time series for output and the employment rate predicted by the benchmark model are equal to the data

After an increase in the cost of firm expansion, the *composition* of entrants shifts towards those with smaller returns to scale (top left). The reason why this happens is that firms with higher returns to scale respond more strongly to changes in their costs because of a lower profit margin. This effect is analogous to the "labor leverage" effect discussed in Gourio (2007).

The impact on output (top right) is substantial as all firms (not only entrants) are affected by the shock. However, the effect of entrant composition alone is also sizeable. The difference between the benchmark case and the scenario in which composition shifts are switched off is about 0.1 percentage points.

Furthermore, the economy displays a large amount of propagation. The figure displays this by plotting also the IRF of the expansion cost shock, appropriately scaled. While the shock itself almost dies out after 10 years, output remains further below its steady state and more so when startup composition varies. This is due to a persistent drag on labor productivity. This is due to the effect the cost shock has on existing firms and their growth.

An increase in expansion costs labor leads to a shift of labor away from firms with high returns to scale. This is again because firms with higher returns to scale are more sensitive to cost changes due to their lower profit margins and are therefore held back more in their growth towards their optimal sizes. This leads to a *reallocation* of labor as the employment share of high-returns to scale firms within a given cohort declines which in turn lowers labor productivity (bottom left).

The above effects are very persistent for output and labor productivity (bottom left), but not so for employment (bottom right). Therefore, composition changes among entrants remain to have sizeable aggregate effects even in general equilibrium through their impact on the allocation of labor across heterogeneous firms and the implied effects on labor productivity. These effects are not relevant for entrants alone, but apply to all firms. Reallocation of labor among existing businesses then persistently affects aggregate output and labor productivity.

5.4 Alternative parameterizations

We now compare the benchmark model to two extreme alternative parameterizations for the initial employment endowment, $\theta = 0$ (lower bound) and $\theta = 1$ (upper bound). Table (5) presents the mean and maximum difference between data series for employment and output and simulated series in the counterfactual model with a fixed composition of entrants with respect to technology types. Under each of the three parameterizations, shocks fed to the counterfactual model are estimated by imposing that the corresponding exactly reproduces observed data series for aggregate output, aggregate employment and average entrant size.

The lower and upper bounds indicate that the maximum entrant composition effects ranges on average from 0.06 ($\theta = 0$) to 1.74 ($\theta = 1$) percent of aggregate output. For employment we obtain a similarly range around the effect in our benchmark model. To understand why setting $\theta = 1$ gives much larger effects, note that this parameterizations implies that the choice employment choices of firms in their initial years are equal to their employment endowment levels. In other words, individual firms do not expand at all in the initial year under this parametrization and it follows that the adjustment cost in the initial year, ζ_0, approaches infinity. Consequently, *all* observed fluctuations in entrant size are purely due to composition effects when $\theta = 1$.

Table (5) also shows that the value of $\theta = 1$ has a strong effect on the volatility of the size within five year old cohorts, relative to the volatility of entrants. Under the upper bound ($\theta = 1$), the model over-predicts these volatilities seven-fold, while under the lower bound ($\theta = 0$) calibration the model under-predicts them roughly 3-4 times. Thus, the parameter θ and thus the strength of the composition effect are powerfully identified by our specific choice of second moments used in the SMM estimation.

6 Conclusion

This paper presented new facts about the job creation patterns of newborn cohorts and their link to the state of the business cycle around the time of their birth. We develop a general equilibrium model of firm dynamics with aggregate uncertainty, endogenous entry, convex adjustment costs and firms which differ in their returns to scale.

Counterfactual simulations reveal that a major driver of composition effects are shocks to adjustment costs of firms. Heterogeneity in returns to scale results in different sensitivity of firms to aggregate shocks. After an unfavorable cost shock fewer firms with high returns to scale start up and labor gets allocated away from naturally large businesses resulting in a persistent drop in labor productivity and output. Therefore, the effects of a deep may be resonating in the years following the recovery.

We estimated reduced-form shocks rather than modeling detailed frictions, allowing us to avoid strong assumptions about the detailed drives of macroeconomic fluctuations. Our results, however are informative about what frictions could be key drivers behind the observed fluctuations. In particular, they indicate that frictions that manifest them selves as shocks to the cost of firm expansion are of key importance for job creation by newborn firms. A deeper friction underlying this cost is possibly a financial friction limiting the expansion of firms, as in for example Bassetto, Cagetti, and De Nardi (2012). Alternatively, the underlying frictions could be a cost of building a customer base, as in Foster, Haltiwanger, and Syverson (2012). In both cases, the friction can be expected to become more severe during recessions, in line with our observed patterns. We leave it for future research to investigate the detailed implications of various frictions mapping into the estimated reduced-form shocks.

References

Acemoglu, D., and W. Hawkins (2010): "Wages and Employment Persistence with Multi-worker Firms," mimeo.

Bassetto, M., M. Cagetti, and C. De Nardi (2012): "Credit Crunches and Credit Allocation in a Model of Entrepreneurshipa," working paper.

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

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 (2010): "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations," mimeo.

Elsby, M., and R. Michaels (2010): "Marginal Jobs, Heterogeneous Firms and Unemployment Flows," mimeo.

Fort, T., J. Haltiwanger, R. Jarmin, and J. Miranda (2012): "How Firms Respond to Business Cycles: The Role of Firm Age and Firm Size," mimeo.

Foster, L., J. Haltiwanger, and C. Syverson (2012): "The Slow Growth of New Plants: Learning about Demand?," NBER working paper no.17853.

Gourio, F. (2007): "Labor Leverage, Firms Heterogeneous Sensitivities to the Business Cycle, and the Cross-Section of Returns," mimeo.

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.

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

Kaas, L., and P. Kircher (2011): "Efficient Firm Dynamics in a Frictional Labor Market," IZA discussion paper no. 5452.

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 (2012): "Entry, Exit, and Plant-level Dynamics over the Business Cycle," mimeo.

Lucas, R. E. (1978): "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 9(2), 508–523.

Moscarini, G., and F. Postel-Vinay (2010): "Stochastic Search Equilibrium," mimeo.

——— (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.

Schaal, E. (2010): "Uncertainty, Productivity and Unemployment in the Great Recession," mimeo.

Sedláček, P. (2011): "Firm Age, Business Cycles and Aggregate Labor Market Dynamics," mimeo.

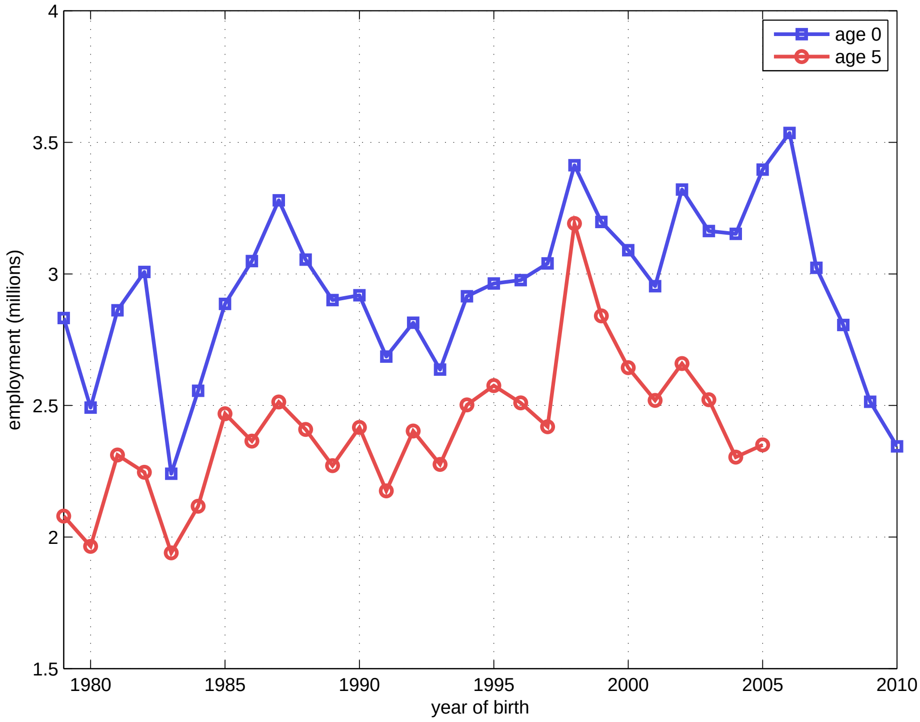
Appendix

.1 Sectoral evidence

To be included.

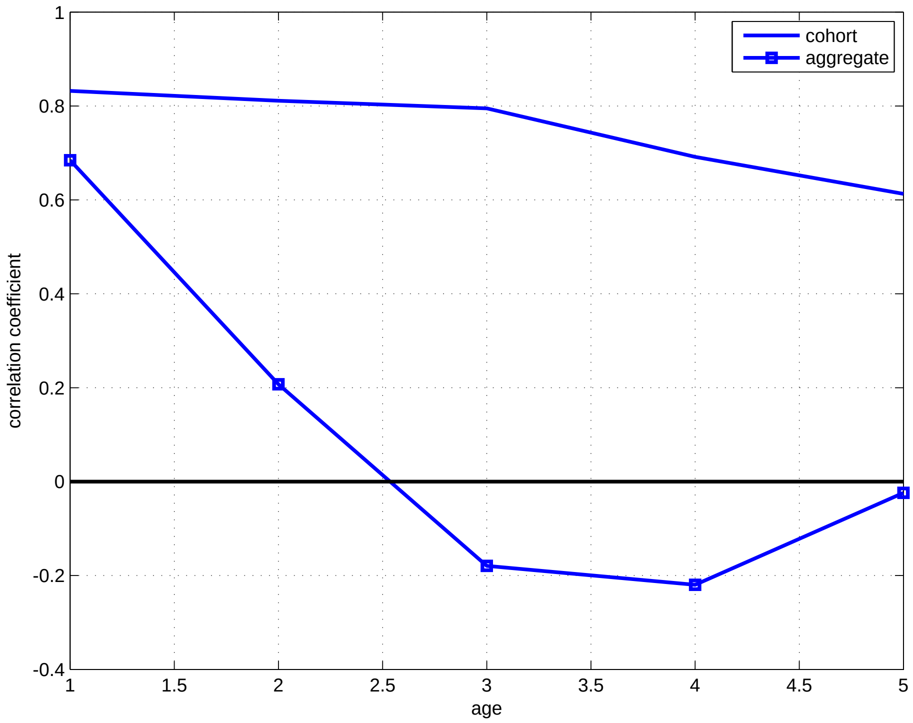
Tables and Figures

Figure 1: Total employment of firm cohorts of age 0 and 5



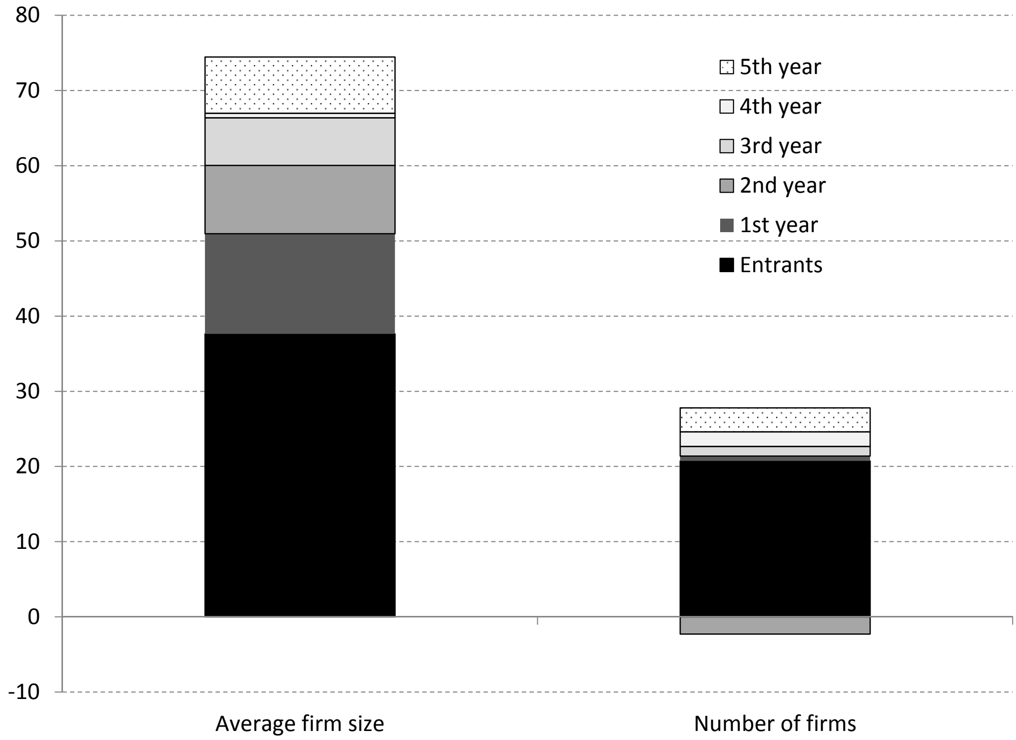
Notes: “age 0” is total job creation by entrants and “age 5” is total job creation of five year old firms. Source: BDS.

Figure 2: Autocorrelations



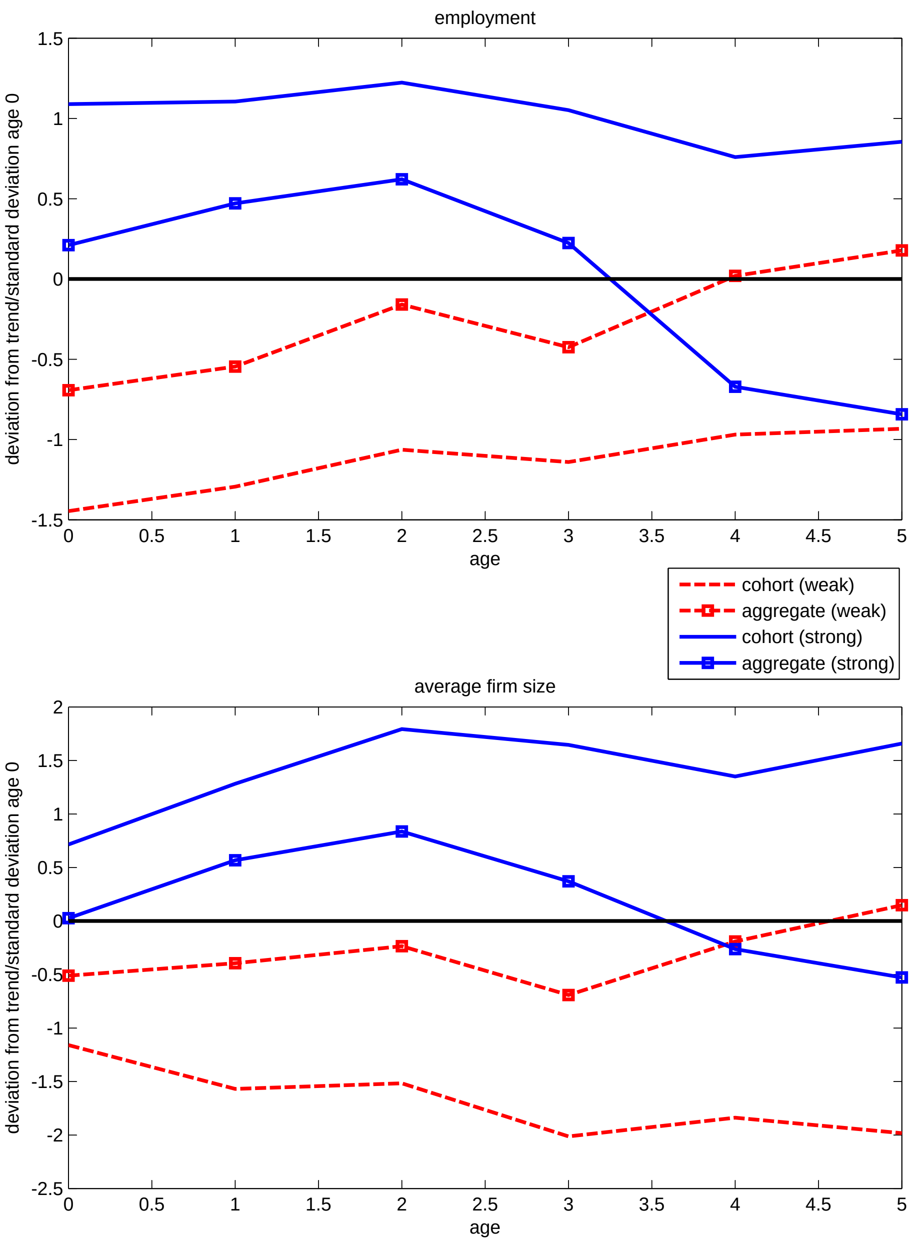
Notes: Correlation coefficients of employment in year $t = 0$ and in year $t + age$, with $age = 1, 2, 3, 4, 5$, at both the level of a cohort born in period $t = 0$ and at the aggregate level

Figure 3: Variance decomposition



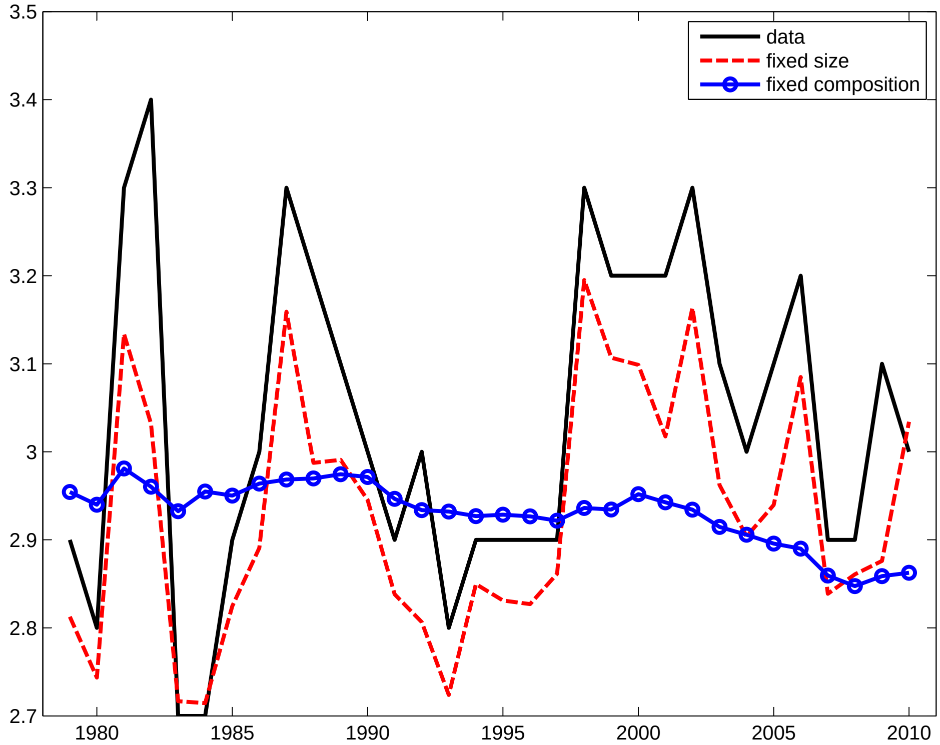
Notes: The figure plots contributions of average firm size and the number of firms to the variation in logged cohort-level employment of five year old firms expressed as percent of the total variation.
Source: BDS and authors' calculations.

Figure 4: Weak versus strong cohorts



Notes: The figure plots the evolution of employment and average firm size between period $t = 0$ and period $t + a$, both at the level of individual cohorts born in period $t = 0$ and at the aggregate level starting in same year. Data points are averages over two subsets of vintages: for the strong years $t = 0$ corresponds to the years 1987, 1998, 1999, 2002 and 2005, while for the weak years $t = 0$ corresponds to 1980, 1983, 1984, 1991, and 1993. The series are plotted as deviations from a linear trend, scaled by the standard deviation of the respective series in year $t = 0$, computed over the entire sample.
Source: BDS, authors' calculations.

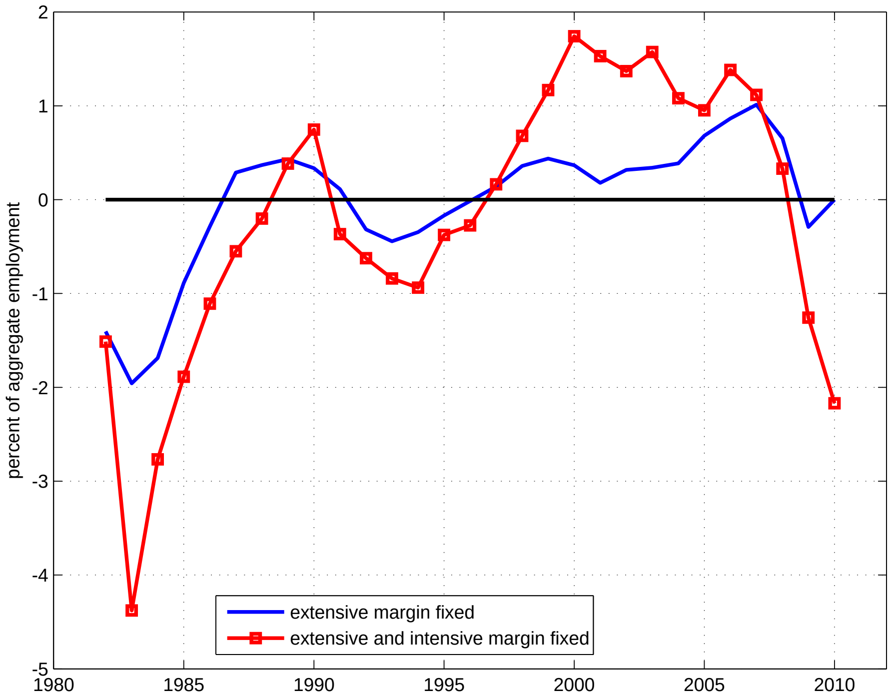
Figure 5: Entrant size: data and two counterfactuals



Notes: The figure plots entrant size ("data") and two counterfactuals. "Fixed composition" assumes that the firm shares within the BDS size categories are fixed and only average size within these categories varies. "Fixed sizes" assumes that the composition of firms varies but the average sizes within the BDS size groups are fixed.

Source: BDS and authors' calculations.

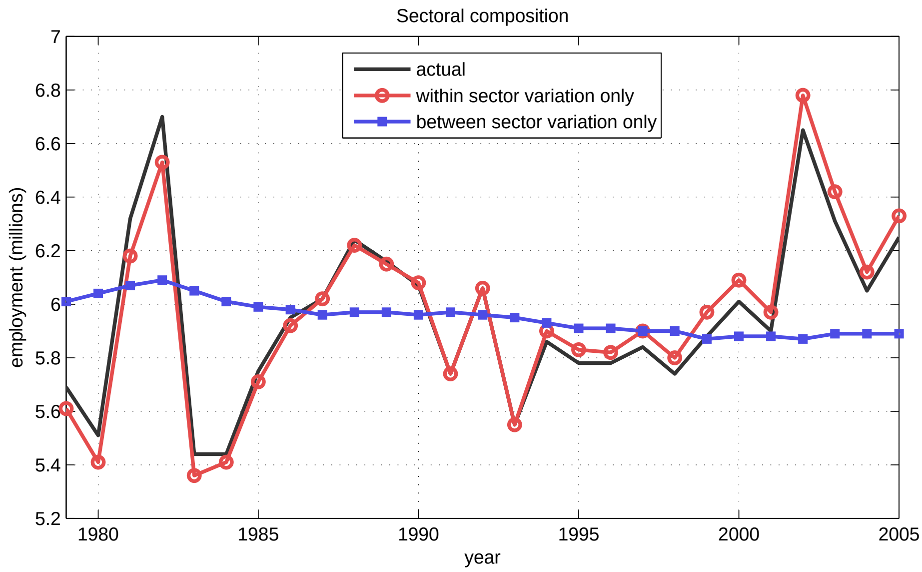
Figure 6: Employment differential

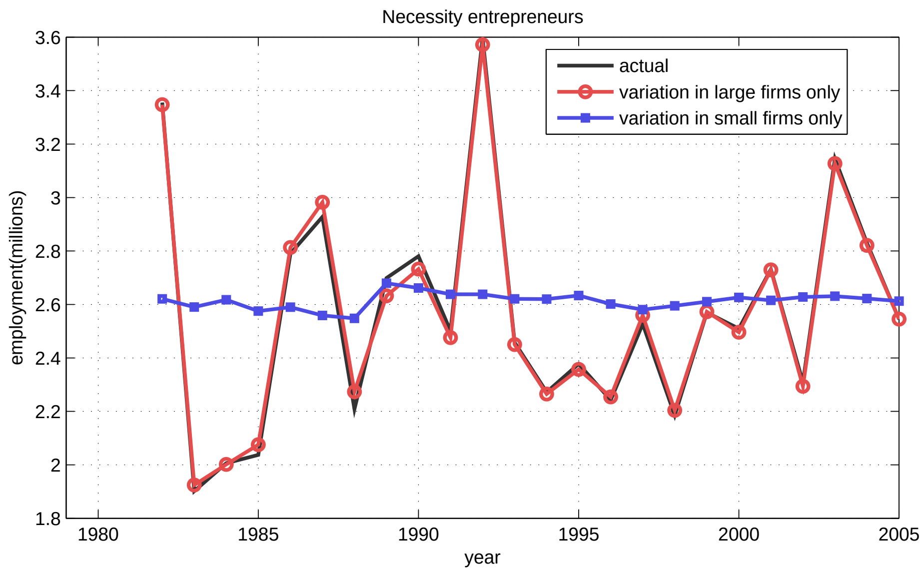


Notes: The figure plots the differential between aggregate employment and employment created under the assumption that the number of firms of age 0 to 5 years is fixed, "extensive margin fixed", and that both the number of firms and average size of firms of age 0 to 5 are fixed, "extensive and intensive margin fixed". The series are expressed in percent of aggregate employment.

Source: BDS and authors' calculations.

Figure 7: Counterfactuals: sectoral shifts and necessity entrepreneurs





Notes:

Source: BDS, authors' calculations.

Figure 8: Steady state: firm size by age and type

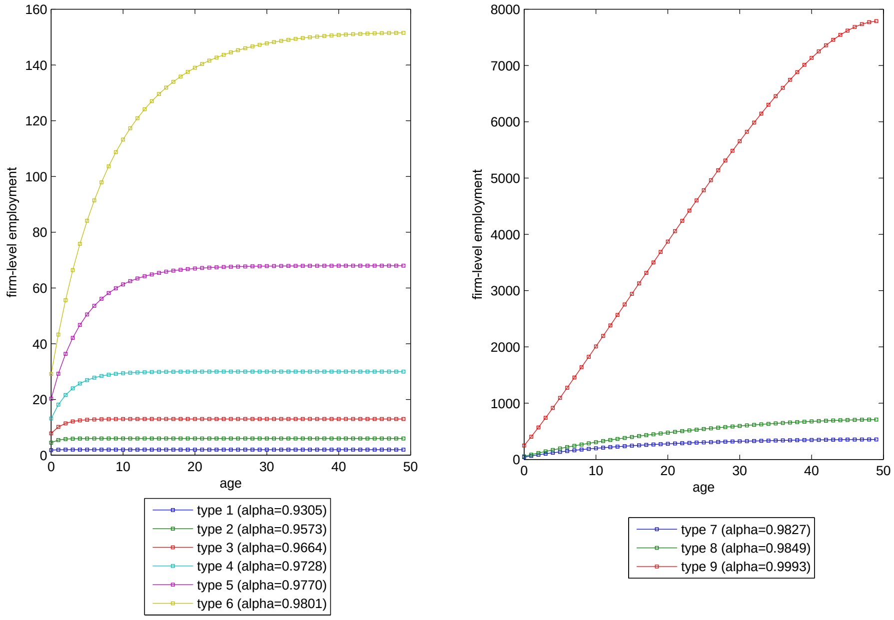


Figure 9: Steady state: contributions to cohort-level employment by age and type

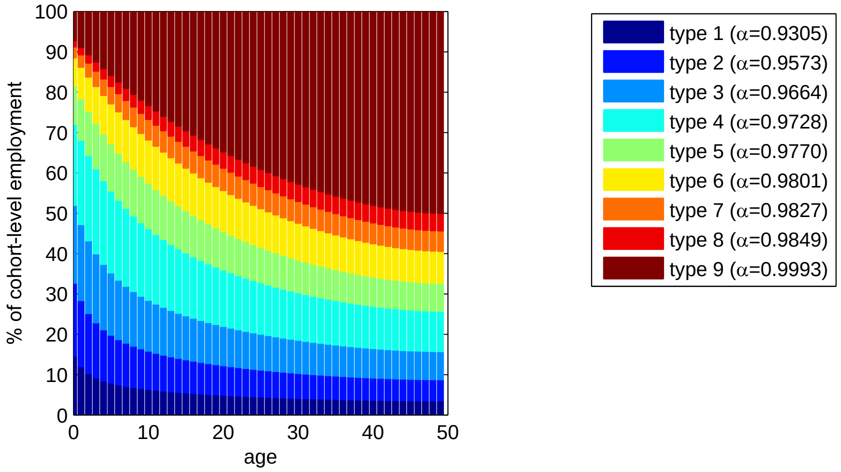


Figure 10: Benchmark model: historical decomposition

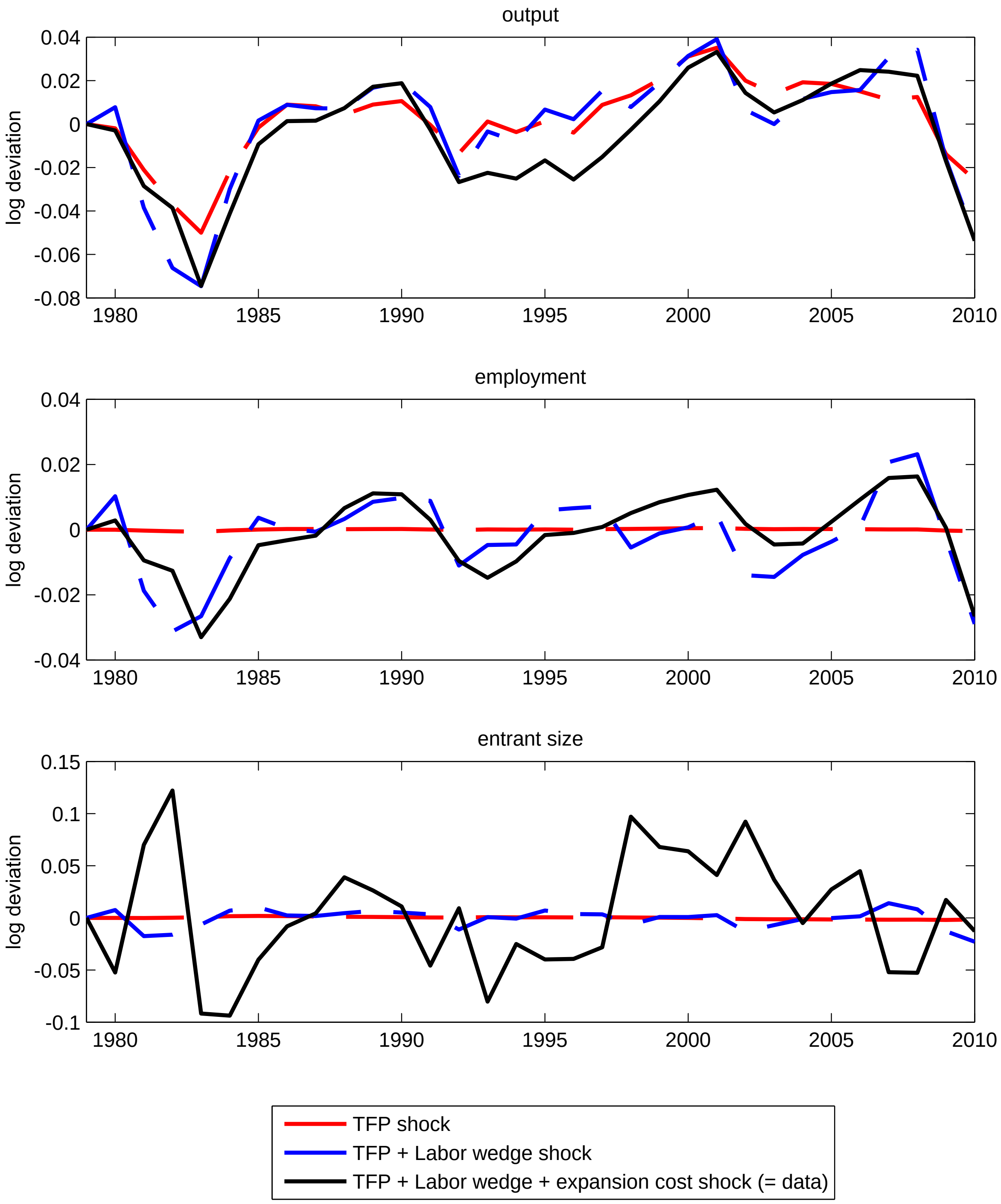
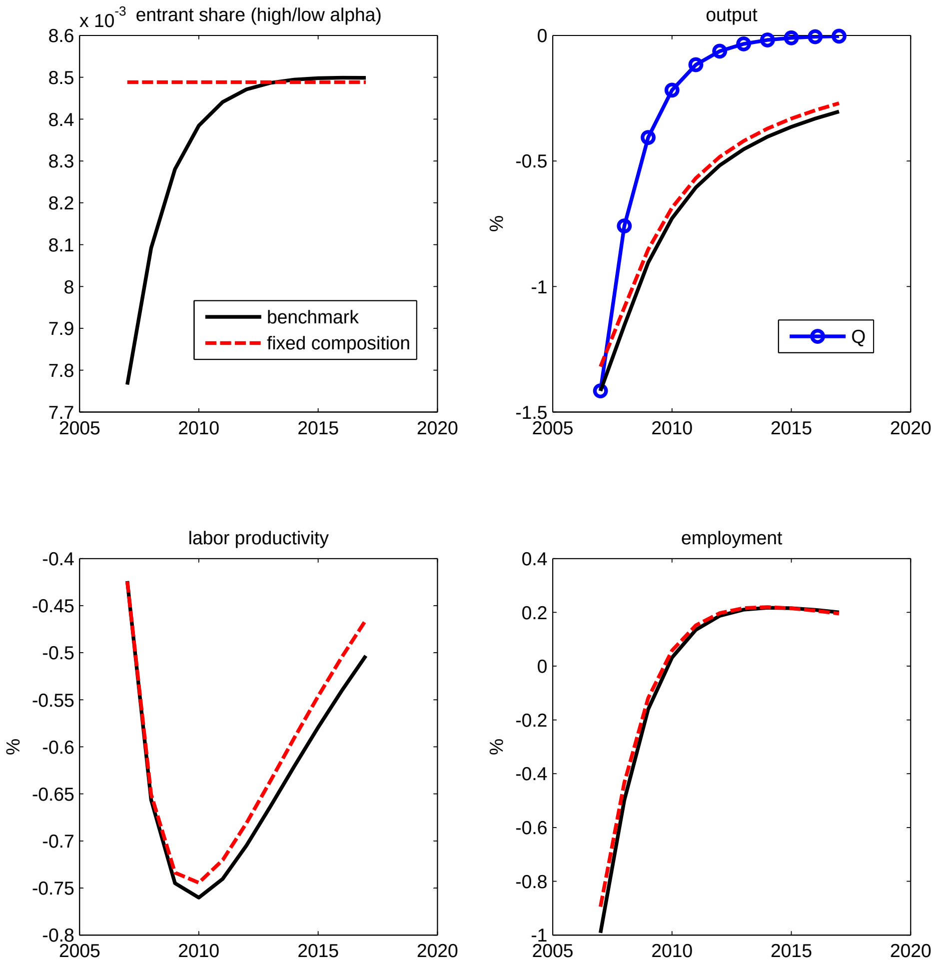


Figure 11: Benchmark and counterfactual model: responses to an expansion cost shock



Notes: The figure shows IRFs to a one-time expansion cost shock, the size of which is equal to the estimated value in 2007. All other shocks are fixed at zero. "Entrant share (high/low alpha)" is the firm share of high returns to scale (types 7-9) over low returns to scale (types 1-3) firms among entrants.

Table 1: Correlations of employment and average size with business cycle indicators in year t

age	Levels	linear trend		CF filter(6,12)	
	erate	erate	gdp	erate	gdp
A. Employment in year $t+a$					
cohort-level					
$a=0$	0.62	0.41	0.43	0.76	0.72
$a=5$	0.59	0.35	0.23	0.84	0.88
aggregate-level					
$a=0$		0.91	0.88	0.96	0.98
$a=5$		-0.07	-0.26	-0.67	-0.55
B. Average size in year $t+a$					
cohort-level					
$a=0$	0.50	0.36	0.33	0.74	0.61
$a=5$	0.44	0.28	0.10	0.74	0.74
aggregate-level					
$a=0$		0.75	0.74	0.76	0.72
$a=5$		-0.17	-0.37	-0.73	-0.65

Notes: The table reports correlation coefficients between the variables in the columns and rows. "CF filter(6,12)" refers to the Christiano and Fitzgerald (1999) filter with frequencies between 6 and 12 years, "erate" refers to the employment rate.

Table 2: Average firm size and firm shares in BDS size categories

	Firm size								
	$1-4$	$5-9$	$10-19$	$20-49$	$50-99$	$100-249$	$250-499$	$500-999$	>1000
average size	2	6	13	30	68	149	335	658	$3,315$
firm shares	47.6%	23.8%	14.6%	9.1%	2.7%	1.4%	0.4%	0.2%	0.2%

Notes: The table reports average firm sizes and firm shares within a given size class.

Table 3: Exit rates by age

	Firm age					
	1	2	3	4	5	> 5
data	0.22	0.15	0.12	0.11	0.10	0.06
model	0.22	0.14	0.11	0.09	0.08	0.06

Notes: The table reports firm exit rates in the BDS database and in the model according to age.

Table 4: Calibrated parameters

	parameter	value	target/estimate
β	discount factor	0.96	annual interest rate 4%
σ	relative risk aversion coefficient	1	log-utility
κ	utility of leisure parameter	1	unit Frisch elastiicty
ζ	adjustment cost, age 1-50	0.007	size of 1 year old firms
ζ_0	adjustment cost, entrants	0.041	size of entrants
ξ_0	exit rate coefficient	0.050	exit rates by age, BDS data
ξ_1	exit rate coefficient	0.170	exit rates by age, BDS data
χ	entry cost	0.930	entry costs = 0.073 GDP
Ψ	measure of business opportunities	0.090	$M = 1$, normalization
ϕ	elasticity in entry function	0.500	std(entry)/std(y)
ρ_A	TFP wedge, persistence	0.795	
σ_A	TFP wedge, standard deviation	0.011	
ρ_Q	adjustment cost wedge, persistence	0.550	
σ_Q	adjustment cost wedge, standard deviation	0.990	
ρ_Z	labor wedge, persistence	0.547	
σ_Z	labor wedge, standard deviation	0.023	

α_i	returns to scale					average size in BDS size classes			
	0.916	0.948	0.959	0.967	0.972	0.976	0.979	0.982	0.999
$P_i = \left(\frac{\psi_i}{x_i}\right)^{1-\phi}$	probability of starting up a type i firm					firm shares in BDS size classes			
	0.799	0.451	0.272	0.153	0.087	0.051	0.030	0.018	0.001

Notes: The table the calibrated 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 (P_i) conditional on paying the startup cost.

Table 5: Model results

	benchmark	lower bound	upper bound
maximum (mean) absolute effect of entrant composition:			
output	0.28(0.06)	0.06(0.02)	1.74(0.61)
employment	0.22(0.10)	0.12(0.05)	1.97(0.58)
std(size5Y) model/data:			
	1.17	0.26	7.10
std(entrants) model/data:			
	0.98	0.30	7.02

Notes: The table shows the maximum and mean absolute difference between the data and the counterfactual with fixed composition for output and employment. Benchmark refers to our preferred strength of composition effects, lower and upper bound refer to the extreme cases of minimum and maximum composition effects as described in the main text. $std(size5Y)$ and $std(entrants)$ refer to the respective standard deviations of the size of five year old firms and the number of entrants.