Entry Decision, the Option to Delay Entry, and Business Cycles

Ia Vardishvili[1]

April, 2022

Click here for the latest version

Abstract

I show that firms' ability to postpone entry has important implications for our understanding of the observed business cycle behavior of start-ups. I use a model that closely replicates the main features of the US firm dynamics to explore and quantify the mechanism. I find that the option to wait endogenously generates a countercyclical opportunity cost of entry: during recessions, a higher risk of failure increases the value of waiting, hence the cost of entry. The mechanism increases the elasticity of entrants to aggregate shocks five times. It is responsible for three-fourths of the observed persistent differences in the recessionary and expansionary cohorts' productivity, survival, and employment. Without the channel, existing models require either large shocks that generate excessive aggregate fluctuations or exogenous mechanisms to reconcile the observed dynamics of entrants. Overlooking this channel may also result in misleading predictions about entrants' responses to different shocks or policies.

Keywords: Entry Decision, Delay, Option Value, Firm Dynamics, Business Cycles

JEL Codes: D25, E22, E23, E32, E37, L25

[1]Email: izv0013@auburn.edu. Address: Department of Economics, Auburn University, 139 Miller Hall, Auburn, AL, United States, 36849. I am very grateful to my advisors Eric Young, Toshihiko Mukoyama, and Sophie Osotimehin for their continuous advice and encouragement. I am also immensely grateful to Nicolas Ziebarth and Veronika Penciakova for great suggestions, comments and feedback. I would also like to thank Alexander Monge-Naranjo, B. Ravikumar, Diego Legal Canisa, Fabio Ghironi, Gaston Chaumont, Ilhan Guner, John Haltiwanger, Jorge Miranda-Pinto, Juan J. Dolado, Juan Sanchez, Julia Thomas, Leland Farmer, Givi Melkadze, Miguel Mascarua, Sergio Ocampo Diaz, Zach Bethune, seminar participants at University of Virginia, St. Louis Fed, UVA-Richmond Fed Jamboree, Auburn University, Midwest Macro at Vanderbilt, EGSC at Washington University in St. Louis, I-85 Conference, Virtual World Congress, 2019 GCER Alumni Conference, EESWM2019, EEA2019, SED 2021, and the 6th Lindau Meeting on Economic Sciences for their comments and feedback. In addition, I thank the Federal Reserve Banks of St. Louis for hosting me during the writing of this paper; I also want to thank Aysegul Sahin and Aparil Franco for their useful feedback and guidance at CSWEP CeMENT Workshop. Finally, I thank Bankard Fund for Political Economy for financial support. All errors are of my own.

1 Introduction

Figure 1 illustrates the business cycle dynamics of the number of entrants together with real GDP and aggregate employment. The number of entrants is three times as volatile as real GDP and four times as volatile as aggregate employment. The recent empirical evidence indicates that the composition of entrants also significantly varies with the initial aggregate conditions. Specifically, cohorts of firms that start operating during recessions employ fewer workers at entry and over time, although they are, on average, more productive and have higher survival rates than their expansionary counterparts.[2] Despite the importance of start-ups for aggregate job creation and economic growth, we lack a microfounded explanation of what drives the observed selection of entrants across initial conditions.

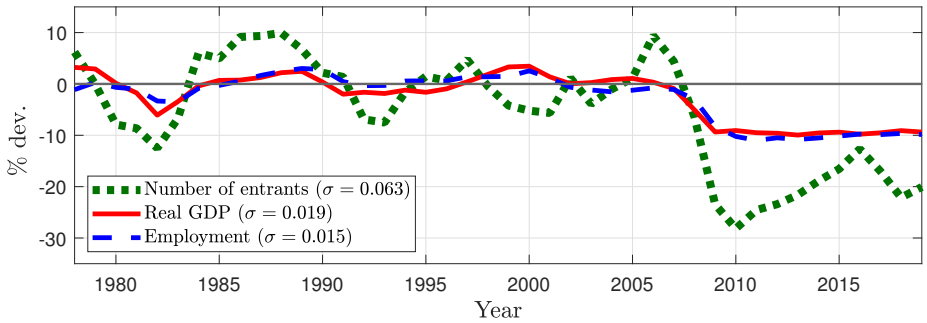
I show that firms' ability to delay entry has important implications for our understanding of start-ups' business cycle dynamics. The theory is motivated by a large body of microeconomics literature, which shows that an option to postpone an irreversible project makes investment especially sensitive to aggregate risks.[3] The neoclassical investment rule – invest in a project when its net present value (NPV) is non-negative – has been widely criticized for ignoring the option. Starting a business is a largely irreversible investment associated with high failure risks that vary with the aggregate states. Therefore, the option to choose initial conditions before committing resources could fundamentally alter firms' entry decisions. However, existing firm dynamics models that find it challenging to reconcile the observed dynamics of entrants have not considered the channel yet. In this paper, I develop a tractable framework to evaluate theoretically and quantitatively the role the option to delay plays in the documented business cycles dynamics of entrants and economic aggregates.

I start by providing empirical support for the option-to-delay channel. First, using the annual Business Dynamics Statistics (BDS) dataset over the period 1978-2019, I document that cohorts that start operating during recessions consist of fewer firms that, on average, survive longer than their expansionary counterparts. The differences in the survival rates do not dissipate over the cohorts' life cycle. The latter finding supports the hypothesis that the initial aggregate conditions have a significant effect on the types of firms that decide to

[2] Lee and Mukoyama (2015), Moreira (2016), and Ates and Saffie (2021) find that firms that are born during recessionary periods are, on average, more productive at entry and over time. Moreira (2016) and Sedlacek and Sterk (2017) document that cohort-level employment is significantly and persistently procyclical. I use annual state-level BDS data over 1978-2019 to additionally document the persistent differences between the cohort-level employment, firms, and survival; see Tabe 8.

[3] For example, see Pindyck (1991), Bernanke (1983), and McDonald and Siegel (1986). See Dixit and Pindyck (1994) for a detailed discussion.

Figure 1



Note. The empirical time series represents the deviations of the log number of entrant establishments, log real GDP and log aggregate employment from their respective trends in the US over the period 1978-2019. To find the cyclical properties of these time series, I use a linear detrending method that allows a structural break in the trend (for details see Appendix D.2). *Source:* BDS, FRED.

enter the market. Second, I provide empirical evidence showing that entrants choose to wait for better aggregate conditions before committing their resources. I use the newly developed Business Formation Statistics (BFS) dataset that allows tracking the timing of the market entry decisions made by aspiring start-ups. I show that the share of firms that postpone starting a business is negatively correlated with the aggregate conditions at entry.

To explore the implications of the option-value channel, I develop a heterogeneous firm dynamics model with endogenous firm entry and exit and aggregate demand volatility. Firms operate in monopolistically competitive markets and make decisions about production and exit. Potential entrants hold heterogeneous signals about their post-entry initial productivity. I deviate from the existing framework and allow entrants to keep their signals over time if they decide to postpone entry after observing the aggregate demand level. Entering today or entering tomorrow are mutually exclusive alternatives, leading to a non-negative option value of delay, which varies with the signal and with the aggregate demand level.

I find that the option to wait leads to an endogenous countercyclical opportunity cost of starting a business, which increases the elasticity of entrants with respect to the initial aggregate conditions. The mechanism works through procyclical variation in survival rates: during recessions, in addition to lower profits, potential entrants expect to lose part of their long-run benefits due to the increased risk of post-entry failure. The higher the expected long-run value, the higher the expected cost of prematurely exiting the market. With the intertemporal choice, the latter value increases the threshold cost of entry, generating a new group of firms that choose to stay outside the market even if the net expected lifetime profits are more than zero.

To evaluate the quantitative implications of the option-value channel, I parameterize the model to match the main features of the US entrants' average life-cycle dynamics. The calibrated model closely replicates the average size, survival, and exit hazard rates of firms up to age 30. It also successfully matches the share of firms by cohort age in the total number of firms and the share of employment by cohort age in aggregate employment. The key parameters that shape the business cycle firm dynamics are the ones that drive the aggregate demand shock process. The exogenous shocks affect incumbent firms' production and exit decisions and determine entrants' expected lifetime profits. At the same time, the shock process shapes the endogenous countercyclical entry cost function, hence the option-value effect. Therefore, I discipline the channel by matching jointly the business cycle dynamics of the total number of firms and entrants in the model and the data.

I show that the calibrated model successfully accounts for the documented persistent and significant differences in the life cycle characteristics of cohorts born at different stages of business cycles. Specifically, due to the increased cost of entry, the group of firms that enter the market during recessions is, on average, more productive than their expansionary counterparts. At the same time, the recessionary cohorts consist of around 12% fewer firms, employ 8% fewer workers and have a 2.1 percentage point higher survival rates at entry. The differences persist over cohorts' life cycles. That is, at age five, the recessionary cohorts consist of around 8% fewer firms, employ 7% fewer workers and have a 1.8 percentage point higher survival rates than their expansionary counterparts. Using the annual state-level BDS dataset about the life cycle dynamics of firms, I show that the model predicted differences are quantitatively very close to the data counterpart.

I find that the selection of entrants across aggregate conditions has an important role in shaping the cohort-level and aggregate dynamics. First, I show that more than two-thirds of the differences between the recessionary and expansionary cohorts are due to the variation in the number and composition of firms at entry rather than the shocks they face over their life cycle. Next, I quantify the role of the persistent cohort dynamics in aggregate fluctuations. Toward the end, I study baseline economy's response to a shock series that matches the variation in the number of entrants from 1978 to 2019 in the data – illustrated in Figure (1), and the model. I find that the simulated dynamics of the economic aggregates closely track the data counterpart. That is, the correlation between the cyclical component of aggregate employment in the model and the data is around 0.93. The variation in the types of firms at entry is responsible for roughly 20% of the total variance in aggregate employment, which is a considerable contribution compared with a small share of entrants' employment.

The option-value channel is quantitatively important in accounting for the observed business cycle dynamics of entrants. Firms decide not to start businesses during recessions because the NPV of entry falls (direct effect), and the cost of entry increases due to the option to wait (indirect effect). To isolate and quantify the indirect effect, I evaluate the model's performance without the intertemporal choice. I find that the endogenous countercyclical opportunity cost of entry increases the elasticity of entrants to aggregate shocks five times and decreases the differences in the productivity composition of entrants by ten times. The direct effect is responsible for only around 20% and 25% in the persistent drop of the recessionary cohorts' firms and employment, respectively. Moreover, the countercyclical survival rates are completely driven by the medium-productivity entrants who choose to delay business formations. As for the aggregate fluctuations, the selection through the option-value channel is responsible for 25% volatility in the model-simulated aggregate employment, which corresponds roughly to 12% volatility in the data counterpart.

Next, I compare the performance of the baseline model to a workhorse firm dynamics model, parameterized to account for the same set of facts. I find that a model without delay requires large shocks to increase the elasticity of start-ups to initial conditions by increasing the variation in the NPV of entry. The response of incumbent firms to the calibrated shock process leads to excessive aggregate fluctuations that, in turn, significantly underestimate the relative importance of entry margin. Existing literature uses various approaches to reconcile the observed selection of entrants for a reasonable aggregate demand shock process. One option is to use exogenous entry cost shock or a cost function that varies with the business cycles, as in Lee and Mukoyama (2018), and Clementi and Palazzo (2016), or introduce entry function, which allows choosing the elasticity of entrants to aggregate shocks, as in Sedlaček and Sterk (2017). In that respect, the option-value channel, which endogenously leads to a countercyclical cost of entry that increases the elasticity of entrants to aggregate shocks, provides a microfoundation for these exogenous mechanisms.

Furthermore, I argue that overlooking the option-to-delay channel may lead to misleading predictions about potential entrants' responses to different shocks or policies. The reason is the following. With the intertemporal choice, the dynamics of entrants depend on how the changes in the aggregate environment affect the relative benefits of entry today versus tomorrow. Whereas the standard frameworks only account for the shock's direct effect. I show that the option-value channel qualitatively and quantitatively alters the existing model's predictions about the responses of entrants to shocks, depending on the magnitude, timing, and duration of the shocks.

Relation to the Literature The paper contributes to the empirical firm dynamics literature that studies the effect of the initial aggregate conditions at entry on the life cycle characteristics of cohorts. Lee and Mukoyama (2015), Moreira (2016), and Ates and Saffie (2021) find that firms that are born during recessionary periods are, on average, more productive at entry and over time. Moreira (2016) and Sedlacek and Sterk (2017) document that recessionary firms are smaller and employ fewer workers at entry and over time compared to their expansionary counterparts. I additionally document that the survival rates of cohorts' of establishments/firms are negatively correlated with the aggregate conditions at the time of entry. The differences persist over cohorts' life cycles. Interestingly, I find no statistically robust relationship between cohorts' average size by age and the aggregate conditions at entry.

The paper is related to the existing firm dynamics literature that explores the mechanism behind the observed significant variation in the number and composition of entrants. Samaniego (2008) finds that entry and exit are insensitive to productivity shocks of a reasonable magnitude. Lee and Mukoyama (2018) show that generating the documented significant selection of entrants in Hopenhayn and Rogerson (1993) framework is a puzzle that can be solved by introducing an entry cost that varies over the cycles in a particular way. Others use exogenous entry cost shock (e.g., Clementi and Palazzo, 2016), or introduce entry function, which allows choosing the elasticity of entrants to aggregate shocks (e.g., Sterk, Sedlaćek, and Pugsley, 2021). The expected life cycle profits are insensitive to a mean-reverting aggregate demand shock process in existing frameworks. Thus, the traditional entry decision does not account for the significant variation in the number and composition of entrants unless the magnitude of the shocks is large. Interestingly, the empirical microeconomics literature also finds that the standard entry decision rule does not explain much variation in entry, as the expected profits do not vary much over time (e.g., O'Brien, Folta, and Johnson (2003), Geroski, 1995).

The theory is motivated by a considerable theoretical and empirical microeconomics literature that emphasize the importance of the option to postpone an irreversible project in explaining the investment behavior under aggregate uncertainty (e.g., Pindyck (1991), Bernanke (1983) and McDonald and Siegel (1986). For detailed review see Dixit and Pindyck, 1994). Pindyck (2009) shows that various risks to post-entry profits magnify the cost of entry. Motivated by the theory, I use the newly developed BFS dataset to provide empirical evidence showing that start-ups choose to wait for better aggregate conditions before committing their resources. Second, I extend the workhorse Hopenhayn (1992) firm dynamics model

to account for the option of timing in the entry decision. I find that the ability to postpone entry and the variation in the risk of failure with the initial aggregate conditions generates an endogenous countercyclical cost of entry that significantly increases the elasticity of entrants to aggregate shocks. That said, the paper also relates to the macroeconomics literature that studies the role of real options in shaping aggregate dynamics (e.g., Jovanovich (1993), Veracierto (2002), Bloom, 2009). Fajgelbaum, Schaal and Taschereau-Dumouche (2017) show that in a framework where agents learn from the actions of others, high uncertainty about fundamentals discourages investment through the option-value channel.

The selection of entrants through the option-to-delay channel complements existing literature that explores the forces that drive the observed persistent and significant differences in the recessionary and expansionary cohorts. Sedlaćek and Sterk (2017) reconcile these differences through the variation in the share of niche and mass product firms, while Smirnyagin (2021) emphasizes the importance of the share of high-target size firms across cohorts. Despite the different mechanisms, all these papers find that the selection of types of firms at entry, drives the persistent cohort-level differences over the business cycles. Pugsley, Sedlaćek, and Sterk (2016) also document that the heterogeneity across cohorts' life cycle dynamics is due to the pre-entry selection of firms rather than post-entry shocks. My paper is also closely related to Gourio, Messer and Siemer (2015) that emphasize the role of the 'missing generation' in propagating the cohort-level and aggregate dynamics.

The paper contributes to the firm dynamics literature that evaluates the role of entry margin in shaping aggregate fluctuations. Lee and Mukoyama (2008), Bilbiie, Ghironi, and Melitz (2012), Clementi et al. (2014), and Clementi and Palazzo (2016) find that endogenous dynamics in the entry and exit margins significantly propagate aggregate shocks. Haltiwanger et al. (2013) emphasize the importance of accounting for the life-cycle demographics of entrants in measuring and understanding the total contribution of the entry margin to economic growth. Using a model that closely replicates the life cycle characteristics of cohorts of firms in the US, on average, and over the business cycles, I show that the persistent differences in cohorts' characteristics significantly amplify and propagate aggregate shocks. This paper relates to the literature that studies the role of entry in the post-Great Recession slow recovery (e.g., Gourio, Messer and Siemer (2016), Siemer (2016), Khan, Senga, and Thomas (2016), Sedlaćek, 2020) and to the literature that investigates the propagation mechanism of standard business cycle models (e.g., Cogley and Nason (1995), King and Rebelo, 1999).

2 Empirical Evidence

In this section, I provide empirical support for the option-to-delay channel. First, I document that cohorts that start operating during recessions, on average, survive longer than their expansionary counterparts. This finding supports the hypothesis that the composition of firms that decide to enter the market varies with the initial aggregate conditions. Second, I provide empirical evidence showing that entrants choose to wait for better aggregate conditions before committing their resources.

2.1 Aggregate Conditions at Entry and Cohorts' Survival Rates

To study how firms' survival rates vary with the aggregate conditions at entry, I use the US- and state-level annual time series about the number of establishments/firms by age from the Business Dynamics Statistics (BDS) dataset.[4] I measure a survival rate of a cohort of age g at year t as

$$S_{g,t} = \frac{N_{g,t}}{N_{0,t-g}},$$

where $N_{g,t}$ measures the number of establishments (firms) in a cohort of establishments (firms) of age g at year t; $N_{0,t-g}$ measures the number of establishments (firms) in the same cohort at the time of entry (age 0).[5] In this analysis, I consider cohorts' survival rates for up to age 5.[6] I measure the aggregate conditions at entry using the cyclical component of log annual real GDP.[7] I find the latter using the HP filter with a smoothing parameter of *100*.

Figure 2 provides binned scatter plots of pooled cohorts' life cycle survival rates at the US-level against the business cycle indicators at the time of entry. The binned scatter plots include age-specific and year-specific fixed effects. The latter controls for the sequence of aggregate shocks cohorts face after entry. Panel (a) of Figure 2 shows that the business cycle conditions at entry is negatively associated with cohorts' survival rates. Panel (b) of Figure 2 shows that the negative relationship is robust if we consider cohorts of firms rather than

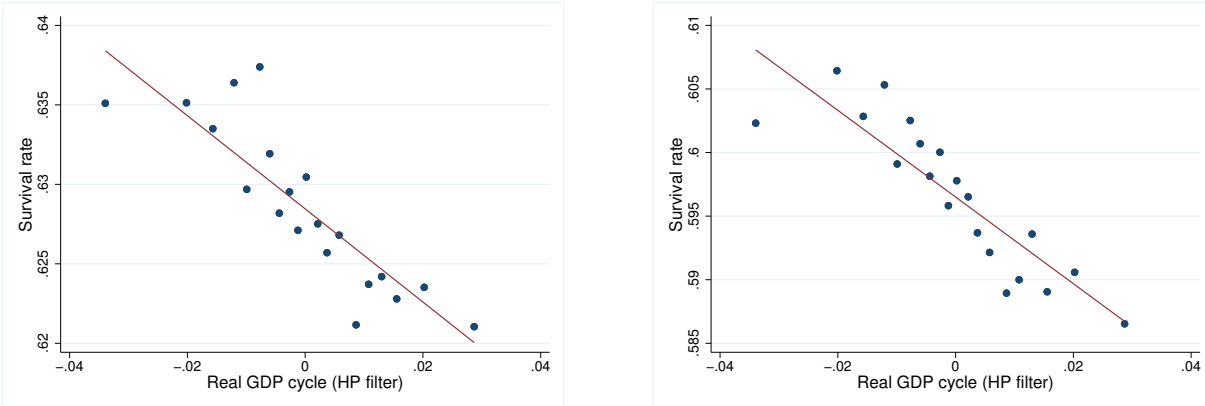
[4]The BDS dataset covers the universe of employer businesses in the US and provides yearly measures of business dynamics aggregated by the establishment and firm characteristics. An establishment is defined as a fixed physical location where economic activity is conducted. A firm might span multiple physical locations and consist of one establishment or many establishments.

[5]Employer businesses are identified as start-ups (age 0) based on their first payroll information in the Longitudinal Business Database.

[6]The publicly available part of the BDS dataset only provides information about cohorts from age 0 to age 5. Information about cohorts above age 5 is binned into 5-year age groups.

[7]I annualize the quarterly real GDP data so that it's consistent with BDS timing. The source and the construction of the annual real GDP data are described in Appendix D. For more information see the link.

Figure 2: Correlation between the survival rates and aggregate economic conditions at entry



(a) Cohort of establishments

(b) Cohort of firms

Note: Each panel plots a binned scatterplot of the survival rates up to age five against the aggregate conditions at entry. I measure the latter using the cyclical component of HP-filtered log real GDP. The time series is at the US level. Bin scatter controls for year-fixed effects and age-fixed effects.

establishments.

Next, I use the state-level variation in the life cycle dynamics of new businesses to further investigate the relationship. I estimate the following regression:

$$S_{c,g,s,t} = \alpha + \beta Z_c + \eta_g + \theta_t + \gamma_s + \varepsilon_{g,s,t}, \tag{1}$$

where $S_{c,g,s,t}$ is a survival rate of a cohort c at age g, in state s, at time t; Z_c represents the economic conditions at the time when the cohort first entered the market.[8] η_g, θ_t, γ_s represent age-, year-, and state-fixed effects, respectively. That said, β measures a percentage point change in the cohorts' average survival rates due to the variation in the business cycle conditions at entry. The coefficient should capture the effect of the initial economic conditions on the cohorts' average survival rates after controlling for the age, year and state fixed effects.

Panel A of Table 1 reports the results of the regression equation (1) when the unit of analysis is a cohort of establishments. Column (1) indicates that cohorts born during good economic conditions are characterized by lower survival rates over their life cycle. Specifically, a 1-percentage point increase in real GDP above the trend decreases cohorts' average survival rates by 0.28 percentage point. For robustness, I additionally consider the following business cycle indicators. Column (2) uses a business cycle indicator that refers to a year as a recession

[8]The specification is also similar to the age-period-cohort model where cohort effects are a proxy for economic conditions at birth. See Moreira (2005) for more details.

Table 1: The survival rates and aggregate economic conditions at the time of entry.

	Panel A. Establishment			**Panel B. Firm**		
	Y_{HP}	$Y_{HP,I}$	$NBER$	Y_{HP}	$Y_{HP,I}$	$NBER$
	(1)	(2)	(3)	(1)	(2)	(3)
β	-0.28***	-0.013***	-0.015***	-0.33***	-0.014***	-0.010***
	(0.03)	(0.001)	(0.001)	(0.03)	(0.001)	(0.001)
Age FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓	✓
Observations	9,945	9,945	9,945	9,945	9,945	1,989
R^2	0.603	0.959	0.959	0.955	0.956	0.596

Note: Robust standard errors clustered at the state-level are in parentheses. The dependent variable represents cohorts' survival rates up to five years of operation. Panel A uses cohorts of establishments and Panel B uses cohorts of firms as a unit of analysis. Columns (1)-(3) use different business cycle indicators at entry. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

if the cyclical component of the log real GDP is below trend ($Y_{HP,I}$). Column (3) uses the NBER-based indicator of a recession that spans the period following the peak through the trough ($NBER$).[9] The indicator equals -1 if the year is indicated as recession, 0 otherwise. Columns (2) and (3) show that cohorts born during recessions, on average, have higher survival rates compared to their expansionary counterparts. Panel B of Table 1 shows that the results hold if I use cohort of firms as a unit of analysis rather than establishments.

To additionally investigate whether the effects of the initial aggregate conditions disappear over cohort's life cycle, I consider the regression specification where I interact business cycle conditions at entry with cohort age:

$$S_{c,g,s,t} = \alpha + \sum_{g=1}^{5} \beta_g D_g Z_c + \eta_g + \theta_t + \gamma_s + \varepsilon_{g,s,t}, \tag{2}$$

where D_g is an indicator variable that takes the value of one if the business establishments/firms are g years of age. The coefficient β_g measures the change in the survival rates of a cohort at age g with the variation in the business cycle conditions at entry.

Panel A of Table 2 reports the regression results. $1_{\{age=g\}} \times Z$ describes the interaction of the business cycle indicators with the cohort of age g. Column (1) of Panel A shows that the aggregate conditions at entry have a statistically significant and persistent effect

[9] The latter indicator specifies the peak and trough dates on a monthly frequency. Using the monthly data, I define a year t as a recession if at least four months from April in year $t-1$ to April t are indicated as recessionary periods. Based on the definition, the recessionary years are 1981, 1982, 1983, 1991, 2002, and 2009. All other years are defined as expansionary.

Table 2: Survival rates by age and aggregate economic conditions at entry

	Panel A. Establishment			Panel B. Firm		
$Z =$	Y_{HP}	$Y_{HP,I}$	$NBER$	Y_{HP}	$Y_{HP,I}$	$NBER$
	(1)	(2)	(3)	(1)	(2)	(3)
$1_{\{age=1\}} \times Z$	-0.20***	-0.007***	-0.024***	-0.35***	-0.010***	-0.023***
	(0.04)	(0.001)	(0.001)	(0.04)	(0.001)	(0.001)
$1_{\{age=2\}} \times Z$	-0.25***	-0.011***	-0.016***	-0.33***	-0.013***	-0.013***
	(0.04)	(0.001)	(0.001)	(0.04)	(0.001)	(0.001)
$1_{\{age=3\}} \times Z$	-0.31***	-0.016***	-0.013***	-0.33***	-0.015***	-0.008***
	(0.03)	(0.001)	(0.001)	(0.03)	(0.001)	(0.001)
$1_{\{age=4\}} \times Z$	-0.33***	-0.017***	-0.010***	-0.34***	-0.016***	-0.002***
	(0.03)	(0.001)	(0.001)	(0.03)	(0.001)	(0.001)
$1_{\{age=5\}} \times Z$	-0.33***	-0.016***	-0.008***	-0.32***	-0.015***	-0.002**
	(0.02)	(0.001)	(0.001)	(0.03)	(0.001)	(0.001)
Age FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓	✓
Observations	9,945	9,945	9,945	9,945	9,945	9,945
R^2	0.959	0.956	0.959	0.955	0.956	0.955

Note: Robust standard errors clustered at the state-level are in parentheses. The dependent variable represents cohorts' survival rates by age. Panel A uses a cohort of establishments, and Panel B uses a cohort of firms as a unit of analysis. Columns (1)-(3) use different indicators for characterizing the business cycle conditions at entry. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

on cohorts' survival rates: cohorts of establishments that start operating during recessions are characterized by higher survival rates at entry and over time. Moreover, the results are robust to alternative business cycle indicators. The results also hold if we use a firm as the unit of analysis rather than an establishment.

To interpret the results, note that the initial economic conditions have two counteracting effects on new cohorts' survival rates. On the one hand, unfavorable economic conditions directly decrease cohorts' survival rates due to higher failure rates. On the other hand, cohorts' survival rates could go up due to the selection of better firms at entry during bad aggregate conditions. The finding that cohorts' average survival rates are countercyclical supports the hypothesis that the initial aggregate conditions significantly affect the selection of firms at entry.

2.2 The Evidence of Entry Timing

Next, I document that part of the business cycle variation in the number of start-ups comes from the firms' option to delay entry. Identifying the latter requires information about the dynamics and decisions made by aspiring start-ups before they enter the market. The newly

developed Business Formation Statistics (BFS) dataset provides a subset of the information.

The BFS dataset is based on applications for Employer Identification Numbers (EINs) submitted in the US, known as IRS Form SS-4 filings.[10] Information provided in the EIN application is used to identify a subset of applications associated with the start of new businesses, referred to as business applications (BA).[11] The BAs are matched to the set of firms in the BDS dataset identified as new employer businesses based on payroll information. The matching process is straightforward because both of the datasets contain Information about EINs. The publicly available part of the BFS dataset provides the US- and state-level time series about the number of employer start-ups that form businesses within the first eight quarters from the date of the EIN application ($\boldsymbol{F8Q}$). This group of businesses covers more than 80% of the total number of entrants each year in the US.[12]

In the analysis, I consider the time series of the number of applications that form businesses within the *first* four ($\boldsymbol{F4Q}$) and *second* four ($\boldsymbol{S4Q}$) quarters from the date of the application. To identify the business cycle dynamics of start-ups due to the option to delay entry, I construct a times series about the share of the applications that form businesses with one year delay, $\mathbf{S}4Q/F8Q$. I refer to this variable as the *share of late start-ups*.[13] I use the latter time series to test the following hypothesis. Suppose the aggregate state has a significant effect on the number of start-ups through the option-to-delay channel. Then, the share of the applications that form businesses with one year delay should increase if the aggregate conditions at the time of the applications are unfavorable.[14]

Table 3 reports the summary statistics of the share of late start-ups. At the state level (Panel

[10]The EIN is a unique number assigned to most of the business entities. The EIN is required when the business is providing tax information to the Internal Revenue Service (IRS).

[11]The EIN application contains information about reasons for applying, type of entity, business start date, the expected maximum number of employees, the first wage pay date, the principal activity of a business, and etc.

[12]For more details see Appendix A.3, Figure 21.

[13]Information about the raw number of EIN applications alone does not help identify delays in business formation due to the following reasons. On the one hand, potential entrants who delay entry might not apply for the EIN applications. Thus, they are not included in the BFS dataset. On the other hand, some parts of the EIN applications might not be for employer business start-ups. In fact, the data about the raw applications is quite noisy about the business formation. For example, out of the total number of business applications, we see that only 14% become employer businesses within two years from the date of the application. Even after considering the subset of the applications with higher rates of employer business births (Business Applications with Planned Wages, Business Applications from Corporations, High-propensity Business applications), the transition rate does not exceed 36%. Bayard et al. (2018) argue that a significant share of the business applications ends up becoming non-employer businesses.

[14]In Appendix A.3.3, I discuss the relevance of the information provided by the BFS for identification the option-to-delay channel. Figure 22 gives a diagram to illustrate the relationship between the BFS, the BDS and potential entrants in the model.

Table 3: Summary statistics

	Variable	Obs	Mean	Std. Dev.	Min	Max
Panel A. State-level	*Share of late start-ups*	2,142	0.13	0.03	0.02	0.26
	*Dur**F**4Q*	2,142	1.02	0.16	0.55	1.99
	*Dur**S**4Q*	2,142	5.43	0.20	4.83	6.21
	*Dur**F**8Q*	2,142	1.58	0.27	0.81	2.62
Panel B. US-level	*Share of late start-ups*	42	0.14	0.02	0.11	0.18
	*Dur**F**8Q*	42	1.66	0.17	1.37	1.96
	*Dur**S**4Q*	42	5.46	0.26	4.95	5.75
	*Dur**F**4Q*	42	1.06	0.09	0.88	1.21

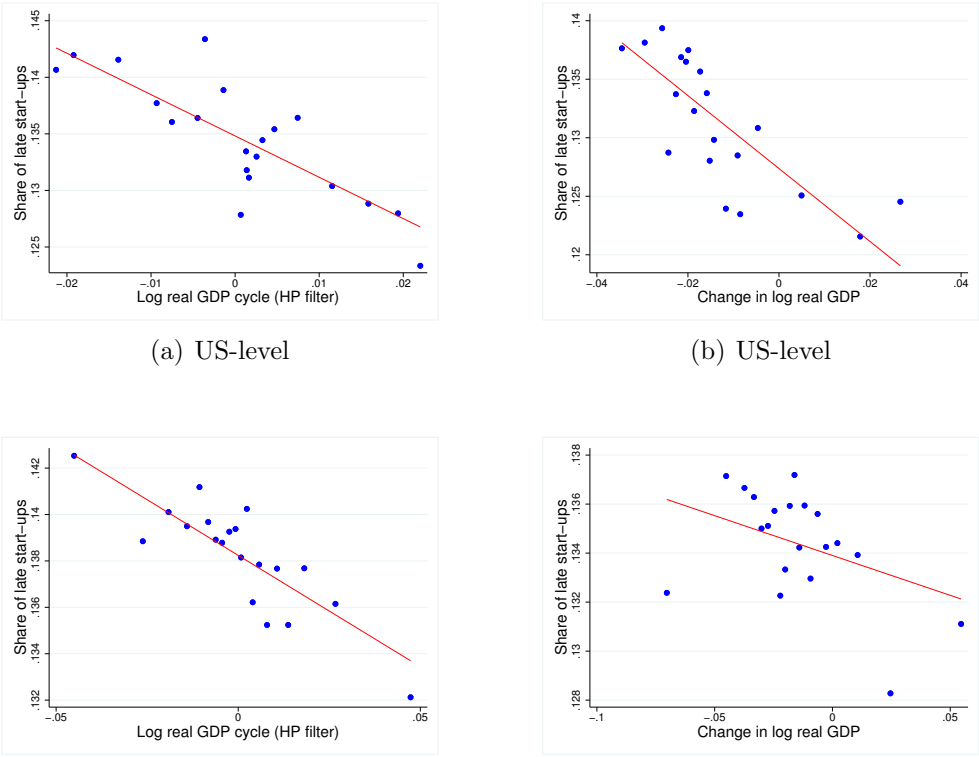
A), the average share of late start-ups equals 13%. That is, out of the total applications that form businesses in the first eight quarters, 13% start businesses with one-year delay. This variable varies from 2% to 26% across time and states, with the overall standard deviation around 3 percentage point. At the US level, the average share of late start-ups equals 14% and varies from 11% to 18% across time. The table also includes variables that describe average duration (in quarters) from a business application to formation conditional on business formation within the first four quarters ($Dur\boldsymbol{F}4Q$), eight quarters ($Dur\boldsymbol{F}8Q$), and the second four quarters ($Dur\boldsymbol{S}4Q$). These time series are quarterly and span the period $2004Q3$-$2015Q4$. Business formation among $\boldsymbol{F}4Q$ happens within the first two quarters. Similarly, business formation among the applications that become start-ups within the *second* four quarters happens between the fifth and sixth quarters from the quarter of the application. Panel B of Table 3 reports the same statistics for the aggregate data. Appendix A.3 provides a detailed description of the dataset.

To assess economic conditions at the time of the application, I use the following business cycle indicators: (1) The cyclical component of the quarterly log real GDP. To find the cyclical component of the yearly log real GDP I apply the HP filter with a smoothing parameter of *1600*. (2) Change in the log annual real GDP between t and $t+1$. I measure the latter as a change in the rolling sum of the consecutive four quarters starting from the quarter of the application.[15] The positive value of this variable indicates that the economic condition today is expected to be better compared to tomorrow. I construct both of the indicators at the state- and country-level.

Figure 3 illustrates the binned scatter plots of the share of late start-ups against the business

[15]For example, if the applications date is 2010Q3, I calculate annual GDP as $Y_{2010Q3}+Y_{2010Q4}+Y_{2011Q1}+Y_{2011Q2}$ and then calculate the difference as $log(Y_{2010Q3}+Y_{2010Q4}+Y_{2011Q1}+Y_{2011Q2})-log(Y_{2011Q3}+Y_{2011Q4}+Y_{2012Q1}+Y_{2012Q2})$.

Figure 3: The share of late start-ups against the aggregate economic conditions at the date of the application



(a) US-level

(b) US-level

(c) State-level

(d) State-level

Note: Each panel illustrates a binned scatterplot of the share of late start-ups against the aggregate conditions at the time of the application. Panels (a) and (b) display correlations at the US level. Panels (c) and (d) illustrate correlations at the state level. All of the plots control for linear and quadratic time trends and include quarter-fixed effects. In Panels (c) and (d), I also control for the state-level fixed effects.

cycle indicators at the time of applications. Panels (a) and (b) illustrate correlations at the US level, while Panels (c) and (d) display this relationship at the state level. The figures show that the share of late start-ups increases if the aggregate conditions at the time of the applications are below trend, measured by the HP-filtered real GDP. And, the share of late start-ups increases if the outlook for tomorrow is better, measured by the change in real GDP. These relationships hold at the country as well as state level. Based on this analysis, we can conclude that the share of the applications that form businesses with one year delay is negatively correlated with the economic conditions at the time of the application.

Next, I use the state-level variation in the share of late start-ups to further investigate the mechanism behind the relationship. I estimate the following regression

$$y_{s,t} = \alpha_0 + \beta Z_{s,t} + \alpha_1 Dur\mathbf{F}4Q + \alpha_2 Dur\mathbf{S}4Q + \alpha_3 \mathbf{F}8Q + \alpha_4 WBA + \gamma_s + \eta_q + \varepsilon_{s,t},$$

where $y_{s,t}$ describes the share of late start-ups in state s at time t. $Z_{s,t}$ describes business cycle conditions in state s at time t. Additionally, I include variables that could lead to the variation in the share of late start-ups that are not due to the waiting for better aggregate conditions. For example, obtaining credit to finance start-up activity might take more time during recessions, which could automatically increase the share of late start-ups. To account for the latter effect, I control the variation in the average duration from a business application to formation within the first ($Dur\boldsymbol{F}4Q$) and second ($Dur\boldsymbol{S}4Q$) four quarters. To control for the variation in the total number of business formation and applications, I include the total number of applications that become employer businesses within the first eight quarters ($\mathbf{F}8Q$). I also include the total number of wage-based business applications (WBA) to control for the variation in the composition of applications. The latter is a subset of business applications that indicate the intention of paying wages. Finally, γ_s and η_q control for the state- and quarter-fixed effects, respectively. That said, the coefficient β measures a percentage point change in the share of late start-ups due to the variation in the business cycle conditions at the time of the application, that are not due to changes in the average duration of the application and the variation in the total number of business applications.[16]

Table 4 reports the results of the regression equation (3). Panel A considers state-level variation in the share of late start-ups. Column (1) uses the state-level HP-filtered log real GDP as a business cycle indicator. The result shows that improving aggregate conditions at the date of the application decreases the share of late start-ups. Specifically, a 1 percentage point increase in the real GDP above the trend decreases the share of late start-ups by 0.063 percentage points. Column (2) considers the state-level change in the log real GDP as a business cycle indicator. The estimate implies that the improving aggregate economic conditions tomorrow relative to today has a statistically significant and positive effect on the share of late start-ups. Overall, Columns (1) and (2) support the original hypothesis that part of the business cycle variation in the start-ups is due to entrants' option to delay entry.

Finally, to check the robustness of these estimates, I consider the following exercises. Columns (3) and (4) of Panel A use business cycle conditions at the US level rather than the state level. Panel B of Table 4 runs the same regression using the US-level time series of the share of late start-ups. Again, we see that deteriorating aggregate conditions have a statistically significant and positive effect on the share of late start-ups. To conclude, the results show that part of the variation in the new business formation as a response to the changes in the

[16] In the appendix, I also consider a regression specification that includes interactions of the control variables with the business cycle indicators. The results of the coefficients are highly robust to the latter specification.

Table 4: The option to delay entry and business cycles

	Panel A. State-level share of late start-ups				**Panel B.** Country-level share of late start-ups	
$Z =$	$Y_{HP\ s,t}$	$\Delta Y_{s,t}$	$Y_{HP\ t}$	ΔY_t	$Y_{HP\ t}$	ΔY_t
	(1)	(2)	(3)	(4)	(1)	(2)
β	-0.062***	-0.068***	-0.187***	-0.154***	-0.326***	-0.165***
	(0.017)	(0.017)	(0.026)	(0.036)	(0.084)	(0.037)
$Dur\boldsymbol{F}4Q$	0.042***	-0.042***	0.037***	-0.027***	0.087***	-0.071**
	(0.006)	(0.005)	(0.006)	(0.006)	(0.025)	(0.027)
$Dur\boldsymbol{S}4Q$	-0.062***	0.020	-0.062***	0.010	-0.201**	0.086
	(0.008)	(0.016)	(0.007)	(0.016)	(0.096)	(0.125)
$\boldsymbol{F}8Q$	0.006*	0.016***	0.004	0.001	0.017	0.013
	(0.003)	(0.003)	(0.004)	(0.001)	(0.013)	(0.009)
WBA	-0.021***	0.011***	-0.014***	0.014***	-0.013	0.021***
	(0.001)	(0.005)	(0.003)	(0.004)	(0.010)	(0.007)
State FE	✓	✓	✓	✓	✓	✓
Quarter FE	✓	✓	✓	✓	✓	✓
Observations	2,040	2,040	2 040	2,040	39	35
R-squared	0.815	0.699	0.818	0.720	0.967	0.757

Note. Robust standard errors are in parenthesis. In Panel A the robust standard errors are clustered at state-level. The table reports results from a linear regression with a dependent variable the *share of late start-ups*. *** significance at 0.01 level, ** significance at 0.05 level, * significance at 0.10 level.

aggregate conditions at entry is due to firms' option to delay entry.

3 The Model

The model builds on Moreira (2016), which features endogenous firm entry and exit in the style of Hopenhayn (1992). The exogenous aggregate demand shock that affects firms' profitability and selection of entrants is the model's only source of business cycles. The economy consists of incumbent firms and potential entrants. Incumbent firms produce differentiated products and are heterogeneous over idiosyncratic productivity and customer capital. They make decisions about production and exit. Potential entrants hold heterogeneous signals about their initial post-entry productivity. I deviate from the original framework and allow potential entrants to keep the signals over time until they enter the market. The modification gives potential entrants the option to delay entry after observing the aggregate state. A detailed description of the framework is given below.

3.1 Firms

Technology At the beginning of each period, a positive measure of heterogeneous firms produce differentiated products on a monopolistically competitive market using the following production function:

$$y_i = s_i n_i.$$

The production function is linear in labor n_i. Labor supply is infinitely elastic. Wage is exogenous and constant. s_i is a time-varying idiosyncratic productivity specific to a firm i and evolves according to a persistent $AR(1)$ process:

$$log(s_i') = \rho_s log(s_i) + \sigma_s \varepsilon_i,$$

where $\varepsilon_i \sim i.i.d.\ N(0,1)$. Idiosyncratic productivity is distributed independently across firms. Every period, firms that are operating in the market incur fixed cost $c_f > 0$, drawn from a time-invariant log normal distribution $c_f \sim G(c_f)$ with mean μ_f and standard deviation σ_f. The fixed cost is distributed independently across firms.

Demand In each period, demand for firm i's differentiated good is determined according to the following demand function

$$y_i = p_i^{-\rho} b_i^{\eta} \alpha z,$$

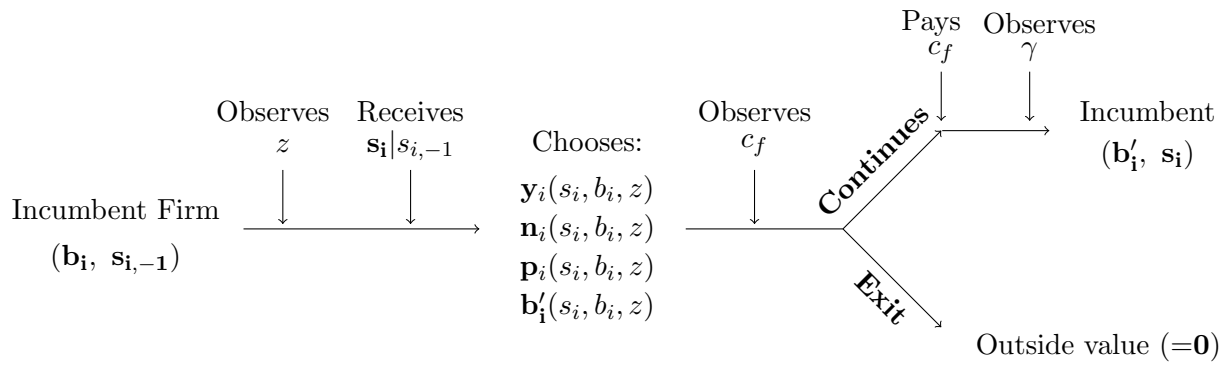
where p_i is the price set by firm i, and $\rho > 1$ is the price elasticity of demand. $\eta \in (0,1)$ measures the elasticity of demand with respect to customer capital b_i, which evolves according to:

$$b_i' = \begin{cases} (1-\delta)b_i + (1-\delta)p_i y_i & \text{incumbent firm } i \\ b_0 & \text{entrant firm,} \end{cases}$$

where b_0 is the initial level of customer capital, common across all entrants. $\delta \in (0,1)$ is the depreciation rate of customer capital. The process of customer capital that is tied to past sales hinders firms' ability to freely adjust their demand over time, which creates persistence in the dynamics of production and employment.[17] z represents a common aggregate demand

[17] The channel is motivated by the growing literature that emphasizes the demand-side factors in understanding the firm-level and aggregate dynamics. For example, Foster et al. (2016) find that the differences between young and mature firms are due to individual demand dynamics rather than differences in productivity. Sedlacek and Sterk (2017), and Moreira (2016) emphasize the demand-side factors in accounting for thee persistent procyclical variation in cohorts-level employment.

Figure 4: Incumbent firm's timing



shock that evolves as a persistent $AR(1)$ process,

$$log(z') = \rho_z log(z) + \sigma_z \epsilon,$$

where $\epsilon \sim i.i.d.N(0,1)$. $\alpha > 0$ is a scale factor.

3.1.1 Incumbent Firms

At the beginning of each period, an incumbent firm i, with predetermined customer capital b_i, observes aggregate demand shock z, and idiosyncratic productivity s_i. Using the information, the incumbent firm makes decisions about the optimal production level, price, and the next period's customer capital. At the end of the period, the incumbent firm draws fixed cost c_f and makes the continuation decision. Even if the firm decides to stay in the market, it may be hit by a random exit shock with probability $\gamma \in (0,1)$. The outside value is normalized to zero.[18] Firms discount future profits at the time-invariant factor β.

The incumbent firm solves the following functional equation:

$$V^I(b,s,z) = \max_{y,p,b'} \quad \left(p - \frac{w}{s}\right) y + \int max\left\{0, -c_f + \beta(1-\gamma)E[V^I(b',s',z')|s,z]\right\} dG(c_f),$$

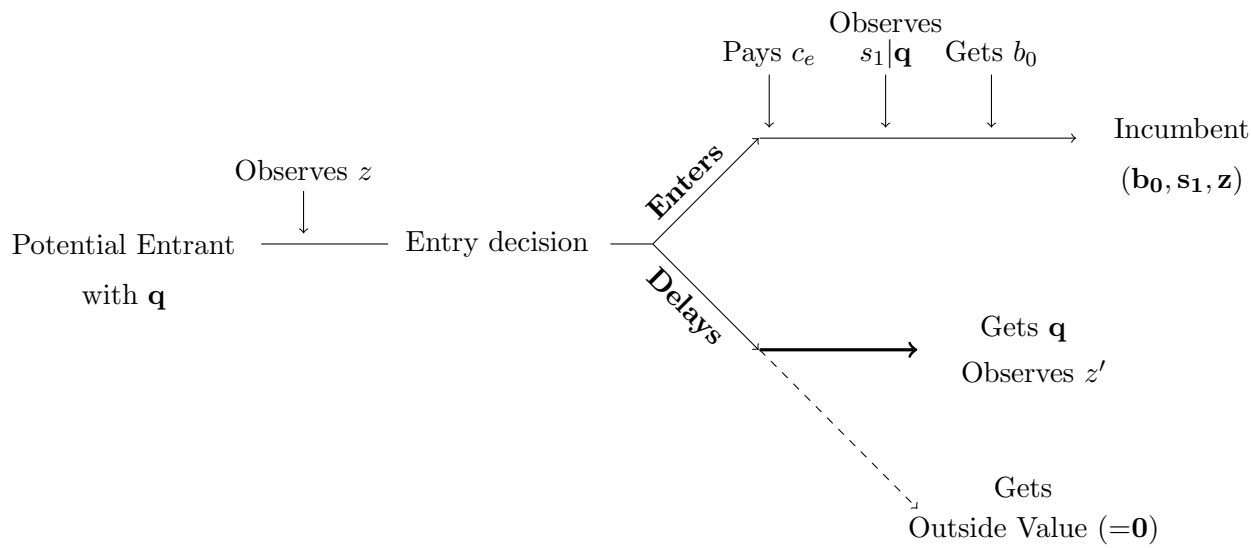
$$\text{s.t.} \quad b' = (1-\delta)(b+py),$$

$$y = \alpha p^{-\rho} b^{\eta} z.$$

The summary of the incumbent firm's timing is illustrated in Figure 4.

[18] I assume that if the incumbent firm decides to exit, the probability that the firm receives an initial productivity signal and becomes a potential entrant again is zero.

Figure 5: Potential entrant's timing



3.1.2 Potential Entrants

At the beginning of every period, there is a constant mass of potential entrants M. Potential entrants are endowed with heterogeneous signals q about their first-period idiosyncratic productivity. For a given signal, the idiosyncratic shock in the first period of operation is normally distributed and follows the process $log(s) = \rho_s \; log(q) + \sigma_s \; \epsilon$, where $\epsilon \sim N(0, 1)$.[19]

The aggregate distribution of potential entrants over signals is time invariant and is given by the Pareto distribution $W(q)$ with location parameter $\underline{q}$ and Pareto exponent $\xi > 0$.[20] The potential entrant's timing is described below and is summarized in Figure 31.

At the beginning of every period, each potential entrant with a signal q observes an aggregate state of the economy z and makes an entry decision. A firm can either enter the market today or wait until tomorrow. Entry into the market is subject to a fixed entry cost c_e.

[19]The ex-ante heterogeneity of potential entrants is crucial for the option-value channel. With ex-ante homogeneous potential entrants, the interior entry solution requires the option value of delay to equal zero. For example, see paper Bilbiie, Ghironi, and Melitz (2012).

[20]Underling the restriction is an assumption that the number of business ideas that can be implemented in the market in each period is limited. This assumption is used throughout the literature (e.g, see Sedlacek and Sterk (2017), Sedlaćek (2020), Lee and Mukoyama (2018)). Fajgelbaum, Schaal, and Taschereau-Dumouche (2017) assume a constant mass of entrants in a model where firms make decisions between entry and waiting. In Appendix B.1, I extend the entry phase that justifies the constant mass of potential in this framework. In Appendix B.2, I show that the main results of the paper are robust if I extend the model and allow the accumulation of potential entrants over time.

Entrant solves the following Bellman equation

$$V^e(q,z) = \max\{\ V^w(q,z), \quad -c_e + V^{gross}(q,z)\}\,, \tag{3}$$

where V^{gross} is the value of entering after paying the entry cost c_e and $V^w(q,z)$ is the value of waiting.

If a firm decides to enter the market today, the firm observes actual idiosyncratic productivity (s), receives the initial customer capital stock (b_0), and behaves like an incumbent with state variables (b_0, s, z). Therefore, the value of entry today is

$$V^{gross}(q,z) = \int_s V^I(b_0,s,z)dH_e(s|q).$$

If the firm waits, it starts the next period with the same signal q, but observes a new aggregate demand level z'. Therefore, the value of waiting is

$$V^w(q,z) = \beta \int_{z'} V^e(q,z')dF_z(z'|z).$$

3.2 Recursive Competitive Equilibrium

Denote the distribution of incumbent firms across productivity and customer capital by $\Omega(s,b)$. Then, at the beginning of every period, the vector of the aggregate state variables is given by $\Gamma = \{\ z,\ \Omega(b,s),\ W(q)\}$. For a given Γ_0, a recursive equilibrium consists of the following: (i) value functions $V^I(b,s,z)$, $V^e(q,z)$; (ii) policy functions $y(b,s,z)$, $p(b,s,z)$, $n(b,s,z)$, and $b'(b,s,z)$; and (iii) distribution of operating firms $\{\Omega_t\}_{t=1}^{\infty}$, such that

1. $V^I(b,s,z)$, $y(b,s,z)$, $p(b,s,z)$, $n(b,s,z)$ and $b'(b,s,z)$ solves incumbent's problem; and
2. $V^e(q,z)$ solves the entrant's problem.

4 Inspecting the Mechanism

The goal of the section is twofold. First, in Section 4.1, I study the optimal timing of market entry decisions made by heterogeneous potential entrants. Then, in Section 4.2 I illustrate how the option modifies the selection of entrants across different aggregate conditions.

4.1 Entry Timing and the Value of Waiting

To understand how the option to wait alters firms' entry decisions, consider the following modification of 5 equation (3)

$$V^e(q,z) = \max\{\ dV^w(q,z), \quad -c_e + V^{gross}(z,q)\ \}, \tag{4}$$

where d describes a dummy variable that takes value one if potential entrants have the option to delay entry. If $d = 0$, the option value equals 0, which happens when the initial productivity signals are 'use it or lose it' type – an entrant loses the signal if he postpones market entry decision until tomorrow. In this case, the model reduces to a standard framework where firms enter the market if the expected value of entry net of the fixed entry cost is more than 0. If $d = 1$, entrants problem coincides with the baseline model. The decision to become an incumbent today is an irreversible choice as the potential entrant gives up the option to exercise the signal in the future. Hence, with $d = 1$, the option value of delay creates an opportunity cost that must be added to the direct fixed cost of entry while making an entry decision. Result 4.2 uses the numerical methods to summarize the properties of the option value of delay, $V^w(z,q)$.

Result 4.1. *(i) $V^w(q,z)$ is non-negative for all q and z; (ii) For a given aggregate demand level z, $V^w(q,z)$ is a weakly increasing function of the signal q; and (iii) For a given signal q, $V^w(q,z)$ weakly increases with the aggregate demand level z.*

Solving the entry timing problem for a potential firm with signal q consists of finding a 'trigger' or a threshold aggregate demand level $\underline{z}^d(q)$ for which the firm enters the market if $z \geq \underline{z}^d(q)$. Using the numerical solutions, Result 4.2 formally characterizes the optimal entry rule and the threshold aggregate demand level across heterogeneous firms.

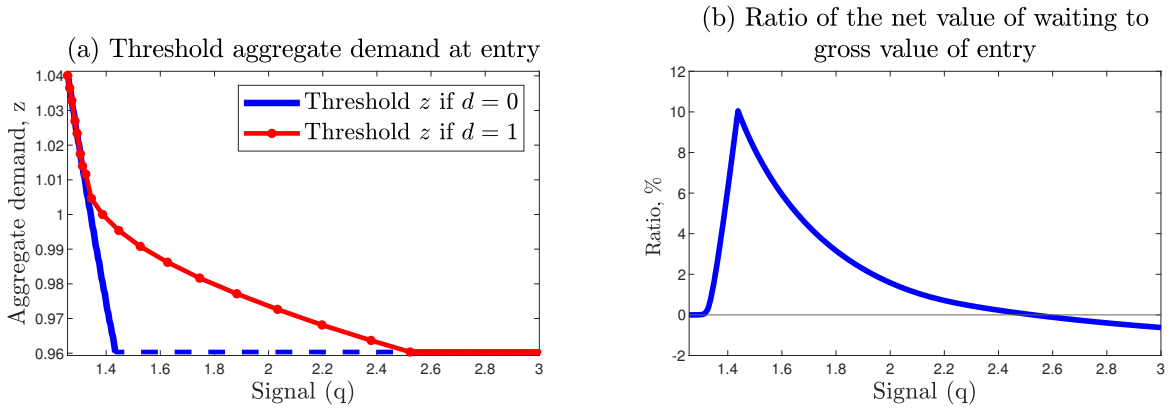
Result 4.2. *Suppose for a signal level q, exists an aggregate demand level $\underline{z}^d(q)$ such that*

$$V^{gross}\left(\underline{z}^d(q), q\right) - c_e = dV^w(\underline{z}^d(q), q);$$

Then, a potential entrant with signal q decides to enter the market for all $z > \underline{z}^d(q)$, otherwise chooses to stay outside the market.

Figure 6(a) compares the threshold aggregate demand levels $\underline{z}^d(q)$ with and without the option to delay entry across signals. The red-circled line shows the threshold aggregate demand levels for the baseline model, whereas the solid-blue line displays the optimal entry decisions assuming that firms do not have the option to wait ($d = 0$ case). The blue-dash line

Figure 6: Entry decision and the value of waiting



indicates scenarios for which $\underline{z}^{d=0}(q) < z_{min}$, where z_{min} represents the minimum grid point for the aggregate demand level identified by the numerical solution. The figure shows that when the option is available, the threshold aggregate demand is weakly higher for each signal level. Specifically, the high- and low-productivity entrants decision to enter the market does not change with the option to delay entry, whereas firms with medium-range productivity signals find it strictly profitable to wait for higher aggregate demand levels. Interestingly, even the potential firms whose net present value of entry is more than zero for all reasonable aggregate demand levels – illustrated by the blue-dashed line in Figure 6(a) – choose to wait if the option is available.

To quantify the option to wait, I calculate the net present value of benefits that firms give up by ignoring the option and following the traditional investment decision rule. I define the *net value of waiting* as the difference between the net present value of entering the market today and the net present value of keeping the option to enter 'alive'. That is,

$$\text{Net value of waiting}(q, z) = V^{w}(q, z) - \left[V^{gross}(q, z) - c_e\right].$$

By definition, when the aggregate demand level is $\underline{z}^{d=0}(q)$, the net present value of entry today equals zero for a potential entrant with signal q, and the net value of waiting equals to the option value of delay $V^{w}(q, \underline{z}^{d=0})$. Figure 6(b) illustrates this value for each signal level relative to the value of entry. The figure shows that the net value of waiting for medium productivity firms can be as high as 10 percent of the present value of life time benefits. Conversely, the figure shows that high-productivity firms lose profits if they postpone entering the market. Finally, waiting has zero value for entrants with low productivity signals.

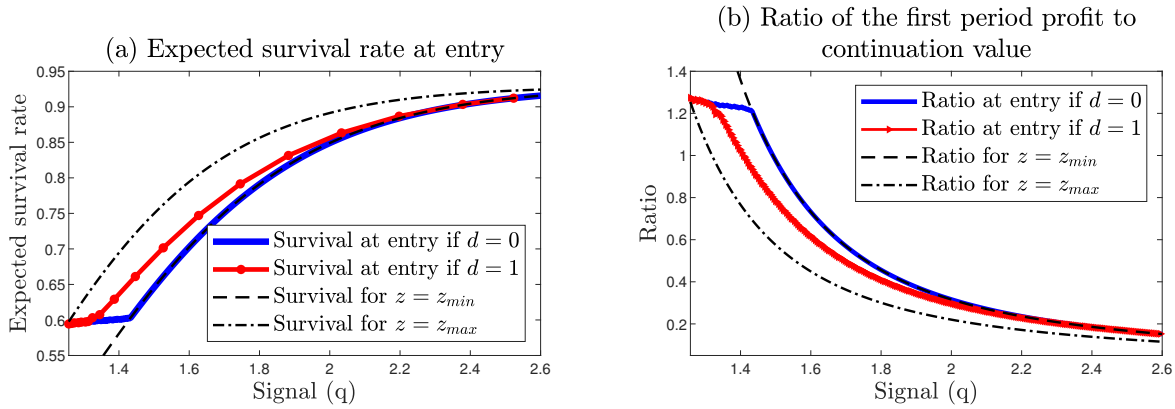
$$
\begin{aligned}
&V^{\text{gross}}(b_o, q, z) = \\
&\quad = \int_s \left(\Pi(b_o, s, z) + \int_{c_f} max\left\{0, -c_f + \beta(1-\gamma)E[V^I(b', s', z')|s, z]\right\} dG(c_f) \right) dH_e(s|q) \\
&\quad = \underbrace{\int_s \Pi(b_o, s, z) \;\; dH_e(s|q)}_{\textit{Expected first period profit}} + \qquad (5) \\
&\quad + \underbrace{\int_s \beta \underbrace{(1-\gamma)G\left(c_f^*\right)}_{\textit{Survival rate}} \left[E(V^I(b', s', z')|s, z) - \frac{1}{(1-\gamma)\beta} E(c_f \mid c_f \leq c_f^*) \right] \; dH_e(s|q)}_{\textit{Expected continuation value}},
\end{aligned}
$$

where $c_f^* = \beta(1-\gamma)E[V^I(b', s', z')|s, z]$.

Next, I investigate what contributes to the heterogeneous option-value effect. Consider Equation (5), which decomposes potential entrants' gross value of entry into the expected first-period profit and continuation value. The latter value depends on the probability that a potential entrant stays in the market after the first period, described by the expression $(1-\gamma)G(c_f^*)$. Figure 7(a) illustrates the expected survival rates for the expansionary (black-dotted line) and recessionary (black-dashed line) periods. The figure shows that the lower aggregate demand level at entry is associated with the lower expected survival rates for entrants, which is due to the combination of the persistent aggregate demand shock process and the persistent customer capital accumulation. Thus, during the recessions, alongside lower first-period profits, firms' also lose part of their long-run benefits due to the increased probability of irreversible exit. The option to choose the time of entry allows entrants to endogenize the latter cost: Firms can stay outside the market until the expected survival rates are high enough to compensate for lower demand levels in the first several years of operation.

An increase in the entrant's productivity signal has two counteracting effects on the value of wait: The weight of the continuation value, hence, the cost of prematurely exiting the market, rises with the signal level. However, at the same time, the procyclical variation in the firms' post-entry failure rates, hence, the benefits of waiting, decreases with the signal level; As a result, a trade-off between the first-period profit and long-run value leads to a positive value of waiting for some but not all potential entrants.

Figure 7: The risk of post-entry failure and the option to wait



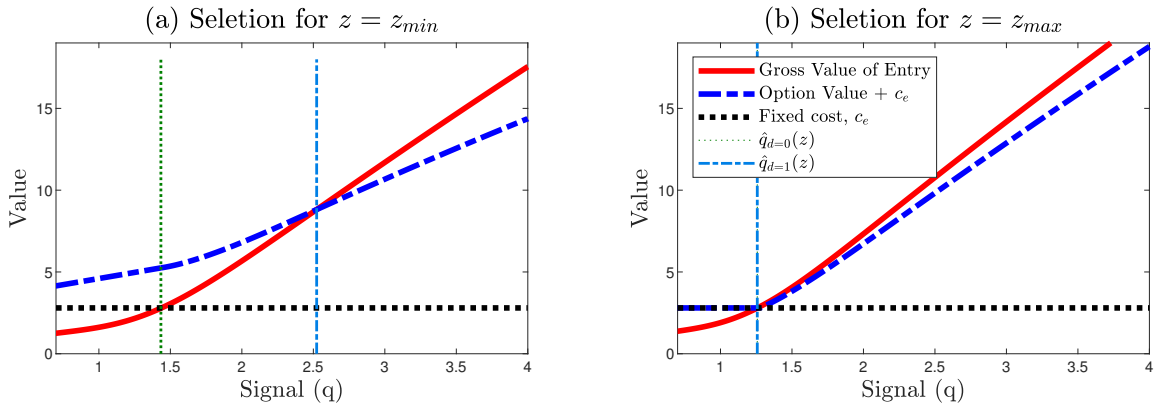
I use the numerical solution to graphically illustrate the trade-off. In Figure 7, the blue-solid lines display the expected survival rates and the ratio of the expected first-period profit to the continuation value at the optimal threshold aggregate demand levels without the option to wait, $\underline{z}^{d=0}(q)$. In contrast, the red-circle lines display the same values over the threshold aggregate demand levels that account for the option to wait, $\underline{z}^{d=1}(q)$. The figure illustrates that medium-productivity firms delay entering the market and wait for aggregate conditions that promise higher expected survival rates and higher continuation value. Entrants with low-productivity signals, on average, have higher risks of post-entry failure, and the first period profits represent a significant share of their entry value. They form businesses during the highest aggregate demand levels when the waiting has zero value. Finally, high-productivity entrants' failure rates do not vary with the aggregate conditions. By postponing entry, they lose more in terms of the short-run profits than gain by an increase in the long-run value – leading to the negative net value of waiting.

Keeping the mechanism in mind, in Section 5.3, I show that the option to wait is a qualitatively important channel to generate a selection of entrants over the business cycles with the countercyclical survival rates – an empirical fact documented in Section 2. Without the option, the optimal entry decision leaves the expected survival rates almost unchanged across $\underline{z}^{d=0}(q) < z_{min}$ (illustrated by the flat segment of the blue-solid line in Figure 7a).

4.2 Aggregate Selection of Entrants

After having discussed the mechanism of how the option to wait modifies entry decisions at the micro-level, I study macro-level implications of the option. That is, I show how the modified cost of entry due to the option to wait affects the elasticity and selection of entrants

Figure 8: Mechanism: selection of entrants



across different aggregate demand levels.

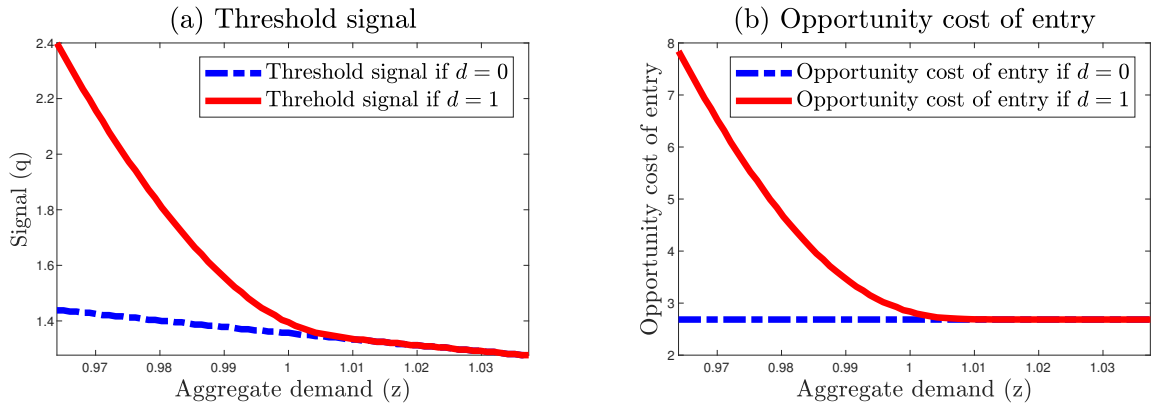
In the model, all potential firms get the same level of customer capital b_0 and observe the same aggregate demand level z at entry. Therefore, we can characterize the selection of firms for each aggregate demand level based only on a signal level q. I use the numerical solution to graphically illustrate the result. Figure 8(a) displays the gross value of entry, the fixed entry cost, and the option value of delay across the signal for an aggregate demand level z. In the baseline model, firms decide to start operating today if the expected benefits cover the total cost of entry – sum of the fixed cost and the option value of delay. In Figure 8(a), these are the firms who hold signals $q \geq \hat{q}_{d=1}(z)$; the rest stays outside the market. I refer to $\hat{q}_{d=1}(z)$ as a threshold signal for an aggregate demand level z for $d = 1$ case.

If firms do not have the option to wait, they enter the market if the expected post-entry benefits cover the fixed entry cost. In Figure 8(a), these are the firm who hold signals $q \geq \hat{q}_{d=0}(z)$; again, the rest stay outside the market. Comparing the threshold signal levels across these cases identifies the selection only through the option-to-delay channel. Specifically, for an aggregate demand level z, the option generates a new group of firms with signals $q \in [\hat{q}_{d=0}(z),\ \hat{q}_{d=1}(z)]$ who decide to stay outside the market despite the positive net expected benefits of entry. Figure 8(b) shows that for the highest aggregate demand levels, the group of potential entrants that decide to enter the market is same with or without the option to delay entry: during the peak, nobody finds it optimal to delay entry. Result 4.3 formally summarizes the threshold signal levels for the rest of the aggregate states using the numerical solution methods:

Result 4.3. *Suppose for an aggregate demand level z, exists a signal $\hat{q}_d(z)$ such that*

$$V^{gross}(z,\ \hat{q}_d(z)) - c_e = dV^w(z, \hat{q}_d(z));$$

Figure 9: Selection of entrants across different aggregate conditions



Then, all potential entrants with $q \geq \hat{q}_d(z)$ decide to enter the market, whereas the rest stays outside the market.

Figure 9(a) shows that the threshold signal is countercyclical: the group of firms that enter the market during recessions hold a relatively higher range of signals than the group of firms that enter during expansions. For each aggregate demand level, the group of firms that decide to use the option to delay entry despite the positive expected post-entry profits are the ones that hold signals from the following range: $q \in [\hat{q}_{d=0}(z),\ \hat{q}_{d=1}(z)]$. Note that the lower the aggregate demand level, the wider the range of signals that leads to the delay decisions.

The increased elasticity of the threshold signal with respect to aggregate demand level is due to the countercyclical opportunity cost of entry endogenously generated through the option-to-delay entry channel. I define the equilibrium opportunity cost of entry as the minimum level of benefits that potential entrants require to enter the market across various aggregate conditions. For example, in the $d = 0$ case, the value is constant across the aggregate demand levels and equals the fixed entry cost. In the baseline model, the value is endogenous due to the option to delay entry, and it coincides with the threshold signal's $\hat{q}_{d=1}(z)$ opportunity cost of entry.[21] Figure 9(b) illustrates that the latter value is countercyclical: the cost of entry significantly increases above the fixed entry cost during the recessions.[22] In fact, for reasonable parameter values, firms that enter the market during recessions are the ones that

[21]Potential entrants with signal $q > \hat{q}_{d=1}(z)$ enter the market and expect returns that are higher than the threshold signal's $\hat{q}_{d=1}(z)$ total total cost of entry. *Proof:* $V^{gross}(q, z)$ strictly increases with the signal. For an aggregate demand level z, firms with $q > \hat{q}_{d=1}(z)$ enter the market. The following inequality holds: $V^{gross}(z, q) > V^{gross}(z, \hat{q}_{d=1}(z)) = c_e + V^w(z, \hat{q}_{d=1}(z))$.

[22]In Appandix F.1, I investigate how the value of delay changes if $d \in (0, 1)$. Figure 32 shows that the total opportunity cost of entry, as well as, the threshold signal level significantly increases with d.

expect the present value of entry up to twice the fixed entry cost.

The latter result is the core finding of the paper – the option to delay entry endogenously generates the countercyclical cost of entry in equilibrium that amplifies the elasticity of entrants with respect to the aggregate conditions compared to a case with the fixed entry cost. The higher elasticity of the threshold signal, in turn, leads to the variation in the number and the composition of entrants over the cycles. That is, during recessions, an increased threshold signal leads to fewer but higher-productivity entrants compared with expansionary cohorts. Later in the paper, I use the calibrated model to quantify the option-value channel in reconciling the observed dynamics of entrants over the business cycles.

5 Calibration and Model Performance

In this section, I explain the model's calibration procedure and the strategy I use to discipline the option-value channel. After I show that the model closely replicates the main features of the US firm dynamics at entry and over time, I investigate the business cycle dynamics of cohorts in the model and the data. Consistent with the empirical facts, I find that aggregate conditions have a significant and persistent effect on firms' life cycle dynamics – cohorts born during recessions are, on average, more productive and have higher survival rates, although they employ fewer workers at entry and over time. Finally, I show that these differences build up significant persistence and variance in aggregate variables.

5.1 Calibration

A period in the model corresponds to one year, consistent with the timing of the BDS dataset. The unit of analysis is an establishment. Estimating the model requires calibrating 17 parameters. First, I describe the parameters that I choose based on the estimations in the literature. Then, I jointly calibrate the rest of the parameters to match the cohorts' average life cycle characteristics. The summary of the parameters, identification strategy, and the final values of the parameters are given in Table 5.

I set the time-preference parameter $\beta = 0.96$ to match a 4% percent annualized average riskless interest rate. In the baseline model, the production function, demand function, and the process of the customer capital accumulation follows the specification developed and estimated in Foster et al. (2008), and Foster et al. (2016). Using the establishment-level data from the Census of Manufactures, Foster et al. (2008) estimates that the autocorrelation

Table 5: Calibration

Symbol	Description	Value	Targets/Source
β	Discount rate	0.960	Riskless interest rate
ρ	Price elasticity of demand	1.622	Foster et al. (2016)
η	Elasticity of demand to capital	0.919	Foster et al. (2016)
δ	Depreciation rate of reputation	0.188	Foster et al. (2016)
ρ_s	Idiosyncratic shock – persistence parameter	0.814	Foster et al. (2008)
σ_s	Idiosyncratic shock – SD parameter	0.161	Firm size by age
α	Demand shifter	0.261	Firm size by age
b_o	Initial customer capital level	12.00	Firm size
μ_f	Operating cost – SD parameter	0.621	Firm survival by age
σ_f	Operating cost – SD parameter	0.410	Firm survival by age
γ	Exogenous exit shock	0.071	Firm exit hazard by age
$\underline{q}$	Pareto location	0.700	Firm size at entry
ξ	Pareto exponent	2.478	Employment share at entry
c_e	Fixed entry cost	2.684	Entry rate – mean
ρ_z	Aggregate shock – persistence parameter	0.750	Entry rate – persistence
σ_z	Aggregate shock – SD parameter	0.003	Entry rate – SD

of the establishments' idiosyncratic productivity process equals $\rho_s = 0.814$.[23] Foster et al. (2016) identifies parameters that drive the demand function and the customer-capital-accumulation process by jointly estimating the demand and the Euler equation, using the dataset from Foster et al. (2008). Based on their estimates, I set the price elasticity of demand (ρ) equal to 1.622, the elasticity of demand to customer capital (η) equal to 0.919, and the depreciation rate (δ) equal to 0.188.

I formally calibrate the rest of the parameters $\sigma_s, b_o, \alpha, \mu_f, \sigma_f, \gamma, \underline{q}, \xi, c_e, \rho_z, \sigma_z$ using the minimum distance estimation procedure proposed by Chamberline (1994). That is, I minimize the sum of squared deviations of the eleven moments that characterize firms' life cycle dynamics in the model from its data counterpart. To compute the relevant statistics, I use annual time series about the US-level cohorts of establishments from the BDS dataset covering the period 1977-2015. I choose the following moments to capture cohorts' average characteristics at entry: entry rate, the employment share of entrants', the average size at

[23] Technology in Foster et al. (2008) is linear in inputs and productivity: $q_i = s_i x_i$ where x_i is the input and s_i is producer-specific productivity. Foster et al. (2008) uses establishment-level data of eleven manufacturing products. The data provide detailed information about producer-level quantities and prices for the following census years: 1977, 1982, 1987, 1992, and 1997. Using the dataset, they are able to directly measure total physical factor productivity, defined as $TFPQ_i = \frac{s_i x_i}{x_i} = s_i$. Autoregressive properties of the measured TFPQ imply persistence rate $\rho_s = 0.814$. Foster et al. (2008) finds that persistence of TFPQ is very close to the persistence parameters generated from other measures of total factor productivity (TFP) (e.g., traditional measure of TFP and revenue TFP).

Table 6: Calibration targets and the model-implied counterparts

Statistics	**Data**	**Model**
Firm size	17.0	17.0
Firm size at entry	8.73	10.0
Firm size at age 5	13.9	14.5
Firm size at age 23	21.2	22.1
Employment share at entry	0.56	0.58
Firm exit hazard at age 5	0.10	0.09
Firm survival rate up to age 5	0.49	0.49
Firm survival rate up to age 23	0.15	0.12
Entry rate (%)	9.90	10.0

Note: The moments are calculated using the US-level cohorts of establishments from the BDS dataset covering the the period 1978-2019.

Table 7: Calibration targets for the aggregate demand shock process

Statistics	**Data**	**Model**
Autocorrelation of the cyclical component of establishments	0.70	0.72
SD of the cyclical component of establishments	0.01	0.01
SD of the cyclical component of entry	0.06	0.06

Note: The time series about the entry rate comes from the BDS and covers the period 1978-2019. The cyclical component of the log entry rate is calculated using the HP filter with smoothing parameter 100.

entry, the average size at entry relative to the size of all active establishments.[24] To capture cohorts' post-entry growth, survival, and exit, I target average cohorts' size and survival rate at age 5 and at age 23. Table 6 summarizes the targeted moments and their corresponding values in the data and the model. The model-simulated moments are calculated in the the stochastic steady state.

Although these parameters are jointly estimated, below, I describe how the targeted sample moments help us infer about each of these parameters. The standard deviation of the idiosyncratic productivity shock (σ_s) shapes cohorts' growth rate. The demand parameter (α) affects the scale of the economy. Thus, in the calibration, these two parameters mainly target cohorts' average size at age 5 and age 23. Finally, the exogenous exit probability (γ), alongside the mean (μ_f) and standard deviation (σ_f) of the fixed operating cost, shapes cohorts' life cycle survival and exit rates. Therefore these parameters are estimated using average cohorts' survival rates at age 5 and age 23 and exit hazard rate at age 5.

[24] size is defined as the total employment number by entrants/incumbents/all establishments over the total number of entrants/incumbents/all establishments.

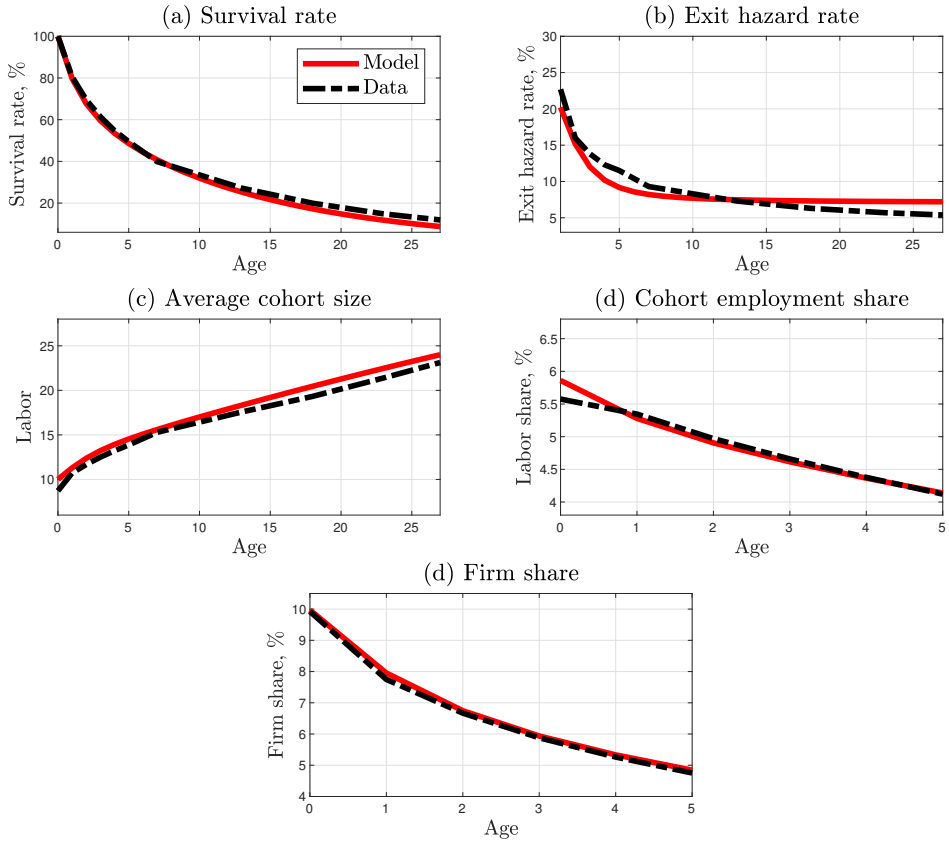
The entry cost (c_e) determines the steady-state mass of entrants, while parameters $\underline{q}$ and ξ shape the potential entrants' distribution over the productivity. I estimate these parameters by targeting the average entry rate, the share of entrants' employment in total employment, and the average size of entrants. The initial level of customer capital (b_0) is calibrated to match the relative size of entrants compared to the average size of all active firms. Note that the parameters $\underline{q}$ and ξ also affect the variation in the number of entrants across aggregate demand levels, as they shape the distribution of entrants across signals. That said, I will partially use these parameters to discipline the business cycle variation in the number of entrants, which I will explain next.

In the model, the key parameters that drive the business cycle firm dynamics are the persistence (ρ_z) and standard deviation (σ_z) of the aggregate demand shock process. The exogenous shock affects incumbent firms' production and exit decisions and, at the same time, determines the business cycle variation in the number and composition of entrants. The aggregate demand affects entry decisions due to the following two reasons. First, it directly determines the NPV of entry through its effect on incumbent firms. And second, as we saw in Section 4.2, the shock also endogenously alters the cost of entry through to the option-value channel. The latter mechanism partially breaks the link between the dynamics of entrants and the aggregate demand shock process and allows me to discipline the option-value channel. That is, I jointly calibrate these parameters to match the autocovariance and the variance of the total number of firms and the variance of the entry rate in the model and the data. To construct the cyclical component of the entry rate and total number of establishments, I apply the HP filter with a smoothing parameter of 100. To calculate the same moments in the model, I simulate the economy over many periods and apply the same detrending method to the model-simulated total number of firms and entry rate. The autocovariance and standard deviation of the time series are reported in the second and third columns of Table 7. The final values of the parameters that generate the match are $\rho_z = 0.75$, and $\sigma_z = 0.003$.

5.2 Cohorts' Average Life Cycle Characteristics

Columns (2) and (3) of Table 6 lists the data moments and their model-implied counterparts, respectively. The model successfully replicates the main features of the US firm dynamics. The average firm employs 17 workers in the data and 16.9 workers in the model. Entrants contribute only around 5.6 percent to total employment in the model and the data. The

Figure 10: Cohorts' average characteristics: Data, model



Note. The data moments are calculated from the US-level annual Business Dynamics Statistics dataset over the period 1978-2019. The unit of analysis is a cohort of establishments by age. The model-simulated moments are calculated at the stochastic steady state.

model does a good job in reproducing the well-known 'up or out' dynamics of firms. About 50% of the entrants fail within the first five years, and by age 23, only around 10% out of original start-ups survive. At the same time, cohorts of firms grow from 9.6 workers at entry to 22 workers by age 23.

Figure 10 goes beyond the moments reported in Table 6 and illustrates the full life cycle profile of firms – moments and statistics not directly targeted in the calibration. Panel (a) show that the model closely replicates the survival rates of firms up to age 30. Panel (b) shows that the model also successfully matches the dynamics of the exit hazard for up to age 30. Panels (c) and (d) further describe growth of cohorts measured by average size and the share of cohorts' employment in total employment. Finally, Panel (e) describes the share of firms by age in the total number of firms. Overall, Figure 10 shows that the model quite closely reproduces average cohorts of establishments life cycle dynamics in the US.

Table 8: Differences in cohorts' characteristics based on the initial conditions: Data

	Age 0	Age 1	Age 2	Age 3	Age 4	Age 5
Firms	2.38***	2.24***	2.01***	1.80***	1.81***	1.87***
	(0.17)	(0.15)	(0.13)	(0.11)	(0.10)	(0.09)
Survival rate		-0.22***	-0.41***	-0.49***	-0.44***	-0.39***
		(0.02)	(0.04)	(0.05)	(0.05)	(0.06)
Employment	3.01***	3.01***	2.23***	1.64***	1.66***	1.95***
	(0.23)	(0.18)	(0.16)	(0.14)	(0.13)	(0.14)
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes

Note. The estimates use the state-level annual BDS dataset over the period 1978-2015. To be consistent with the model calibration, I consider a cohort of establishments as a unit of analysis. I use the cyclical component of the HP-filtered log real GDP to measure the aggregate economic conditions at entry.

5.3 Entry conditions and Persistent Cohort Dynamics

Next, I show that the calibrated model successfully accounts for the documented persistent and significant differences in the average life cycle characteristics of cohorts born at different stages of business cycles.

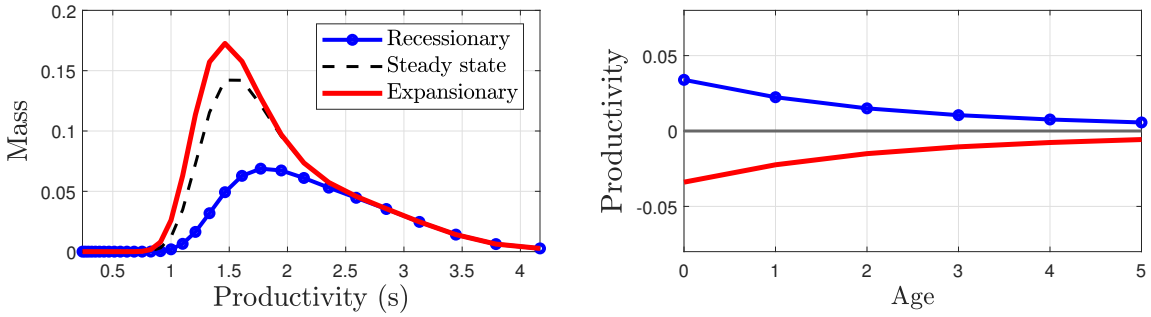
To investigate the effect of the initial aggregate conditions on cohorts' life cycle characteristics, I estimate the following regression equation in the data and the model:

$$X_{g,t} = \alpha + \sum_{g=0}^{T} \beta_g D_g Z_c + \eta_g + \varepsilon_{g,t}, \tag{6}$$

where $X_{c,g,t}$ is a cohort-level outcome variable that varies with cohort age (g) and time (t); Z_c represents the economic conditions at the time when the cohort entered the market. D_g is an indicator variable that take the value of one if a group of firms is g years of age. η_g represents age-fixed effects. To study the effect of the initial aggregate conditions over the cohorts' life-cycle, I consider the interaction of the business cycle conditions at entry with the cohort age. The coefficient β_g measures the average change in the variable of interest at age g with the variation in the business cycle conditions at entry.

To empirically estimate the regression, I use the state-level annual BDS dataset over the period 1978-2015. To be consistent with the model calibration, I consider a cohort of establishments as a unit of analysis. I use the cyclical component of the HP-filtered log real GDP to measure the aggregate economic conditions at entry. I additionally include state-fixed effects in the above regression equation. Table 8 report the results and shows that cohorts

Figure 11: Productivity at entry and over time



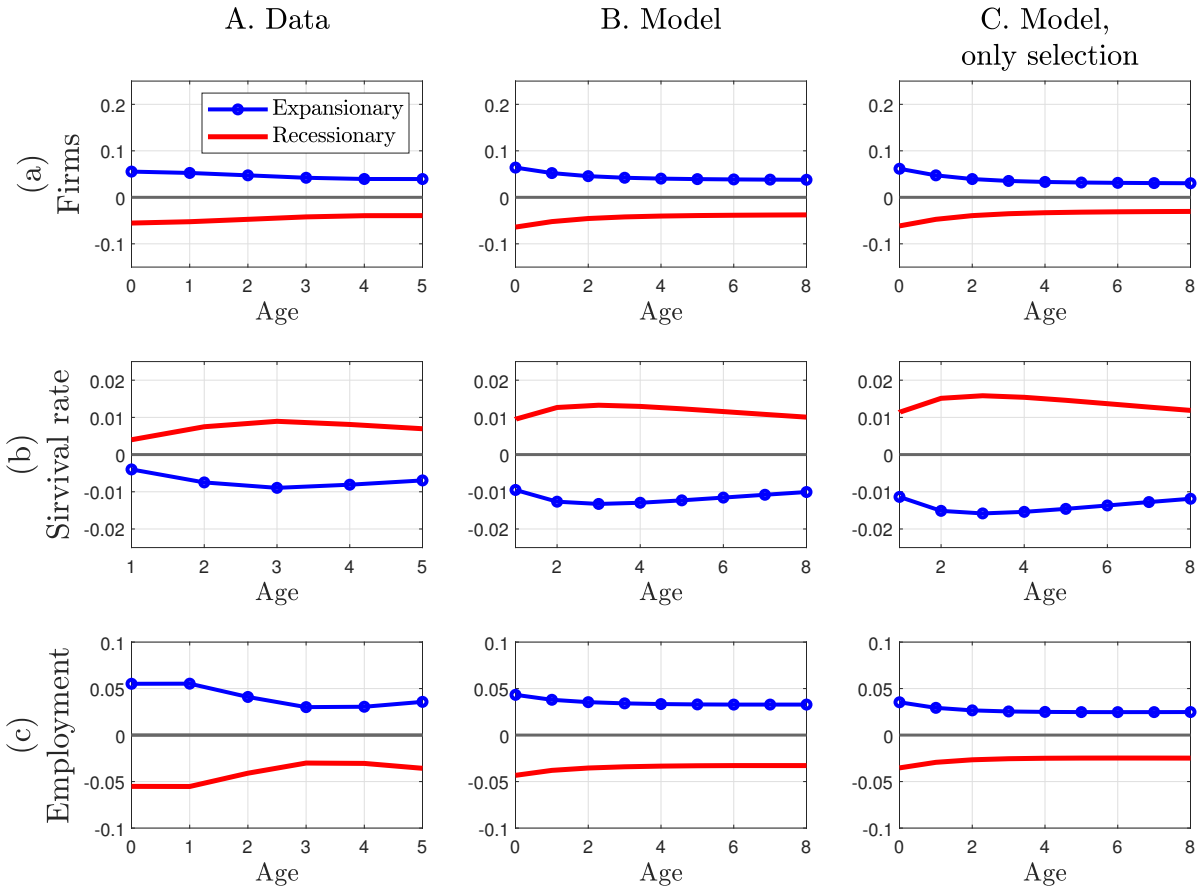
born during recessions consist of fewer firms, employ fewer workers, and have higher survival rates compared to their expansionary counterparts. Moreover, the effects persist over cohorts' life cycles.

To estimate the regression using model economy, I simulate the model over 500 periods and follow each cohort of firms for up to 15 years of operation. I measure Z_c using the log aggregate demand level firms observe at entry.[25] For comparability, I use the estimated coefficients of the model and the data to construct the simulated life cycle characteristics of cohorts born across different aggregate states. That is, I calculate $\hat{X}_{c,g} = \hat{\alpha} + \sum_{g=0}^{T} D_g \hat{\eta}_g + \sum_{g=0}^{T} D_g \hat{\beta}_g Z_c$. Based on the expression, cohorts' average life cycle characteristics vary across the aggregate conditions at entry Z_c. I assign Z_c the following three values $\{-\sigma_z, 0, \sigma_z\}$, where σ_z is a standard deviation of the respective business cycle variable. I refer to cohorts of firms as recessionary (expansionary) if they are born when the aggregate conditions are one standard deviation below (above) the mean. In Figures 11 and 12, the blue-circle and red lines illustrate recessionary and expansionary cohorts life cycle characteristics relative to the mean cohort ($Z_c = 0$).

Consistent with the empirical findings, in the model, the aggregate conditions at entry have a significant and persistent effect on cohorts' life cycle characteristics. Figure 11 shows that the average productivity of the recessionary cohorts is higher than their expansionary counterparts. The difference persists in later years due to the persistent idiosyncratic productivity process. The result is a direct implication of the countercyclical variation in the threshold signal discussed in Section 5.3, as it leads to the endogenous business cycle variation in the number and composition of entrants.

[25] The results are robust if I define business cycles using the deviations from the average log employment (output) or the cycle component of the HP-filtered log employment (output).

Figure 12: Entry conditions and persistent cohort dynamics



Note. The figure illustrates the deviations of the recessionary and expansionary cohorts' characteristics from the mean (steady state) cohorts. Recessionary cohorts refer to groups of firms that start operating when the aggregate conditions are one standard deviation below (above) the mean. Panel A ('Data') uses the state-level annual Business Dynamics Statistics dataset. The unit of analysis is a cohort of establishments. Panel B ('Model') describes the same statistics using simulated model data. Panel C ('Model, only selection') considers a counterfactual version of the model where the shocks affect the selection of firms at entry and not post entry shocks.

Panels A and B of Figure 12 describe the recessionary and expansionary cohorts' life cycle charachteristics - total number of firms, survival rates, and employment – in the data and the model, respectively.

Number of firms In the model, the cohort of firms that start operating during recessions consists around 12% fewer firms at entry compared to expansionary cohorts. The difference equals 8% at age 5 and persists over time. The result is very close to the data counterpart. That is, in the US, the recessionary cohorts consists with around 11% and 8% fewer establishments at age 0 and age 5, respectively. These dynamics are illustrated in figures 12A(a) and 12B(a).

Survival rate In the data and the model, the firms that start operating during the recessionary periods have, on average, higher survival rates compared to their expansionary counterparts. In the model, the difference equals 1.8% at entry and 2.0% at age 5. In the data, the corresponding statistics equal 0.8% and 1.5%, respectively. These statistics are illustrated in figures 12A(b) and 12B(b).

Employment In the model, the recessionary cohorts employ around 8.6% fewer workers at entry and 6.4% at age five. In the data, the estimates are statistically significant and equal to 11% and 7%, respectively. These statistics are illustrated in figures 12A(c) and 12B(c). Note that while the model matches the differences between the expansionary and recessionary cohorts' employment starting from age 3, it underestimates the effect of the initial aggregate conditions on cohort-level employment at age 0. That is, in the model, the difference between the recessionary and expansionary cohorts' employment is three percent less than the data counterpart. The latter suggest that other channels, such as the variation in the niche and mass product firms (Sedlaćek and Sterk, 2017), and variation in the share of the high-target size firms (Smirnyagin, 2021) over the business cycles might be needed to fully account for the effect of the initial conditions on the cohorts' employment.[26]

The Role of Selection The persistent differences between recessionary and expansionary cohorts could be due to the selection – the change in the number and composition of firms at entry and/or the shocks they face after entry. For example, group of firms that enter during expansions are not only different in composition but might also face higher aggregate demand levels during the first couple of years of operation. To decompose the effect of the selection and post-entry shocks, I consider the dynamics of the counterfactual economy where the shocks only affect selection of firms. Panel C of Figure 12 illustrates cohorts' demography in the counterfactual scenario. Comparing Panel C to Panel B shows that the significant (more than 80%) differences in the cohorts' characteristics are due to the selection of firms across aggregate conditions. The model mechanism is consistent with the recent firm dynamics literature, which documents that the significant share of the life-cycle differences across cohorts is due to the pre-entry selection of firms rather than post-entry shocks (Sterk,

[26]The model produces the average size that is higher for the recessionary cohorts compared to expansionary cohorts. In Appendix A.1, I study how the average size of cohorts varies with the business cycle conditions at inception. I find no statistically robust relationship between average size and aggregate conditions at entry for the cohorts of establishments. However, I find a statistically negative relationship between aggregate conditions at entry and cohorts of firms' average size dynamics over time. That is, cohorts of firms that start operating during recessions have larger average sizes at entry and over time than their expansionary counterparts. Although, the effect dissipates over time when the cohorts age.

Sedlaćek, and Pugsley, 2021).[27]

5.4 Aggregate Fluctuations

Finally, I use the good fit of the model to the life cycle firm dynamics to evaluate how the persistent and significant differences in cohorts' characteristics over the business cycles shape the aggregate fluctuations.

To find the cyclical properties of the log number of entrant establishments, log total number of establishments, and log aggregate employment over the period 1978-2019, I follow Fajgelbaum, Schaal, and Taschereau-Dumouche (2017) and use a linear detrending method that allows a structural break in the trend.[28]

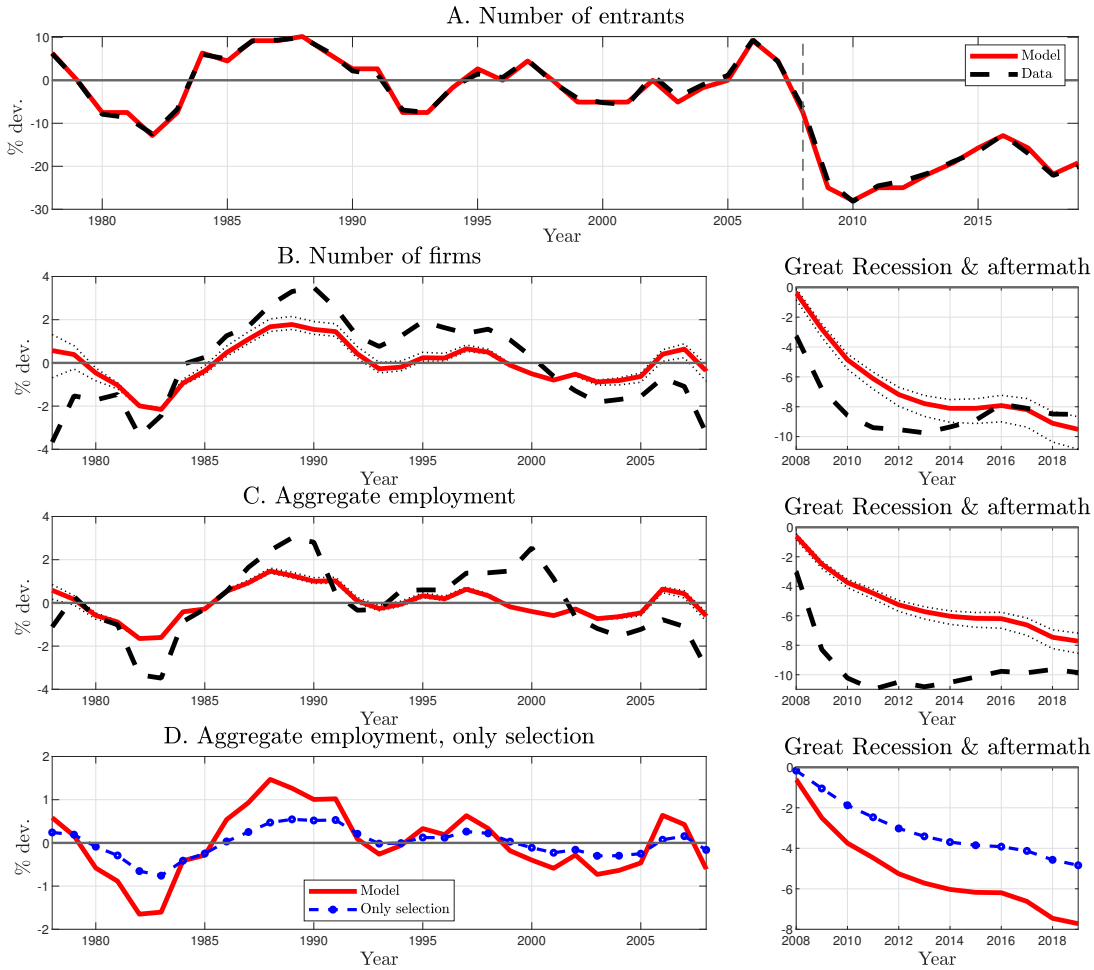
To study the model's implications for the aggregate fluctuations, I construct an aggregate demand shock series that matches the deviations of the log number of entrants from the trend over the period 1978-2019 in the model and the data. I use the shock process to generate 1000 simulations of the economy over 300 periods. In each simulation, the economy faces the same constructed shock process in the last 42 periods, corresponding to 1978-2019. I apply a linear trend to find the cyclical component of each variable, after which I average each time series across simulations. The results are illustrated in Figure 13. For visual readability, I separately present the time series for the pre-Great Recession (1978-2008) and the Great Recession & aftermath (2008-2019) periods.

I find that the correlation between the cyclical component of the total number of establishments in the model and the data is 0.93. This statistic equals 0.94 for aggregate employment. The high correlations between the model simulated time series and the data counterpart show that the model closely tracks the observed aggregate fluctuations. Table 9 compares the properties (standard deviations and autocorrelations) of these time series in the data and the model. The variance of the aggregate employment in the baseline model is 0.008, 50% of what is observed in the data. The autocovariance of the aggregate employment equals 0.76 and 0.91 in the model and the data, respectively.

[27] In Appendix A (Table 12), I show that after controlling for the time fixed effects in regression equation (6), the role of the initial aggregate conditions on the persistent differences in cohorts' characteristics (reported in Table 8) stays statistically significant. Moreover, the magnitude of the effect increases over the cohorts life cycle.

[28] In Appendix D.2, Figure 29 compares the trend and cyclical component of aggregate variables after applying the HP-filter with smoothing parameter 100, a linear trend and a linear trend that allows a break point in trend. The HP-filters predicts that the Great Recession was a strong downturn after which the economy recovered quickly. On the other hand, the linear trend exaggerates the severity of the recession. I

Figure 13: Aggregate fluctuations



Note. The empirical time series represents the deviations of the log number of entrant establishments, log number of establishments, and log aggregate employment from their respective trends in the US over the period 1978-2019. To find the cyclical properties of these time series, I use a linear detrending method that allows a structural break in the trend. In Panel D, 'only selection' describes a counterfactual economy where aggregate shocks only affect the selection of firms at entry and not decisions made by firms after entry.

To evaluate the role of the variation in the number and composition of entrants in the aggregate fluctuations, I construct a counterfactual scenario ('Only selection') in which the exogenous shock affects only the selection of entrants. Panel C of Figure 13 compares the dynamics of the counterfactual employment to the baseline model. The correlation of the counterfactual aggregate employment with the baseline and the data equals 0.99 and 0.92, respectively. Row 'Baseline, only selection' of Table 9 shows that the variation in the types of firms at entry and resulting persistent differences in cohort-level employment build up around 20% of the volatility observed in the data. The latter is a big number when compared with

show that the predictions of the model is robust across these filters.

Table 9: Business cycle moments: Data, baseline, and alternative scenarios

	Entry	**Firms**	**Employment**
Panel A: Standard deviation			
Data	0.062	0.019	0.016
Baseline	0.063	0.010	0.008
Baseline, only selection	0.063	0.009	0.004
Baseline $d = 0$	0.014	0.003	0.006
Baseline $d = 0$, only selection	0.014	0.002	0.001
Model w/o delay	0.064	0.012	0.027
Model w/o delay, only selection	0.064	0.009	0.003
Panel B: Autocorrelation (1st lag)			
Data	0.624	0.850	0.809
Baseline	0.624	0.815	0.755
Baseline, only selection	0.624	0.784	0.818
Baseline $d = 0$	0.587	0.880	0.726
Baseline $d = 0$, only selection	0.587	0.490	0.139
Model w/o delay	0.636	0.850	0.713
Model w/o delay, only selection	0.636	0.801	0.951

Note. These empirical time series describe the deviations of the log number of entrant establishments, log number of establishments, and log aggregate employment from their respective trends in the US over the period 1978-2008. To find the cyclical properties of these tune series, I use a linear detrending method in the data and the model. 'Only selection' describes a counterfactual economy where the aggregate shock process only affect the selection of firms and not their post-entry decisions.

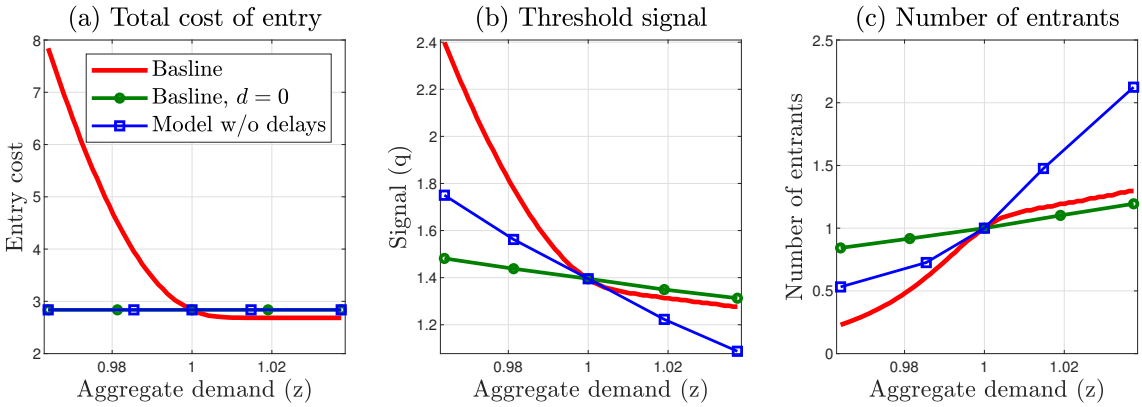
a small share of entrants' employment in aggregate employment (see Figure 10).

The entire model accounts for around 45% of the depth and more than 85% of the slow recovery in aggregate employment during and aftermath of the Great Recession (Panel C). Panel D shows that while the changes at the entry margin do not explain much of the depth of the recession, the cumulative contribution of the persistent drop in the new cohorts' employment accounts for around half of the slow recovery in aggregate employment by 2019. To sum up, the changes in the number and composition of entrants over the business cycles have a significant and persistent effect not only on the cohort-level outcomes, but also on the aggregate economy.

6 The Option–Value Channel

So far, we have seen that the model closely replicates the life cycle characteristics of cohorts of firms in the US, on average, and over the business cycles. Also, I have shown that the persistent differences in cohorts' characteristics significantly amplify and propagate aggregate shocks. In the rest of the paper, I explore the quantitative and qualitative importance of the option-value channel in accounting for the observed micro- and macro-level fluctuations.

Figure 14: The option-to-delay channel



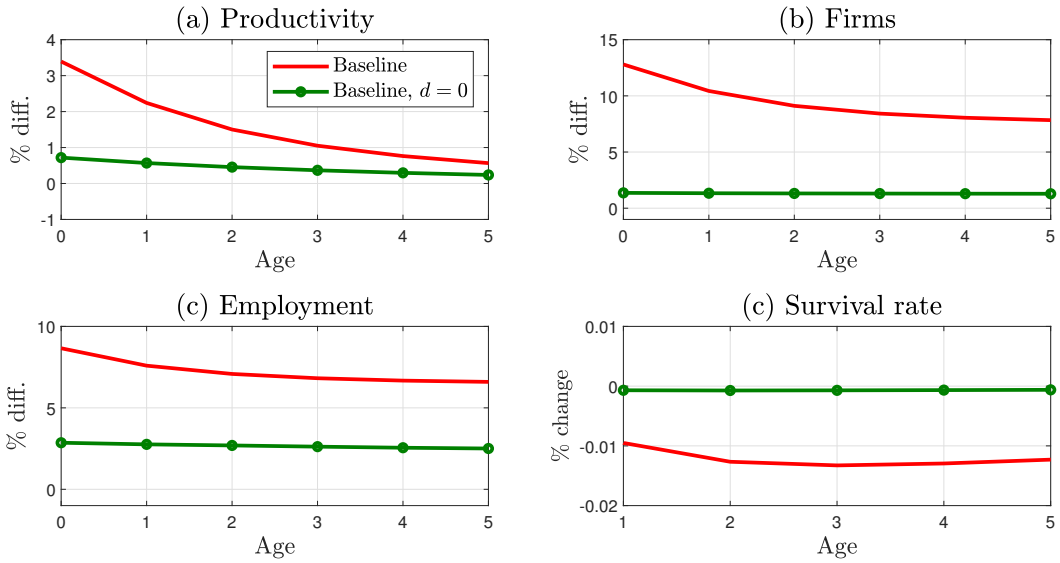
6.1 Quantitative Evaluation

To isolate the selection of entrants through the option-value channel, I evaluate the performance of the model without the option. Toward the goal, I set $d = 0$ in the baseline model. I refer to the latter version of the model as the baseline with $d = 0$ case. The baseline and the alternative scenario are identically parameterized except the fixed entry cost, which is adjusted to ensure that the baseline with $d = 0$ case coincides to the main model in the stochastic steady state.[29] Figure 14 shows that the opportunity cost of entry, the threshold signal, and the number of entrants is the same across these cases in the steady state.

The differences between the baseline and the baseline with $d = 0$ case come when we go beyond the steady state. The option-value channel that leads to the endogenous countercyclical variation in the entry cost increases the elasticity of the threshold signal with respect to aggregate conditions in the baseline model. For a visual illustration, compare the opportunity cost of entry and the threshold signal across these scenarios in Figures 14(a) and 14(b), respectively. The range of signals between the green-circle line and the red-solid line describes an additional group of firms that decide not to enter the market if they have the option to wait. Since the opportunity cost of entry only affects the selection – the number and composition of entrants, and has absolutely no effect on the post-entry decisions of firms, comparing these scenarios completely isolate the option-to-delay channel in the selection and post-entry dynamics of cohorts.

[29] To achieve the goal, the fixed entry cost in the baseline with $d = 0$ case should equal to the steady state total opportunity cost of entry in the baseline model. That is, $c_{d=0} = c_{d=1} + V^w(q^*_{d=1}(z_{ss}), z_{ss})$. The Column (b) of Table 15 summarizes the parameter values used in the baseline with $d = 0$ case.

Figure 15: Entry conditions and persistent cohort dynamics: The option-value effect



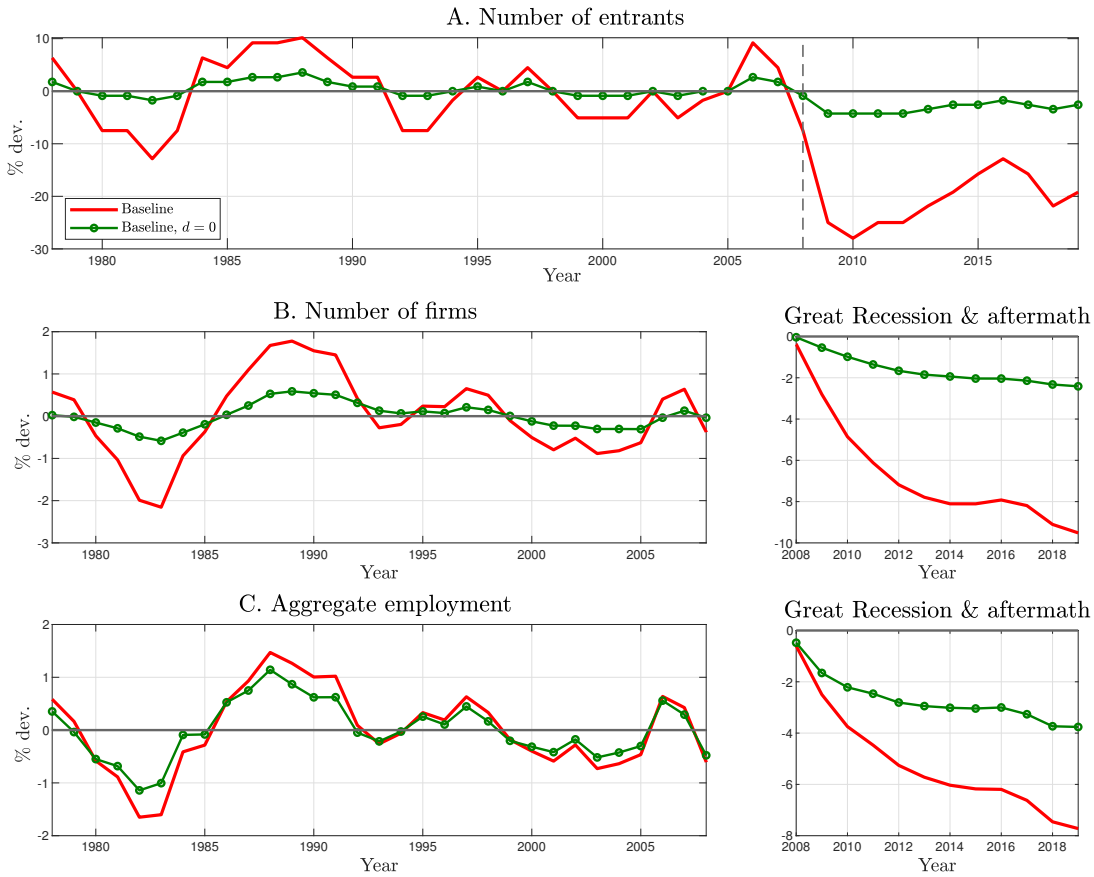
Note. The figure illustrates the percentage differences between the expansionary and recessionary cohorts' characteristics in the baseline and the baseline with $d = 0$ case. Recessionary cohorts refer to groups of firms that start operating when the aggregate conditions are one standard deviation below (above) the mean.

Selection and Persistence in Cohorts Dynamics To explore the quantitative importance of the option, I re-estimates the regression equation (6) using the simulated data from the baseline with $d = 0$ case. Figure 15 illustrates the percentage differences between expansionary and recessionary cohorts characteristics in the baseline and the alternative case. Panel (a) shows that without the option, the elasticity of the threshold signal decreases so much that the differences in average productivity falls from 3% to *0.5*%. Figures 15(b) and 15(c) show that in the baseline with $d = 0$ case, the recessionary cohorts consist of around 2% fewer firms and employ around 3% fewer workers at entry and over time compared to their expansionary counterparts. Comparing the results to the baseline model and the data show that 80% of the drop in the number of firms during recessions is due to the firms that decide to delay entry. Moreover, the option-value channel is responsible for around 70% in the persistent drop of the recessionary cohorts' employment.

Finally, Figure 15(c) shows that without the option to wait, the differences in the survival rates become economically negligible. As we saw in Section 4, in the baseline model, the countercyclical variation in the cohorts' survival rates is driven by firms who decide to postpone entry and wait for aggregate conditions that guarantee higher survival rates.

Aggregate Fluctuations To evaluate the aggregate implications of the option-value channel, I compare the response of the baseline and the baseline with $d = 0$ case to the shock

Figure 16: Aggregate fluctuations: The option-value channel



Note. The time series describe the response of the economy to a constructed shock process that matches the dynamics of the number of entrants in the baseline model and the data. I apply a linear trend to find the cyclical properties of the number of firms and aggregate employment. The baseline, $d = 0$ case represents a version of the baseline model without the option-value channel.

process that matches the dynamics of entrants in the baseline model.[30] In Figure 16, the differences between the red-solid and the green-circle lines are completely driven by the selection of entrants through the option-value channel. The properties of these time series are summarized in Table 9. I find that the endogenous countercyclical opportunity cost of entry is responsible for around 80% of the variance in the number of entrants and 60% in the total number of firms. Without the option-value channel, the volatility of the aggregate employment drops by 25% in the model, which corresponds to 12% of the volatility in the data. The baseline with $d = 0$ case is also characterized by lower auto-correlations in the aggregate time series as shutting down the option-value channel also takes away the persistent differences across cohorts. Finally, in contrast to the main model, in the baseline with $d = 0$ case the post-entry shock drives around 90% of the business cycles and only 10% of

[30]For details about the exercise see Section 5.4.

the volatility is accounted by the selection of entrants (see raw 'Baseline with $d = 0$, only selection' in Table 9).

Note that the option-value channel is particularly important during episodes with prolonged adverse aggregate shocks accompanied by the persistent drop in the number of entrants. For example, consider the following periods 1980-1983 and 1978-1993. That said, the channel plays a particularly important role during the Great Recession. According to the counterfactual exercise, over the period 2009-2019, the persistent change in the number and composition of entrants and resulting cumulative drop in their employment accounts for around 40% of the drop in aggregate employment by 2019.[31]

6.2 Standard Case: A Model without Delay

Next, I compare the performance of the baseline model to a workhorse firm dynamics mode. I show that the latter does not account for the observed dynamics of entrants for a reasonable aggregate demand shock process.

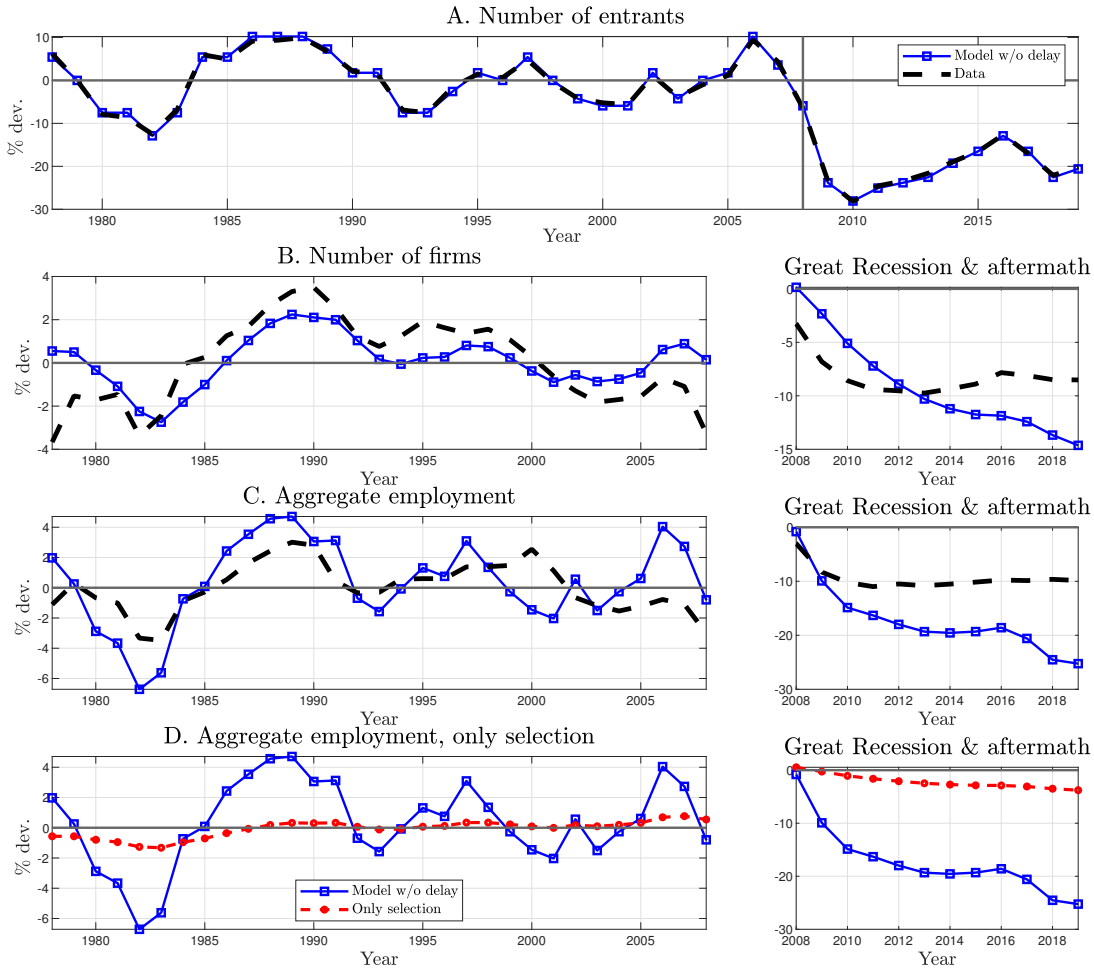
Consider a version (Model w/o delay) of the baseline model with $d = 0$, calibrated to match the same set of facts described in Section 5.1. In this model, the entry decision follows a traditional neoclassical investment rule: enter if the NPV is non-negative. The entry cost is fixed and does not vary over the cycles (Figure 14a). Thus, the aggregate shock can affect the selection of entrants only through its direct effect on potential firms' lifetime profits. I find that producing the observed variation in the number and composition of entrants without the option-value channel requires the standard deviation and the autocorrelation of the aggregate demand shock to be 0.016 and 0.64, respectively.[32] In the baseline model, the numbers are 0.003 and 0.75, respectively. In terms of the unconditional variance, the model w/o delay requires shocks with 5-times higher magnitudes to produce the observed variation in entry than the baseline model. To put it differently, the endogenous countercyclical variation in the cost of entry amplifies the elasticity of entrants to aggregate shocks 5-times.

The exogenous shock process affects not only the selection of entrants, but incumbent firms' production and continuation decisions. Thus, the different exogenous shock processes in the

[31] In Appendix A.4, Figure 23 shows that the share of start-ups that formed business with one year delay during and aftermath of the Great Recession increased by 5 percentage point, from 12% in year 2007 to 17% by year 2014.

[32] Appendix E provides a detailed description of the model w/o delay's calibration procedure. In Appendix E, Table 15 summarizes the parameter values, and tables 16, and 17 summarize how the moments targeted in the Standard model compare with the data counterpart and the baseline model.

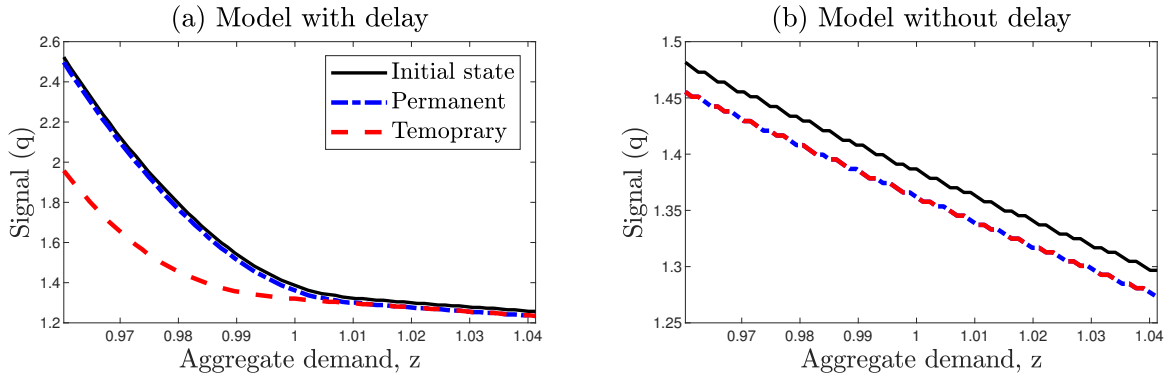
Figure 17: Aggregate fluctuations: Model without delay



Note. These empirical time series describe the deviations of the log number of entrant establishments, log number of establishments, and log aggregate employment from their respective trends in the US over the period 1978-2019. To find the cyclical properties of these time series, I use a linear detrending method that allows a structural break in the trend. In Panel D, 'only selection' describes a counterfactual economy where aggregate shocks only affect the selection of firms at entry and not decisions made by firms after entry.

baseline and the model w/o delay have different implications about the role of entry margin in aggregate fluctuations. Brief summary of the main conclusions are as follows. First, the model w/o delay leads to excessive aggregate fluctuations. Row 'model w/o delay' of Table 9, shows that the variance of the aggregate employment is around *1.7* times higher than the data counterpart. To put the number into perspective, Panel C of Figure 17 shows that a shock process that matches the dynamics of entrants predicts the drop in aggregate employment that is more than 2-times larger compared to the one observed in the data in the aftermath of the Great Recession. Second, in the model w/o delay, the post-entry shocks account for around 90% of the volatility in aggregate variables; see Panel D of Figure 17 and row 'model w/o delay, selection' of Table 9. As a result, the model w/o delay significantly

Figure 18: Permanent/temporary decline in the fixed entry cost



underestimate the relative role of the observed variation in the entry margin in shaping the aggregate fluctuations.

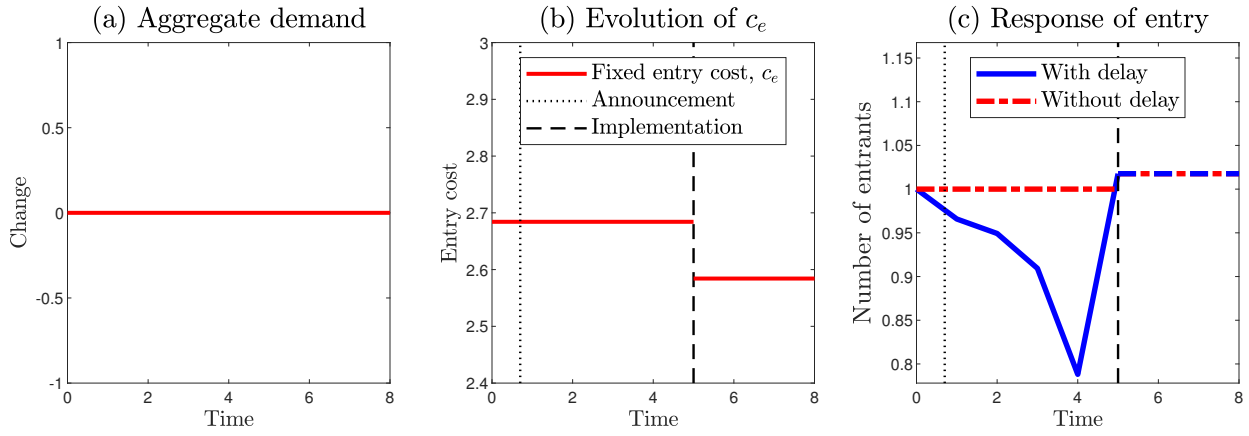
Existing literature uses various approaches to reconcile the observed variation in the number and composition of entrants for a reasonable aggregate demand shock process. One possibility is to use exogenous entry cost shock or a function that varies over the cycles as in Lee and Mukoyama (2018) and Clementi and Palazzo (2016), or introduce entry function, which allows choosing the elasticity of the number of entrants with respect to aggregate shocks as in Sedlaček and Sterk (2017). In that respect, one could also think about the option-to-delay channel as a microfoundation for these exogenous mechanisms – it endogenously generates the countercyclical cost of entry that increases the elasticity of entrants with respect to aggregate demand.

6.3 Policy Implications

Finally, I argue that not accounting for entry timing may lead to misleading predictions about potential entrants' responses to different shocks or policies. The reason is the following. With the option to delay entry, the dynamics of entrants depend on how the changes in the aggregate environment affect the relative benefits of entry today versus tomorrow. Whereas the standard frameworks only account for the shock's direct effect. Thus, depending on the type, magnitude, timing, and duration of the shocks, the standard framework may lead to imprecise predictions about the response of potential entrants. To illustrate the point, I demonstrate how the option to wait alters potential entrants' responses to the permanent, temporary, or future reductions in the entry cost.

Permanent versus Temporary Policy Figures 18(a) and 18(b) contrast the changes in the

Figure 19: News about the decline in the fixed entry cost



threshold signal level as a response to a permanent and a temporary decrease in the fixed entry cost. First, consider a model with the option to delay entry. If the goal is to increase the number of entrants, the temporary decline in the fixed entry cost does a better job during recessions, and has the same effect during expansions compared with a permanent decline in the fixed entry cost. Moreover, marginal entrants who respond to the reduction of the fixed entry cost are mostly high-productivity firms during recessions and low-productivity firms during expansions. Without the option to delay entry, the responses of entrants do not vary across these policies.

News Shock Consider the response of potential entrants to an anticipated decline in the fixed entry cost after five periods from today. Figures 19(a) and 19(b) show the actual change in the aggregate demand and the level of entry cost, respectively, over time. I find that the threshold signal in the news scenario is weakly higher than in the baseline (no-news) scenario. The magnitude of the change depends on the distance between today and the policy's actual time. Figure 19(c) describe the response of entrants to the news with and without the option to delay entry. If the time of the actual decrease in the entry cost is close enough (small T), the indirect effect of the news that decreases the number of entrants today is quantitatively more significant than the increase in the number of entrants at time T as a response to the lower fixed entry cost. In the standard firm dynamics models, the news would have altered the dynamics of entrants today only through general equilibrium effects.[33] However, as the exercise illustrates, the response of entrants to the policy announcement through the option-value-of-delay channel could be quantitatively more important.

[33] Constantini and Melitz (2008) also show that potential entrants respond differently to the news about trade liberalization depending on the timing and the implementation of the policy.

7 Conclusions

In this paper, I show that firms' option to delay entry, missing in existing frameworks, has important implications for our understanding of entrants' business cycle dynamics. I provide a firm dynamics model with endogenous entry and exit, which allows firms to postpone business formation after observing the initial aggregate conditions. I find that the option to wait endogenously generates a countercyclical opportunity cost of entry: During recessions, a higher risk of failure increases the value of waiting, hence the cost of entry. The calibrated model successfully accounts for the life cycle dynamics of cohorts in the US on average and over the business cycles.

The option-value channel is quantitatively important in accounting for the observed business cycle dynamics of entrants. I find that the endogenous countercyclical entry cost increases the variation of entrants over the business cycles five times. This channel accounts for around 80% of the observed differences in the recessionary and expansionary cohorts' number of firms, employment, and productivity. The variation in the medium productivity firms who choose to postpone entry produce cohorts with countercyclical survival rates. The option-value channel also builds significant persistence in the dynamics of economic aggregates and contributes to 10% of the volatility in aggregate employment observed in the data.

The option-to-delay channel provides a microfoundation for endogenously reconciling the observed significant effect of the initial aggregate conditions on the selection of entrants. Without the mechanism, existing models require either large shocks that generate excessive aggregate fluctuations or exogenous mechanisms to reconcile the observed dynamics of entrants. I also argue that overlooking this channel may also result in misleading predictions about entrants' responses to different shocks or policies.

Further Applications The framework provides an interesting avenue for future research. For example, using the framework, one can study how the changes in the ability to delay entry over time could explain the decreasing business dynamism in the US. Another possibility is to explore how the heterogeneity in the ability to postpone entry explains the variation in the entry rates across sectors. Additionally, one can re-examine, study and quantify the effect of different policies (e.g., labor adjustment tax, entry subsidies, R&D subsidies) on the response of entrants and the dynamics of the aggregate variables or investigate stabilization policies. While in the paper, I examined how the option to wait alters entry decisions, explaining the dynamics of potential entrants after they use the option (e.g., whether they actually come

back to start a business) is also left for future research. I believe that with the development of the Business Formation Statistics dataset, the framework can be very useful to uncover further details about the dynamics of entrants over time.

References

[1] Ates, S.T. and Saffie, F.E., 2021. Fewer but better: Sudden stops, firm entry, and financial selection. American Economic Journal: Macroeconomics, 13(3), pp.304-56.

[2] Bayard, K., Dinlersoz, E., Dunne, T., Haltiwanger, J., Miranda, J. and Stevens, J., 2018. Early-stage business formation: An analysis of applications for employer identification numbers (No. w24364). National Bureau of Economic Research.

[3] Bernanke, B.S., 1983. Irreversibility, uncertainty, and cyclical investment. The quarterly journal of economics, 98(1), pp.85-106.

[4] Bilbiie, F.O., Ghironi, F. and Melitz, M.J., 2012. Endogenous entry, product variety, and business cycles. Journal of Political Economy, 120(2), pp.304-345.

[5] Clementi, G.L., Khan, A., Palazzo, B. and Thomas, J.K., 2014. Entry, exit and the shape of aggregate fluctuations in a general equilibrium model with capital heterogeneity. Unpublished Working Paper.

[6] Clementi, G.L. and Palazzo, B., 2016. Entry, exit, firm dynamics, and aggregate fluctuations. American Economic Journal: Macroeconomics, 8(3), pp.1-41.

[7] Cogley, T. and Nason, J.M., 1995. Output dynamics in real-business-cycle models. The American Economic Review, pp.492-511.

[8] Costantini, J.A. and Melitz, M.J., 2008. The dynamics of firm-level adjustment to Trade Liberalization. The organization of firms in a global economy, p.107.

[9] Decker, R., Haltiwanger, J., Jarmin, R. and Miranda, J., 2014. The role of entrepreneurship in US job creation and economic dynamism. Journal of Economic Perspectives, 28(3), pp.3-24.

[10] Dixit, R.K. and Pindyck, R.S., 2012. Investment under uncertainty. In Investment Under Uncertainty. Princeton university press.

[11] Fajgelbaum, P.D., Schaal, E. and Taschereau-Dumouchel, M., 2017. Uncertainty traps. The Quarterly Journal of Economics, 132(4), pp.1641-1692.

[12] Folta, T.B. and O'Brien, J.P., 2004. Entry in the presence of dueling options. Strategic Management Journal, 25(2), pp.121-138.

[13] Fort, T.C., Haltiwanger, J., Jarmin, R.S. and Miranda, J., 2013. How firms respond to business cycles: The role of firm age and firm size. IMF Economic Review, 61(3), pp.520-559.

[14] Foster, L., Haltiwanger, J. and Syverson, C., 2008. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?. American Economic Review, 98(1), pp.394-425.

[15] Foster, L., Haltiwanger, J. and Syverson, C., 2016. The slow growth of new plants: Learning about demand?. Economica, 83(329), pp.91-129.

[16] Geroski, P.A., 1995. What do we know about entry?. International journal of industrial organization, 13(4), pp.421-440.

[17] Gourio, F., Messer, T. and Siemer, M., 2016. Firm entry and macroeconomic dynamics: a state-level analysis. American Economic Review, 106(5), pp.214-18.

[18] Messer, T., Siemer, M. and Gourio, F., 2016. A missing generation of firms? Aggregate effects of the decline in new business formation. In 2016 Meeting Papers (No. 752). Society for Economic Dynamics.

[19] Haltiwanger, J., Jarmin, R.S., Kulick, R. and Miranda, J., 2016. High growth young firms: contribution to job, output, and productivity growth. In Measuring entrepreneurial businesses: current knowledge and challenges (pp. 11-62). University of Chicago Press.

[20] Haltiwanger, J., Jarmin, R.S. and Miranda, J., 2013. Who creates jobs? Small versus large versus young. Review of Economics and Statistics, 95(2), pp.347-361.

[21] Hopenhayn, H.A., 1992. Entry, exit, and firm dynamics in long run equilibrium. Econometrica: Journal of the Econometric Society, pp.1127-1150.

[22] Hopenhayn, H. and Rogerson, R., 1993. Job turnover and policy evaluation: A general equilibrium analysis. Journal of political Economy, 101(5), pp.915-938.

[23] Jovanovic, B. (2008). Jovanovic, B., 2009. Investment options and the business cycle. Journal of Economic Theory, 144(6), pp.2247-2265.

[24] Khan, A., Thomas, J. and Senga, T., 2019. Business Formation and Economic Growth Beyond the Great Recession. In 2019 Meeting Papers (Vol. 1453). Society for Economic Dynamics.

[25] King, R.G. and Rebelo, S.T., 1999. Resuscitating real business cycles. Handbook of macroeconomics, 1, pp.927-1007.

[26] Lee, Y. and Mukoyama, T., 2015. Entry and exit of manufacturing plants over the business cycle. European Economic Review, 77, pp.20-27.

[27] Lee, Y. and Mukoyama, T., 2008. Entry, exit and plant-level dynamics over the business cycle.

[28] Lee, Y. and Mukoyama, T., 2018. A model of entry, exit, and plant-level dynamics over the business cycle. Journal of Economic Dynamics and Control, 96, pp.1-25.

[29] McDonald, R. and Siegel, D., 1986. The value of waiting to invest. The quarterly journal of economics, 101(4), pp.707-727.

[30] Moreira, S. (2015). Moreira, S., 2016. Firm dynamics, persistent effects of entry conditions, and business cycles. Persistent Effects of Entry Conditions, and Business Cycles.

[31] O'Brien, J.P., Folta, T.B. and Johnson, D.R., 2003. A real options perspective on entrepreneurial entry in the face of uncertainty. Managerial and Decision Economics, 24(8), pp.515-533.

[32] Pindyck, R., 1991. Irreversibility, uncertainty, and investment. Journal of Economic Literature, 29, pp.1110-1148.

[33] Pindyck, R.S., 2009. Sunk costs and risk-based barriers to entry (No. w14755). National Bureau of Economic Research.

[34] Pugsley, B.W. and Şahin, A., 2019. Grown-up business cycles. The Review of Financial Studies, 32(3), pp.1102-1147.

[35] Sterk, V., Sedlaćek, P. and Pugsley, B., 2021. The nature of firm growth. American Economic Review, 111(2), pp.547-79.

[36] Samaniego, R.M., 2008. Entry, exit and business cycles in a general equilibrium model. Review of Economic Dynamics, 11(3), pp.529-541.

[37] Sedláček, P. and Sterk, V., 2017. The growth potential of startups over the business cycle. American Economic Review, 107(10), pp.3182-3210.

[38] Sedláček, P., 2020. Lost generations of firms and aggregate labor market dynamics. Journal of Monetary Economics, 111, pp.16-31.

[39] Siemer, M., 2014. Firm entry and employment dynamics during the great recession. Federal Reserve Board: FEDS Working paper, 56.

[40] Smirnyagin, V., 2021. The Impact of Aggregate Fluctuations on Superstar Business Formation. SSRN working paper 3514855.

[41] Veracierto, M.L., 2002. Plant-level irreversible investment and equilibrium business cycles. American Economic Review, 92(1), pp.181-197.

[42] Young, E.R., 2010. Solving the incomplete markets model with aggregate uncertainty using the Krusell–Smith algorithm and non-stochastic simulations. Journal of Economic Dynamics and Control, 34(1), pp.36-41.

Appendix (For Online Publication)

Table of Contents

A Appendix for Empirical Findings

A.1 Aggregate Conditions at Entry and Cohorts' Average Size

I use the US- and state-level annual time series about cohorts of establishments/firms by age from the BDS. The BDS dataset covers the universe of employer businesses in the US and provides yearly measures of business dynamics aggregated by the establishment and firm characteristics. An establishment is defined as a fixed physical location where economic activity is conducted. A firm may consist of one establishment or many establishments and often spans multiple physical locations. The dataset covers the period $1978-2019$. I measure an average size of a cohort of age g at year t as

$$\bar{L}_{g,t} = \frac{L_{g,t}}{N_{g,t}},$$

where $L_{g,t}$ and $N_{g,t}$ measure the total employment and total number of establishments (firms) in a cohort of establishments (firms) of age g at time t. $\bar{L}_{g,t}$ measures the average size of a cohort of establishments (firms) of age g at year t. In this analysis, I consider cohorts' for up to age 5.[34] I measure the aggregate conditions at entry using the cyclical component of log annual real GDP. I find the latter using the HP filter with a smoothing parameter of *100*.

Figure 20 provides binned scatter plots of pooled cohorts' life cycle averages sizes at the US level and state level against the business cycle indicators at the time of entry. Panel (a) of Figure 20 shows that the business cycle conditions at entry are positively associated with average sizes of cohorts of establishments at the US level. However, Panel (b) shows that the correlation becomes negative at the state level. Panels (c) and (d) consider average size of cohorts of firms rather than cohorts of establishments. In both cases, the aggregate conditions at entry are negatively correlated with cohorts' average size. That is, better aggregate conditions at entry are associated with groups of firms with smaller average size.

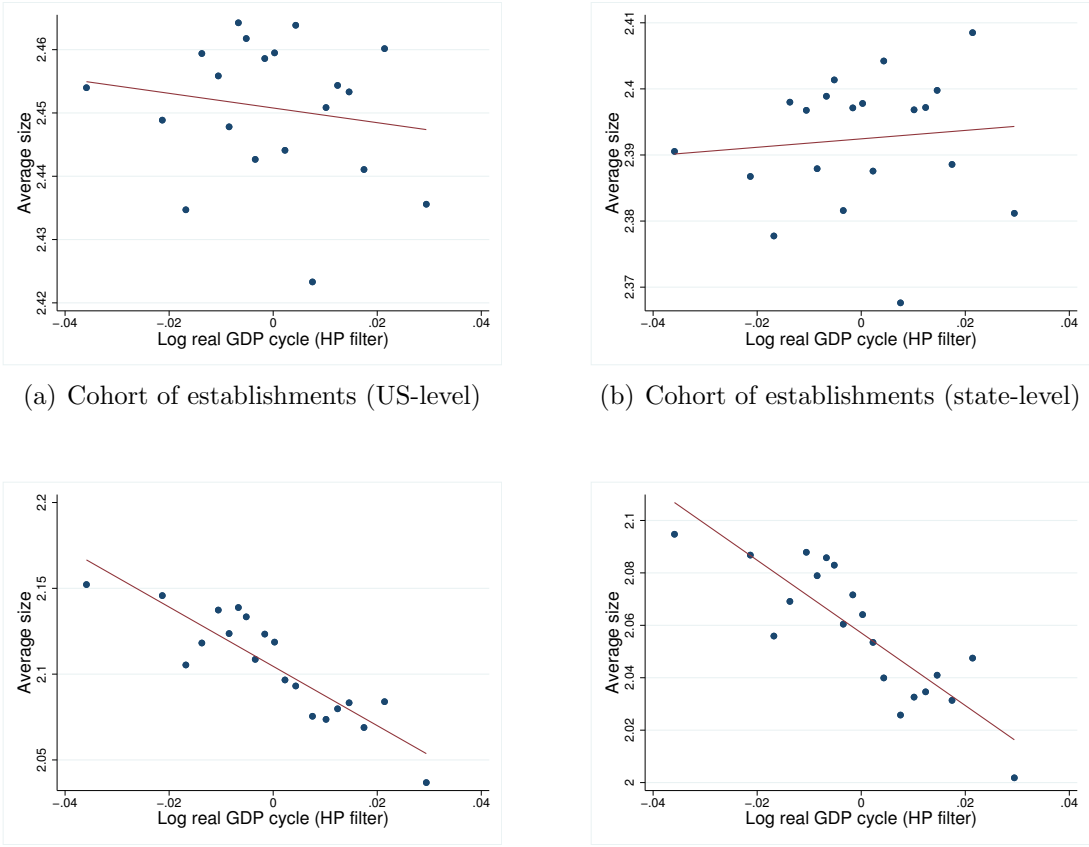
Next, I use the state-level variation in the life cycle dynamics of new businesses to further investigate the relationship. I estimate the following regression:

$$log(\bar{L}_{c,g,s,t}) = \alpha + \beta Z_c + \eta_a + \theta_t + \gamma_s + \varepsilon_{g,s,t}, \tag{7}$$

where $log(\bar{L}_{g,t})$ is a log average size of a cohort at age g, in state s, at time t; Z_{t-g} represents

[34] The publicly available part of the BDS dataset only provides information about cohorts from age 0 to age 5. Information about cohorts above age 5 is binned into 5-year age groups.

Figure 20: Cohorts' average size against the aggregate economic conditions at the entry



(a) Cohort of establishments (US-level)

(b) Cohort of establishments (state-level)

(c) Cohort of firms (US-level)

(d) Cohort of firms (state-level)

Note: Each panel displays a binned scatterplot of cohorts' average sizes against the aggregate conditions at the time of entry. I measure the latter using the cyclical component of the HP-filtered log real GDP with a smoothing parameter of 100. The bin scatterplots control for year- and age-fixed effects. Panels (b) and (d) also control state-fixed effects.

the economic conditions at the time when the cohort first entered the market.[35] η_a, θ_t, γ_s represent age-, year-, and state-fixed effects, respectively. The year-fixed effects controls for the sequence of aggregate shocks cohorts face after entry. That said, β measures a percentage point change in cohorts' average size due to the variation in the business cycle conditions at entry. The coefficient should capture the effect of the initial economic conditions on cohorts' average size after controlling for the age, year and state fixed effects.

Panel A of Table 10 summarizes the estimates of regression equation (7) with a unit of analysis of a cohort of establishments. Column (1) indicates that cohorts born during good economic conditions are characterized by a larger average size over the life cycle, but the

[35]The specification is also similar to the age-period-cohort model where cohort effects are a proxy for economic conditions at birth. See Moreira (2005) for more details.

Table 10: Cohorts' average size and aggregate economic conditions at the time of entry

	Panel A. Establishment			**Panel B. Firm**		
	Y_{HP}	$Y_{HP,I}$	$NBER$	Y_{HP}	$Y_{HP,I}$	$NBER$
	(1)	(2)	(3)	(1)	(2)	(3)
β	0.064	-0.006	0.031***	-1.38***	-0.040***	0.045***
	(0.106)	(0.004)	(0.007)	(0.135)	(0.005)	(0.007)
State FE	✓	✓	✓	✓	✓	✓
Age FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Observations	12,087	12,087	12,087	12,087	12,087	12,087
R^2	0.807	0.807	0.808	0.786	0.783	0.782

Note: Robust standard errors clustered at the state-level are in parentheses. Panel A uses a cohort of establishments, and Panel B uses a cohort of firms as a unit of analysis. Columns (1)-(3) use different indicators for characterizing the business cycle conditions at entry. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

coefficient is not statistically significant. For robustness, I additionally consider the following business cycle indicators. Column (2) uses a business cycle indicator that refers to a year as a recession if the cyclical component of the log real GDP is below trend ($Y_{HP,I}$). Column (3) uses the NBER-based indicator of a recession that spans the period following the peak through the trough ($NBER$).[36] The indicator equals -1 if the year is indicated as recession, 0 otherwise. Columns (2) and (3) show that cohorts born during recessions, on average, have larger survival rates compared to their expansionary counterparts. Panel B of Table 10 shows that the results hold if I use a cohort of firms as a unit of analysis rather than establishments.

I additionally I consider a regression specification where I interact business cycle conditions at entry with the cohort age:

$$log(\bar{L}_{g,t}) = \alpha + \sum_{g=1}^{5} \beta_g D_g Z_{t-g} + \eta_g + \theta_t + \gamma_s + \varepsilon_{g,s,t}, \tag{8}$$

where D_g is an indicator variables that take the value of one if the business establishments/firms are g years of age. The coefficient β_g measures a change in average cohort size at age g with the variation in the business cycle conditions at entry.

Panel A of Table 11 reports the regression results. $1_{\{age=g\}} \times Z$ describes the interaction of the

[36] The latter indicator specifies the peak and trough dates on a monthly frequency. Using the monthly data, I define a year t as a recession if at least four months from April in year $t-1$ to April t are indicated as recessionary periods. Based on the definition, the recessionary years are 1981, 1982, 1983, 1991, 2002, and 2009. All other years are defined as expansionary.

Table 11: Cohorts average size by age and aggregate economic conditions at entry

	Panel A. Establishment			Panel B. Firm		
$Z =$	Y_{HP}	$Y_{HP,I}$	$NBER$	Y_{HP}	$Y_{HP,I}$	$NBER$
	(1)	(2)	(3)	(1)	(2)	(3)
$1_{\{age=0\}} \times Z$	-0.982***	-0.027***	0.072***	-2.24***	-0.043***	0.115***
	(0.160)	(0.005)	(0.007)	(0.16)	(0.005)	(0.008)
$1_{\{age=1\}} \times Z$	-0.447**	-0.013**	0.034***	-1.76***	-0.034***	0.074***
	(0.156)	(0.006)	(0.006)	(0.15)	(0.006)	(0.007)
$1_{\{age=2\}} \times Z$	-0.254*	-0.013***	0.026***	-1.71***	-0.044***	0.058***
	(0.144)	(0.005)	(0.006)	(0.16)	(0.005)	(0.008)
$1_{\{age=3\}} \times Z$	0.151	-0.011**	0.023***	-1.33***	-0.048***	0.032***
	(0.145)	(0.005)	(0.008)	(0.20)	(0.006)	(0.010)
$1_{\{age=4\}} \times Z$	0.706***	0.006	0.018***	-0.87***	-0.041***	0.004
	(0.134)	(0.005)	(0.007)	(0.20)	(0.007)	(0.009)
$1_{\{age=5\}} \times Z$	1.116***	0.021	0.014*	-0.49***	-0.033***	-0.012
	(0.134)	(0.005)	(0.007)	(0.22)	(0.008)	(0.010)
Age FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓	✓
Observations	12,087	12,087	12,087	12,087	12,087	12,087
R^2	0.808	0.807	0.809	0.787	0.783	0.785

Note: Robust standard errors clustered at the state level are in parentheses. Panel A uses a cohort of establishments, and Panel B uses a cohort of firms as a unit of analysis. Columns (1)-(3) use different indicators for characterizing the business cycle conditions at entry. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

business cycle indicators with the cohort of age g. Panel A uses a cohort of establishments as a unit of analysis. Analyzing Columns (1)-(3) show no statistically robust relationship between cohorts' size over the life cycle and aggregate conditions at entry. Panel B considers the same regression if the unit of analysis is a cohort of firms rather than a cohort of establishments. Panel B shows a statistically negative relationship between the aggregate conditions at entry and cohorts' size. That is, cohorts of firms that start operating during recessions are larger at entry and over time than their expansionary counterparts. However, one can see that the effect dissipates over time when the cohort age.

A.2 Selection vs Life Cycle Dynamics

In this section, I investigate differences between cohorts born at different stages of business cycles after controlling the sequence of aggregate conditions they face after entry. Toward the end, in the regression equation (6), I include year-fixed effects.

Table 12: Differences in cohorts' characteristics based on the initial conditions: Data

	Age 0	Age 1	Age 2	Age 3	Age 4	Age 5
Firms	0.14	0.37***	0.99***	1.83***	2.60***	2.93***
	(0.10)	(0.09)	(0.07)	(0.09)	(0.13)	(0.15)
Survival rate		-0.20***	-0.25***	-0.30***	-0.33***	-0.33***
		(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Employment	-0.84***	-0.08	0.74***	1.98***	3.31***	4.05***
	(0.18)	(0.17)	(0.17)	(0.15)	(0.14)	(0.16)
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The estimates use a state-level annual BDS dataset over the period 1978-2015. To be consistent with the model calibration, I consider a cohort of establishments as a unit of analysis. I use the cyclical component of the HP-filtered log real GDP to measure the aggregate economic conditions at entry.

A.3 The Business Formation Statistics

A.3.1 Data Description

The BFS dataset is based on applications for Employer Identification Numbers (EINs) submitted in the United States, known as IRS Form SS-4 filings.[37] EIN application responses include information about reasons for applying, type of entity, business start date, the expected maximum number of employees, the first wage pay date, principal activity of a business, etc. This information is used to identify a subset of applications associated with new businesses, referred to as business applications. Then, the business applications are matched to the set of firms in the BDS identified as new employer businesses based on payroll information. The matching process is straightforward since both datasets contain information about EINs.

In the analysis, I use the following publicly available seasonally adjusted time series at quarterly frequency:

1. *Business formations within 4 quarters (**F**4Q)* - the number of employer businesses that originate from the business applications within four quarters from the quarter of the applications. Time period: 2004Q3-2015Q4. In the analysis, I refer to this time series as **F**4Q.

2. *Business formations within 8 quarters (B8Q)* - the number of employer businesses that

[37] EIN is a unique number assigned to most of the business entities. The EIN is required when the business is providing tax information to the Internal Revenue Service (IRS). Note that EIN applications describe start-up and not establishment-level activities since opening a new establishment does not require a new EIN.

originate from the business applications (BA) within eight quarters from the quarter of the applications. Time period: 2004Q3-2014Q4.

3. *Average duration (in quarters) from business application to formation within 4 Quarters (Dur**F**4Q)* - a measure of delay between business application and formation, conditional on business formation within four quarters. Time period: 2004Q3-2015Q4.

4. *Average duration (in quarters) from business application for formation within eight quarters (Dur**F**8Q)* - a measure of delay between business application and formation, conditional on business formation within eight quarters. Time period: 2004Q3-2014Q4.

   I use these time series to construct the following three variables:

5. *Business formations within the fifth and eighth quarters (**S**4Q):* The number of employer businesses that take between four and eight weeks to transition into employer business from the date of the application. I construct the time series as the difference of $BF8Q - BF4Q$.

6. *Share of late start-ups :* a time series that describes the share of the applications that become employer businesses with one year delay from the date of the application:

$$\textit{Share of late start-ups} = \frac{\mathbf{F}8Q - \mathbf{F}4Q}{\mathbf{F}8Q}$$

7. *Average Duration (in Quarters) from Business Application to Formation from 5 to 8 Quarters (Dur**S**8Q):* a measure of delay between business application and formation, conditional on business formation between the fifth and eighth quarters. I construct the variable using the following formula:

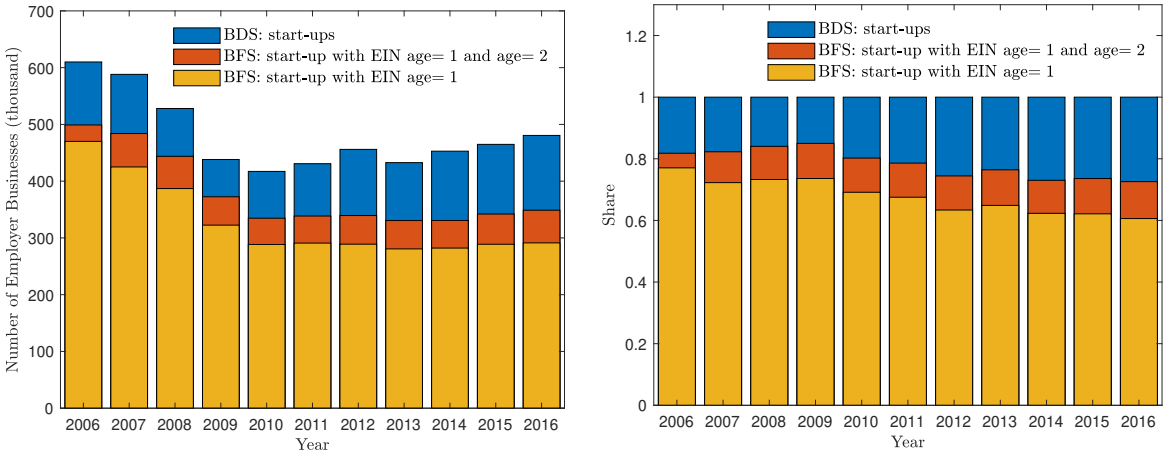
$$Dur\mathbf{S}8Q = \frac{Dur\mathbf{F}8Q\ \mathbf{F}8Q - Dur\mathbf{S}4Q\ \mathbf{F}4Q}{\mathbf{F}8Q - \mathbf{F}4Q}$$

A.3.2 Coverage of the BFS

All firms that show up in the BDS have EINs. Thus, they show up in the BFS dataset before entry.[38] The publicly available part of the BFS dataset allows tracking only the subset of the employer businesses that applied for the EINs within eight quarters before entry. To evaluate the coverage of the publicly available BFS, I compare the information about employer business formation provided by the BFS to the BDS. Since the BDS dataset

[38] There is a small group of employer businesses that get EINs after submitting the first payroll information.

Figure 21: The BDS vs the BFS



(a) Number of business formation

(b) Share of business formation

Note: The figure compares the annual total number of new employer businesses in the BDS and the BFS.

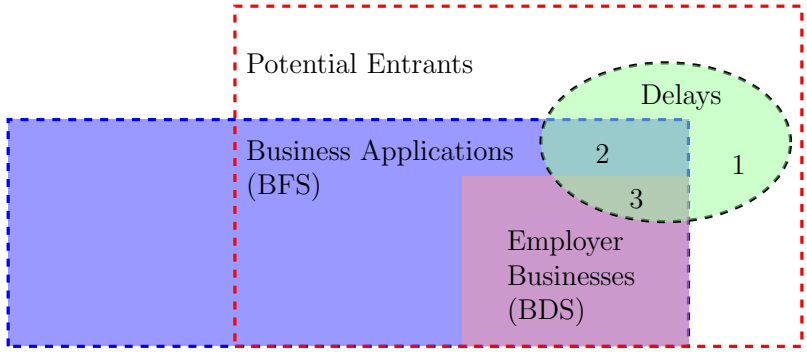
is annual, I convert the quarterly data from the BFS into a yearly time series. Figure 21 shows that the information about the employer businesses provided in the BFS covers more than 80% of the total start-ups in the BDS.

A.3.3 Discussion

The potential entrants that delay entry could belong to the following three groups. First is the group of potential entrants that delay entry and also delay applying for the EIN. Second, the group of potential entrants that apply for the EIN delay starting a business at least for the first eight quarters from the date of the application. Third, the group of potential entrants who apply for the EINs, delay entry in the first year and start businesses in later years. Figure 22 illustrates the potential links between the BFS, the BDS, and potential entrants in the model.

I cannot identify the first and the second groups of entrants using the BFS dataset. On the one hand, potential entrants who choose to delay entry might not apply for the EIN applications. Thus, they are not included in the BFS dataset. On the other hand, some parts of the EIN applications might not be for employer business start-ups. In fact, the data about the raw applications is quite noisy about the business formation. For example, out of the total number of business applications, we see that only 14% become employer businesses within two years from the date of the applications. In particular, 12% become employer businesses in the first four quarters, and an additional 2% become employer businesses after a year.

Figure 22: The potential links between the Business Formation Statistics Dataset (BFS), the Business Dynamics Statistics Dataset (BDS), and potential entrants in the model.



Note: The figure illustrates potential links between the BDS, the BFS datasets, and the potential entrants that could potentially choose to delay entry. Segment 1 corresponds to potential entrants who decide to delay entry and do not apply for the EIN. Segment 2 characterizes potential entrants who apply for the EIN, decide to delay entry, and never start a business. Finally, segment 3 represents a group of potential entrants that apply for the EIN and choose to wait for a year and enter the market with one year delay.

Even after considering the subset of the applications with higher rates of employer business births – business applications with planned wages, business applications from corporations, high-propensity business applications, their transition rate does not exceed 36%. Bayard et al. (2018) claim that a significant share of the business applications ends up becoming non-employer businesses.

Finally, note that by combining information in the BFS and BDS dataset, I can follow the pre-entry and post-entry decisions made by the third group of entrants. Specifically, I can use the variation in the time it takes to become employer businesses for the third group of entrants to identify delays in potential entrants' entry decisions.

A.3.4 Robustness: Annual Data

In this section, I conduct the same analysis using an annualized time series about business formation. I construct the following time series: The annual number of applications that form businesses within a year ($BF1Y$); The annual number of applications that form businesses within two years ($BF2Y$); The annual number of applications that form businesses with one year delay ($BF2Y$); The share of the business applications that form businesses with one year delay ($Share$).

To be consistent with the BDS, I construct annual data by summing up $BF4Q$ and $BF8Q$ time series from the second quarter of the year $t-1$ to the first quarter of the year t. $BF1Y$

Table 13: Summary statistics (in thousands)

	Mean	SD	Min	Max
Firms (BDS)	491.5	70.8	417.2	610.0
BF in 2 years	376.0	62.5	330.8	505.9
First year	326.3	59.7	281.6	462.2
Second year	51.2	4.80	43.70	59.40

Table 14: Correlations between the business applications and aggregate conditions at entry

	X / Corr()	$(X_{hp,t}, Y_{hp,t})$	$(X_{lin,t}, Y_{lin,t})$	$(\Delta X_t, \Delta Y_t)$	$(X_{hp,t}, \Delta u_t)$
Panel A	BF within 2 years (p_val)	0.69 (0.03)	0.75 (0.01)	0.63 (0.07)	-0.71 (0.02)
Panel B	First year (p_val)	0.78 (0.01)	0.84 (0.00)	0.67 (0.03)	-0.83 (0.01)
	Second year (p_val)	0.94 (0.00)	0.95 (0.00)	0.77 (0.02)	-0.98 (0.01)
Panel C	Share (p_val)	-0.83 (0.00)	-0.84 (0.00)	-0.74 (0.02)	0.70 (0.03)

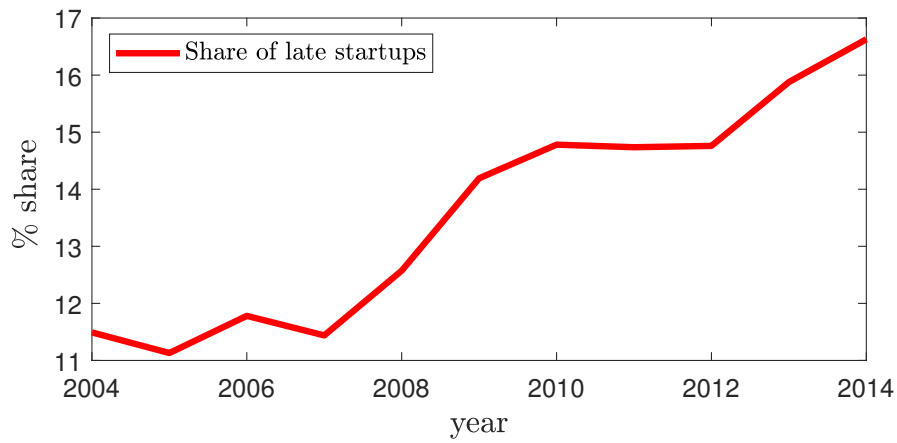
covers the period $2006-2016$, and the time series for $BF2Y$ covers the period $2006-2015$.[39] The summary statistics for the annual time series are given in Table 13. For comparison, the table also reports summary statistics for the employer business start-ups from the BDS dataset.

Cyclicality of the business formation at annual frequency Next, I study the cycle properties of the annual business formation data. Table 14 reports the results. The results imply that the number of applications that form business within a year decreases during the recessionary periods. The subset of the applications that form businesses a year to form a business also decreases during recessions. The share of the applications that form a business with one-year delay increases.

[39] $BF4Q$ and $BF8Q$ data starts the year $2004Q4$. Since I do not have the complete number of applications for the year 2005, I had to drop them from the analysis.

A.4 The Great Recession and the Share of Late Start-ups

Figure 23: Share of late startups



B Model Appendix

In Section B.1, I present an extended description of the entry phase that justifies the assumption about the constant mass of potential entrants. In Section B.2, I describe results from a model that allows the accumulation of potential entrants who delayed entry. In Section B.3, I present a general equilibrium version of the model.

B.1 Extension: Two-Stage Entry Phase

Every period, there is a limited mass of heterogeneous business opportunities that potential entrants can use to enter the market. These business opportunities are characterized by signal q. The signal describes the productivity of a business opportunity after it is implemented in the market. For a given signal q the distribution of the initial period productivity is given by $H_e(s|q)$. The higher the signal, the higher the expected first-period productivity of a business opportunity. The distribution of business opportunities over the signal is time-invariant and is given by $q \sim W(q)$.[40]

Analyzing the Business Formation Statistics dataset shows that, on average, only 14% of the business applications end up becoming employer start-ups. Using this information, I extend the entry phase and model an additional stage that decomposes entrants between aspiring start-ups – those that wish to be entrepreneurs and potential entrants who hold business ideas and enter the market.

[40]The distribution is such that the mass of business opportunities with signal q decreases with q.

The entry phase consists of two stages. During the first stage, an infinite mass of individuals makes decisions about whether to compete or not for the available business opportunities. Individuals need to pay a fixed cost, c_q, to participate in the competition. After which they are free to direct their search for a particular group of business opportunities characterized by a signal q. Since there are a limited number of business opportunities within each signal category, not all aspiring startups receive a signal. During the second stage, those aspiring startups that receive a signal about business opportunities become potential entrants and make entry decisions. The signal is persistent over time, which gives a potential entrant the ability to exercise the business opportunity in the future instead of today.

In what follows, I describe each phase in detail.

Stage 1. The expected value of attempting to seize a business opportunity with a signal q equals to

$$V^o(q, z) = \frac{B_t(q)}{n_t(q)} V^e(q, z_t) - c_q,$$

where $B_t(q)$ is a mass of available business opportunities with quality q at time t.[41] The total mass of available business opportunities equal to the total number of business opportunities within each signal category minus the mass of ideas that is already in the hands of entrants that delayed entry in the previous periods. $n_t(q)$ refers to a number of aspiring startups competing for the business opportunities with the signal q. The ratio in the equation represents a probability by which an individual receives a signal q and becomes a potential entrant.[42] $V^e(q, z_t)$ is a value of a potential entrant with signal q at time t.

If $V^e(q, z_t) < c_q$, individuals do not compete for the business opportunities with signal q. A positive mass of individuals decide to pay fixed cost c_q and compete for a business opportunity with signal q if $V^e(q, z_t) > c_q$. Due to the free entry, the number of individuals $n_t(q)$ competing for each signal q is such that $\frac{B_t(q)}{n_t(q)} V^o(q, z_t) = c_q$.

Denote $\underline{q_t}$ a signal at time t that satisfies $V^e(\underline{q_t}, z_t) = c_q$. Since the value of entry increases with a signal level, aspiring startups choose to compete for the business opportunities with signal level $q > \underline{q_t}$. The total number of individuals attempting to get the business opportu-

[41] $0 < B_t(q) < W(q)$

[42] $0 \leq \frac{B_t(q)}{n_t(q)} \leq 1.$

nities equals

$$N_{t,\text{aspiring startups}} = \int_{\underline{q_t}} n_t(q)dq.$$

Note that while $\underline{q_t}$ is weakly countercyclical (the higher the aggregate demand level, the higher the expected value of entry for all q), the cyclical variation of $N_{t,\text{aspiring startups}}$ depends on the available business opportunities at time t that is determined by the states in the past period.

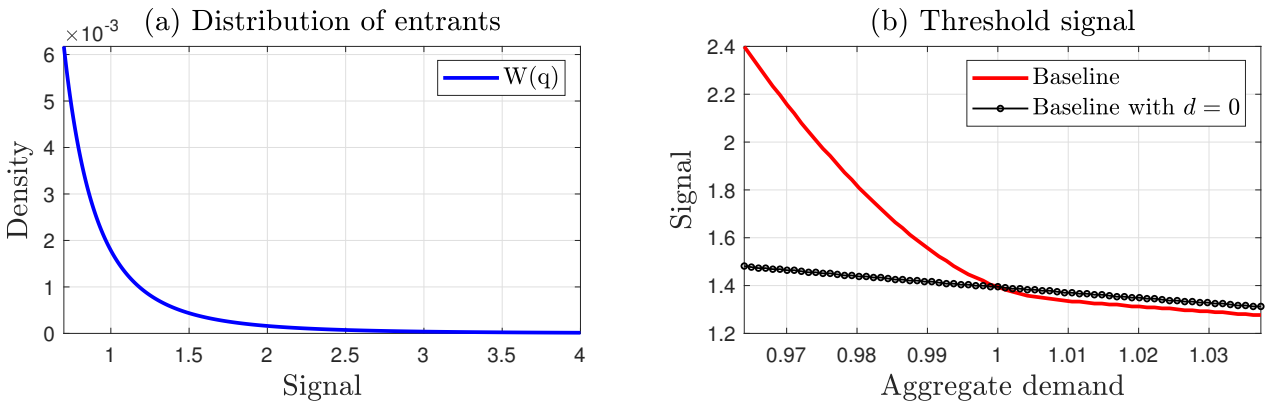
Stage 2. Stage 2, in which potential entrants make entry decisions, follows the same process as described in the 3.1.2.

Calibration To parameterize the entry phase, I use the Business Formation Statistics dataset. As I have noted before, the transition rate from application to business formation is around 12% and 14% within one and two years from the date of the application, respectively. In terms of the model, I consider the number of applications as the number of aspiring start-ups. I choose c_q, the fixed cost that individuals need to pay to become aspiring start-ups so that the share of the actual entrants in the total number of aspiring start-ups is 13%. The value corresponds to $c_q = 0.022$.

The data also shows that only an additional 2% of applications transition into employer businesses with one year delay. In terms of the model, the fact implies that $B(q)$ is close to $W(q)$; only a few potential entrants who choose to delay entry come back to the market next period. The ability to delay entry is an option for a potential entrant and does not require the potential entrant to enter the market in the future. Explaining the reasons behind what makes potential entrants actually come back or not come back in the market after delaying entry is beyond the scope of this paper and is left for future research.

Interestingly, the two-stage entry phase can also be used to reconcile the low transition rates from the business applications to employer businesses observed in the BFS data. In particular, the framework differentiates aspiring start-ups – those who want to start businesses and apply for the EIN, from those who actually hold business ideas and make entry decisions. According to the model, the restricted number of actual business ideas does not allow most aspiring start-ups to enter the market.

Figure 24



B.2 Extension: Model with Signal Accumulation

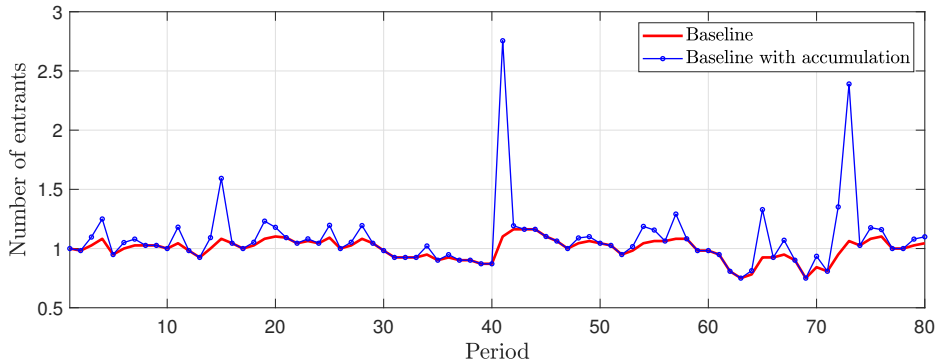
In this section, I relax the assumption that keeps the aggregate distribution of potential entrants constant in the baseline model. I investigate how the accumulation of potential firms modifies cohorts' characteristics over the cycles. I show that the recessionary cohorts have significantly and persistently different characteristics compared to their expansionary counterparts, even after allowing the accumulation of potential entrants over time.

In the baseline model, the aggregate distribution of potential entrants over the signal is time-invariant and is given by $W(q)$. In this section, I relax the assumption in the following way. At the beginning of every period, a constant mass of new potential entrants is born and make entry decisions. The distribution of new potential entrants over the signal is given by $W(q)$, see Figure 24(a). In addition to the new potential entrants, the aggregate distribution also consists of old potential firms. Old potential entrants are the ones who chose to delay entry in the previous periods even though their NPV of entry was non-negative. Figure 24 (b) displays the threshold signal, $\hat{q}_d(z)$ for each aggregate state when $d = 0$ (blue-dashed line) and $d = 1$ (solid red line). For given z, potential entrants that decide to delay entry hold signals in between $[\hat{q}_{d=0}(z), \quad \hat{q}_{d=1}(z)]$.

The distribution of old potential entrants evolves endogenously and depends on the realization of the aggregate states in the previous periods. Denote the mass of old potential entrants with signal q at the beginning of period t with $\Lambda_t^{\text{old entrants}}(q)$.

$$\Lambda_{t-1}^{\text{old entrants}}(q) = \sum_{k=0}^{t} W(q)\ 1\left\{\hat{q}_{d=0}(z_k) \leq q < \hat{q}_{d=1}(z_k)\right\} + \Lambda_0^{\text{old entrants}}(q),$$

Figure 25: Dynamics of entrants with and without accumulation



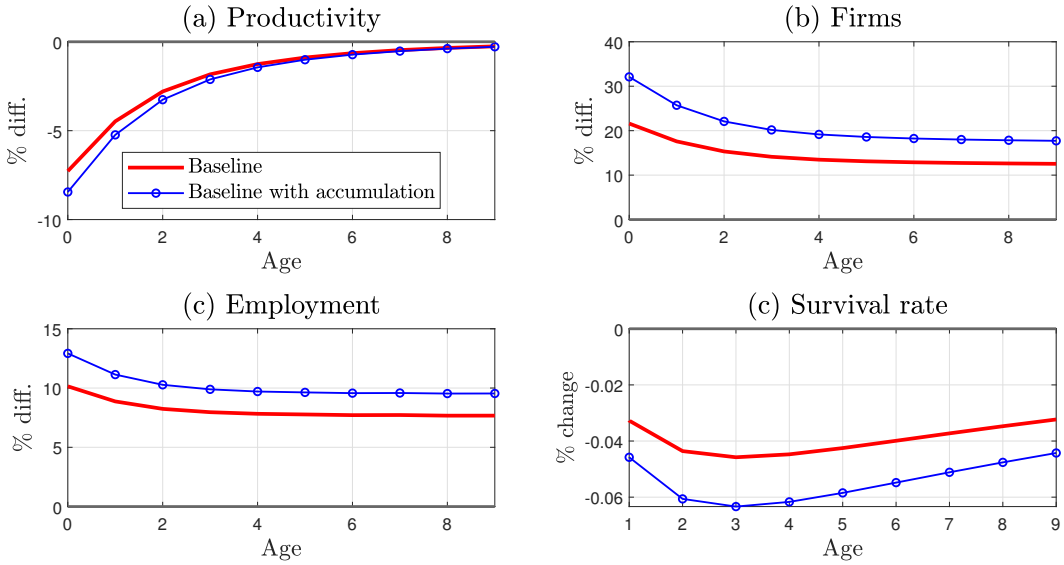
where $\Lambda_0^{\text{old entrants}}(q)$ denote the distribution of old potential entrants at time 0.

Then, the total mass of potential entrants with signal q at the beginning of period t, $\Psi_t(q)$ is given by

$$\Psi_t(q) = W(q) + \Lambda_t^{\text{old entrants}}(q).$$

Figure 25 compares the dynamics of entrants in the baseline and the alternative scenario with entrant accumulation. Note that when the aggregate demand decreases from z_{t-1} to z_t, these two scenarios coincide with each other. However, if the aggregate demand level increases from period $t-1$ to period t in addition to new potential entrants, some of the old potential entrants also decide to enter the market, resulting in a higher number of entrants to the model with signal accumulation compared to the baseline model.

Figure 26: Differences between the expansionary and recessionary cohorts' life cycle characteristics: Baseline with signal accumulation



It turns out that the alternative model produces an entry rate that has a higher variance and is more procyclical. Moreover, Figure 26 describes the differences between the expansionary and recessionary cohorts life cycle characteristics. The figure shows that allowing accumulation of potential entrants over time leads to cohorts with significantly and persistently different characteristics based on the business cycle conditions at entry. That is, cohorts born during recessions are more productive and have higher survival rates. However, they consist of fewer firms and employ fewer workers at entry and over time. Interestingly, the differences are even larger compared to the baseline model, where I did not allow for the signal accumulation.

B.3 General Equilibrium Framework

In this section, I extend the model to the general equilibrium framework. Note that the model presented in the main body of the paper is a reduced form of a general equilibrium model with infinitely elastic labor supply $\chi(L_t) = \psi L_t$ and where the demand of aggregate consumption basket is given by $P_t = C_t^{\rho}$.

B.3.1 Consumers

The economy is populated by a unit mass of atomistic, identical households. At time t, the household consumes the basket of goods C_t, defined over a continuum of goods Ω. At any given time t, the only subset of goods $\Omega_t \subset \Omega$ is available. Let $p_t(\omega)$ denote the nominal price of a good $\omega \in \Omega_t$.

First layer maximization:

$$\max_{\left(C_t, L_t, (c_t(\omega))_{\omega \in \Omega_t}\right)_{t=0}^{\infty}} E_0 \left[\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\sigma} - 1}{1-\sigma} - \chi(L_t) \right],$$

such that

$$P_t C_t = P_t w_t L_t + \Pi_t.$$

Second layer maximization:

$$\max_{(c_t(\omega))_{\omega \in \Omega_t}} C_t = \left(\int_{\omega \in \Omega_t} (\alpha z_t)^{\frac{1}{\rho}} b_t(\omega)^{\frac{\gamma}{\rho}} c_t(\omega)^{\frac{\rho-1}{\rho}} d\omega \right)^{\frac{\rho}{\rho-1}},$$

such that

$$\int_{\omega \in \Omega_t} p_t(\omega) c_t(\omega) d\omega \leq P_t C_t.$$

B.3.2 The Mutual Fund

The household owns shares in the mutual fund. The mutual fund consists of heterogeneous incumbent firms and new entrants. The mutual fund collects profits from all active firms at the end of the period and allocates dividends to households based on their shares. Description of the incumbent firms and potential entrants are similar to the baseline model. Except, I modify parts of the value functions to include aggregate prices and stochastic discount factors. The timing is shortly summarized below.

Incumbent Firms Incumbent firms are distributed over consumer capital (b) and productivity (s). The distribution given by $\Omega_t(s, b)$. At time t, for given aggregate demand level z, an incumbent firm characterized by (s, b) takes solves the following functional equation, while taking as given real wage w and the aggregate price index P.

$$V^I(b, s, z) = \max_{y,p,b} \quad py - Pwn + \int max \left\{ 0, -Pc_f + \tilde{\beta}(1-\gamma) E[V^I(b', s', z')|s, z] \right\} dG(f),$$

s.t.

$$y_t^s = s_t n_t;$$

$$y_t^d = \alpha z_t b_t^\eta \left(\frac{p_t}{P_t} \right)^{-\rho} Y_t;$$

$$b_{t+1} = (1-\delta) \left(\ b_t + y_t p_t \ \right);$$

$$c_f \sim G(f), \ c_f \text{ is in consumption units:};$$

$$log(s_{it}) = \rho_s log(s_{it-1}) + \sigma_s \varepsilon_{it};$$

$$log(z_t) = \rho_z log(z_{t-1}) + \sigma_z \epsilon_t.$$

Potential Entrants Potential entrants are endowed with signal, q that characterize their initial productivity. At any t, density of potential entrants over q is constant and is given by $W(q)$. To enter into the market the potential entrant needs to pay fixed entry cost in consumption units c_e (value $P_t c_e$). Upon entry the potential entrant observes actual idiosyncratic productivity (s), receives fixed initial capital stock (b_0) and behaves like an incumbent with (b_0, s).

At time t, for the given aggregate demand level z, aggregate price P and real wage w potential entrants solve the following problem:

$$V^e(b_0, q, z) = max \left\{ \tau \ \tilde{\beta} E[V^e(b_0, q, z')|z], \quad -Pc_e + \int_s V^I(b_0, s, z) dH_e(s|q) \right\}.$$

Value of the Mutual Fund The value of mutual fund, Λ_t at the beginning of time t, after entry and exit has occurred:

$$\Lambda_t = \int_s \int_b V(s, b, z)\Omega(b, s, z) ds db + \int_{q_*}^{\infty} \int_s V(b_o, s, z) H(s|q) W(q) dq.$$

Denote $N_{e,t}$ be the number of entrants in period t, than: $N_{e,t} = \int_{q_*}^{\infty} W(q) dq$. At the end of the period value of mutual fund is

$$\Lambda'_t = \Pi - N_{e,t} c_e + (\Lambda_t - \Pi).$$

Let $x_t \in [0, 1]$ was the share household decides to hold of the mutual fund in period t. Then, household budget constraint will be

$$\Lambda_t x_t + C_t \leq [\Pi - N_{e,t} c_e + (\Lambda_t - \Pi)] \, x_t + L_t P_t w_t.$$

The optimal solution implies that if $\Lambda_t \geq 0$ then $x_t = 1$. The latter reduces HH budget constraint to

$$P_t C_t + P_t N_e c_e = P_t w_t L_t + \Pi_t.$$

B.3.3 Discussion

In general equilibrium, both wages and the stochastic discount factor become procyclical (Hong, 2018). The procyclical discount factor makes delay favorable, since potential entrants give more weight to high aggregate demand conditions. The procyclical variation in wages makes delay less favorable during recessionary periods. However, the option value of delay is always non-negative due to entrants' ability to get an outside option by not entering the market. As a result, for any initial aggregate states the threshold value of the entry is weakly higher in the model with persistent signal compared to the models without persistent signals.

C Numerical Solution

The following section describes the numerical solution algorithm used to solve the model.

C.1 Incumbent's Value Function

1. Define grid points for the state variables s, z, and b.

   (a) The grids and the transition matrices for the idiosyncratic productivity shock s and the aggregate demand shock z are constructed following the Rouwenhorst (1995)'s method. Denote the number of grid points as I_s and I_z, and the probability transition matrices as $P^s(s'|s)$ and $P^z(z'|z)$, respectively.

   (b) To construct grid points for the customer capital I use equally distributed grid points on a logarithmic scale on the interval $[b_0, \quad b_{max}]$. I choose b_0 to match entrants' average size. I choose b_{max} so that employment by large firms is more than 1000+. The latter corresponds to the highest size bin in the BDS dataset. Denote the number of customer capital grid points as I_b.

2. For all the grid points (b, s, z), guess the incumbent firm's value function $V_0^I(b, s, z)$.

3. Construct a revised guess for the value function $V_1^I(b, s, z)$ by solving:

$$V_1^I(b, s, z) = \max_b \left\{ \Pi(b, s, z) + G(c_f^*) \left(\beta(1-\gamma) E[V_0^I(b', s', z')|s, z] - E[c_f | c_f < c_f^*] \right) \right\},$$

   subject to

$$\Pi(b, s, z) = \left(\frac{b'}{1-\delta} - b \right) - \frac{w}{s} \left(\frac{b'}{1-\delta} - b \right)^{\frac{\rho}{\rho-1}} b^{\frac{-\eta}{\rho-1}} \ (\alpha z)^{\frac{-1}{\rho-1}},$$

$$E[V_0^I(b', s', z')|s, z] = \sum_i \sum_j V_0^I(b', s_i, z_j) P^z(z_j|z) P^s(s_j|s),$$

   where $P^z(z_j|z)$ and $P^s(s_j|s)$ represents probabilities that next periods aggregate shock equals to z_j and idiosyncratic shock equals s_j. c_f^* is the value of the fixed cost which equals to incumbent's expected continuation value $\beta(1-\gamma)E[V^{I*}(b', s', z')|s, z]$. In other words, when an incumbent firm receives c_f^*, the incumbent firm is indifferent between staying or exiting from the market.

4. Stopping criteria: $\left| \frac{V_{n+1}^I(b, s, z) - V_n^I(b, s, z)}{V_n^I(b, s, z)} \right| \leq 10.0^{-8}$.

C.2 Potential Entrants' Distribution

Figure 27: $W(q)$

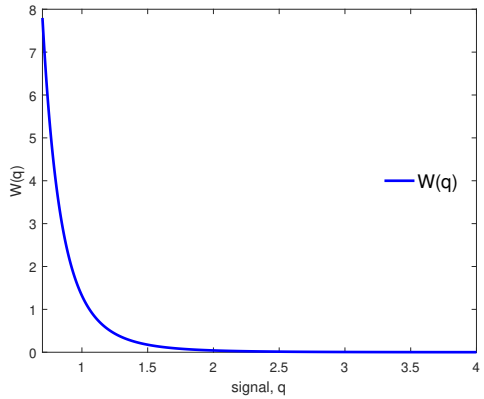
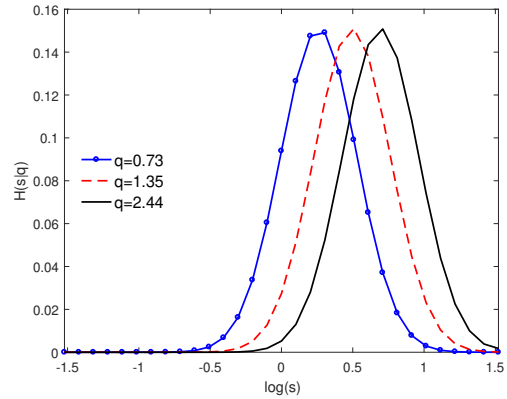


Figure 28: $H(s|q)$



1. Denote the number of grid points for the signal by I_q. I use Gauss-Legendre quadrature method over the interval $[\underline{q},\ q_{max}]$ to generate grid points q and weights w_q for the signal.

2. The aggregate signal distribution $W(q)$ has Pareto Distribution with a location Parameter of $\underline{q}$ and Pareto exponent ξ. I approximate the mass of potential entrants denoted by P_q, at each grid point of signal according to the following equation:

$$P_q(q) = w_q(q)\ \xi \frac{\underline{q}^{\xi}}{q^{\xi+1}}.$$

3. I construct the distribution for the initial idiosyncratic productivity $H(s|q)$ as follows:

   The idiosyncratic shock in the first period of operation follows the normal distribution. For each grid point $q_j \in I_q$ and $s_i \in I_s$, I calculate $F(s_i|q_j)$, the probability that the entrant with signal q_j gets the initial productivity lower than s_i as follows:

$$H(s_i|q_j) = \frac{1}{2}\left(F(s_i|q_j) - F(s_{i-1}|q_j)\right) + \frac{1}{2}\left(F(s_{i+1}|q_j) - F(s_i|q_j)\right).$$

   I construct the initial and the terminal grid points of the productivity based on the following function:

$$H(s_1|q_j) = F(s_1|q_j) + \frac{1}{2}\left(F(s_2|q_j) - F(s_1|q_j)\right),$$

$$H(s_{I_s}|q_j) = \max\left(0, 1 - F(s_{I_s}|q_j)\right) + \frac{1}{2}\left(F(s_{I_s}|q_j) - F(s_{I_s-1}|q_j)\right).$$

I denote the final value function by $V^I(b, s, z)$.

C.3 Entrant's Value Function

1. For all grid points (q_j, z_k) I calculate the gross value of entry as

$$V^{gross}(b_o, q_j, z_k) = \sum_{i \in I_s} \left[H(s_i|q_j) V^I(b_o, s_i, z_k) \right].$$

2. To approximate the entrant's value function and the option value of delay, I use the value function iteration algorithm described below:

   (a) Guess for the values of the entrant value function. $V_0^e(b_0, q, z)$

   (b) Given the guess find value of the option value of delay.

$$V^{Opt}(q, z) = \tau \ \beta E[V_0^e(b_0, q, z')|z] = \tau \ \beta \sum_{z_j \in I_z} V_0^e(b_0, q, z_j).$$

   (c) Update guess for value function of entry.

$$V_1^e(b_0, q, z) = \max \left\{ V^{Opt}(q, z), V^e(b_0, q, z) - c \ \right\}.$$

   (d) Stopping criteria: $|\frac{V_{n+1}^e(b, s, z) - V_n^e(b, s, z)}{V_n^e(b, s, z)}| \leq 10.0^{-8}$.

   Denote the final entry value function by $V^e(b_0, q, z)$ and the final option value of delay function as $V^{Opt}(q, z)$.

D Calibration Appendix

D.1 The Sources of Data

The BDS dataset covers the universe of employer businesses in the US and provides yearly measures of business dynamics for the US economy aggregated by the establishment and firm characteristics. For more information see the link. The establishment is defined as a single physical location, whereas the firm is defined at an enterprise level. The data report establishment/firm-level activity based on the employment status on March 12. Specifically,

at year t, establishment- and firm-level activity describes the period from the second quarter of year $t-1$ through the first quarter of year t. For more information see the link.

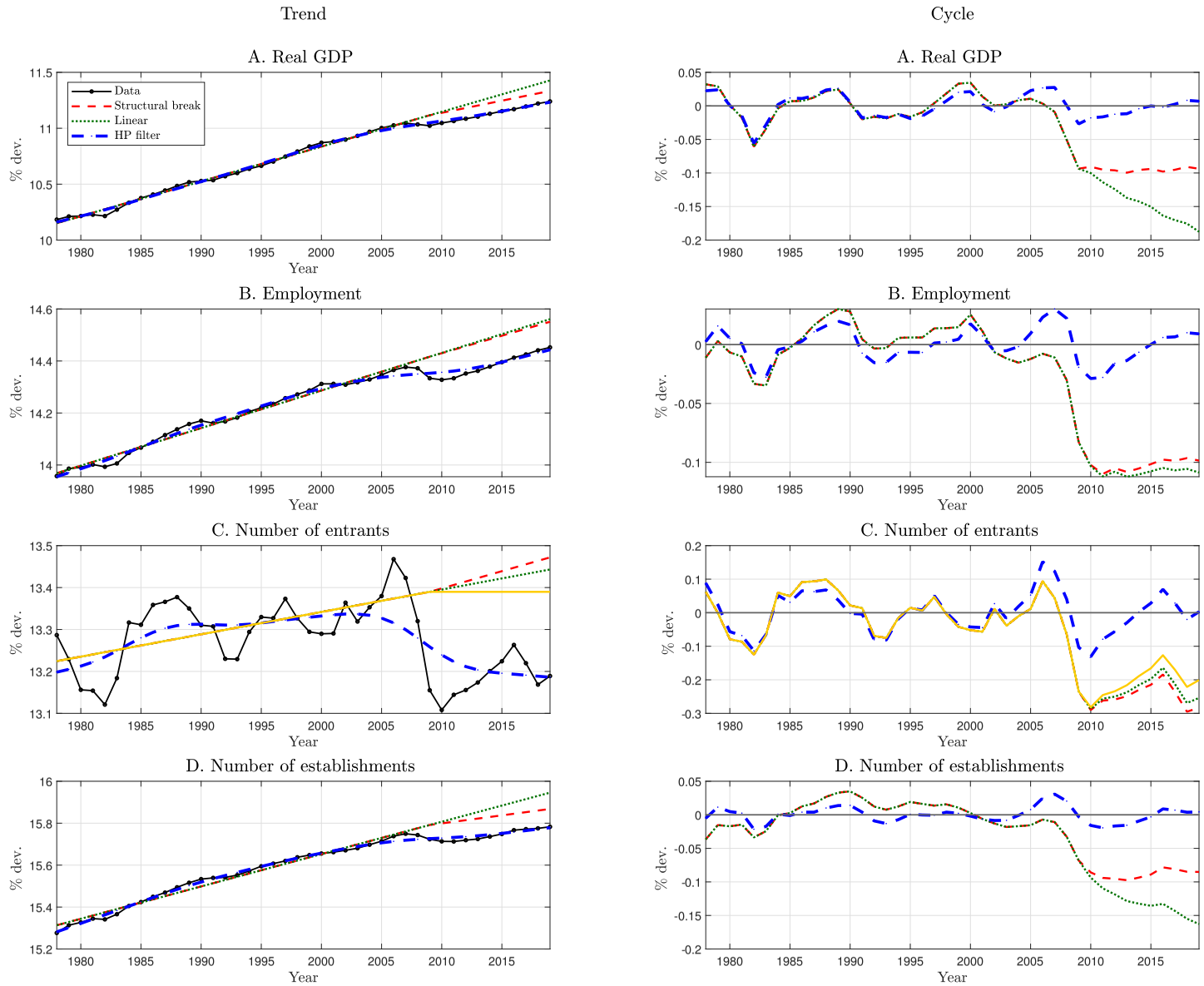
The BDS follows each cohort of establishments for up to 5 years. After five years, the dataset gives information in 5-year bins. More specifically, The data set characterizes cohorts within the following age groups $[0, 1, 2, 3, 4, 5, 6\text{-}10, 11\text{-}15, 16\text{-}20, 21\text{-}25, 26+]$.

Aggregate Time Series To measure aggregate activity I use time series for real GDP, and aggregate employment from the Federal Reserve Economic Data (FRED). While studying the business cycle conditions at entry, I consider the modified versions of the time series that is consistent with the BDS timing. Specifically, in the BDS, establishment-level and firm-level activity at year t covers the establishment activity from March of year $t-1$ to the March of year t. Thus, I construct the annual time series of the aggregate variables as March-to-March averages, to be consistent with the BDS dataset timing.

D.2 Detrending

Figure 29 illustrates the trend and cyclical component of aggregate variables after applying alternative detrending methods. Specifically, I compare the trend and cyclical component of log real GDP, aggregate employment, number of entrants, and number of establishments over the period 1978-2018 after applying (i) a linear trend (ii) a linear trend and a linear trend that allows a break point in trend (iii) the HP filter with smoothing parameter 100. Note that the HP-filters predicts that the Great Recession was a sharp downturn after which the economy recovered quickly. On the other hand, the linear trend exaggerates the severity of the recession.

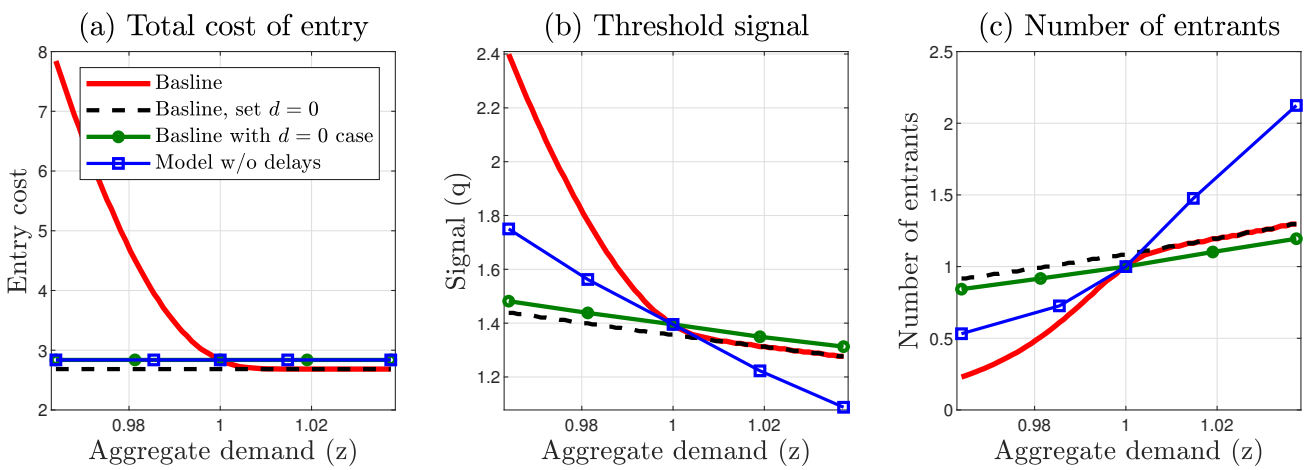
Figure 29


Trend
Cycle
A. Real GDP
B. Employment
C. Number of entrants
D. Number of establishments
Data
Structural break
Linear
HP filter
% dev.
Year

E Alternative Models

In this section, I describe in detail the construction of the alternative scenarios that I use to understand the role of the option to delay entry in driving the micro and macro level fluctuations. First, in Section E.1, I describe the baseline with $d = 0$ case that I use to study how much the option to delay entry amplifies and propagates aggregate shocks. Second, in Section E.2, I compare the performance of the baseline model to a workhorse firm dynamics model (model w/o delay), parameterized to account for the same set of facts.

Figure 30: Alternative scenarios



E.1 Baseline with $d = 0$

To isolate the role of the option to delay entry in the business cycle dynamics of the aggregate variable, I consider a version of the baseline model with $d = 0$ (*Baseline with $d = 0$ case*).

Setting $d = 0$ in the baseline model decreases the opportunity cost of entry by the amount of the option value of delay. As a result, compared to the baseline model, the threshold quality signal is lower in the baseline model with $d = 0$. In the steady state, the threshold signal uniquely determines the distribution of entrants over the initial productivity, which in turn can be mapped uniquely to the invariant firm distribution. Hence these two scenarios exhibit different dynamics in the steady state. To isolate the role of the option value of delay in the business cycle dynamics of entrants, I need to re-calibrate the baseline model with $d = 0$ to match the same set of facts in the steady state, as the baseline model.

$$V^{gross}(z_{ss}, \hat{q}_d) = c_e + dV^w(z_{ss}, \hat{q}_d) \tag{9}$$

Table 15: Calibration of alternative scenarios

Symbol	Description	Baseline	$d = 0$	Mode w/0 delay
β	Discount rate	0.960	0.960	0.960
ρ	Price elasticity of demand	1.622	1.622	1.622
η	Elasticity of demand to capital	0.919	0.919	0.919
δ	Depreciation rate of reputation	0.188	0.188	0.188
ρ_s	Idiosyncratic shock – persistence parameter	0.814	0.814	0.814
σ_s	Idiosyncratic shock – SD parameter	0.161	0.161	0.161
α	Demand shifter	0.261	0.261	0.261
b_o	Initial customer capital level	12.00	12.00	12.00
μ_f	Operating cost – SD parameter	0.621	0.621	0.621
σ_f	Operating cost – SD parameter	0.410	0.410	0.410
γ	Exogenous exit shock	0.071	0.071	0.071
$\underline{q}$	Pareto location	0.700	0.700	0.700
ξ	Pareto exponent	2.478	2.478	2.478
c_e	Fixed entry cost	2.684	2.837*	2.833*
ρ_z	Aggregate shock – persistence parameter	0.750	0.750	0.600*
σ_z	Aggregate shock – SD parameter	0.003	0.003	0.019*
d		1	0	0

The gross value of entry does not vary with d. The threshold signal depends on the total opportunity cost of entry $c_e^d + dV^w(z_{ss}, \hat{q}_d)$. For any d and d', equating the threshold signals in the stochastic steady state $\hat{q}_d(z_{ss}) = \hat{q}_{d'}(z_{ss})$ requires the opportunity cost of entry to equal each other across these scenarios. The value of the option is endogenously determined within the model, while the fixed cost of entry $c_{e,d}$ can be modified to generate the desired level of the threshold signal. Since the fixed entry cost only affects the selection of entrants at entry, equalizing the threshold signal implies that these scenarios will lead to the same dynamics in the stochastic steady state.

Following the argument, I set $c_e^{d=0} = c_e + V^w(z_{ss}, \hat{q}_{d=1})$ in the baseline model with $d = 0$. Figure 30(a) summarizes the difference in the fixed entry cost. The Column (b) of Table 15 summarizes the parameter values used in the baseline with $d = 0$ case and shows that the $d = 0$ case is identically parameterized except the fixed entry cost. Table 16 reports the steady state moments for the $d = 0$ case. Comparing the business cycle dynamics in the baseline model against the $d = 0$ case allows quantifying the role of the option to delay entry in accounting for the observed dynamics of the cohorts over the business cycles.

Table 16: Calibration targets and the model-implied counterparts

Statistics	Data	Baseline	$d = 0$	model w/o delay
Firm size	17.0	17.0	17.0	17.0
Firm size at entry	8.73	10.0	10.0	10.0
Firm size at age 5	13.9	14.5	14.5	14.5
Firm size at age 23	21.2	22.1	22.1	22.1
Employment share at entry	0.56	0.58	0.58	0.58
Firm exit hazard at age 5	0.10	0.09	0.09	0.09
Firm survival rate up to age 5	0.49	0.49	0.49	0.49
Firm survival rate up to age 23	0.15	0.12	0.12	0.12
Entry rate (%)	9.90	10.0	10.0	10.0

Note: The moments are calculated using the US-level cohorts of establishments from the BDS dataset covering the the period 1978-2019.

Table 17: Calibration targets for the aggregate demand shock process

Statistics	Data	Baseline	Baseline with $d = 0$	model w/o delay
Autocorrelation of establishments	0.70	0.72	0.72	0.74
SD of establishments	0.01	0.01	0.01	0.01
SD of entry	0.06	0.06	0.01	0.06

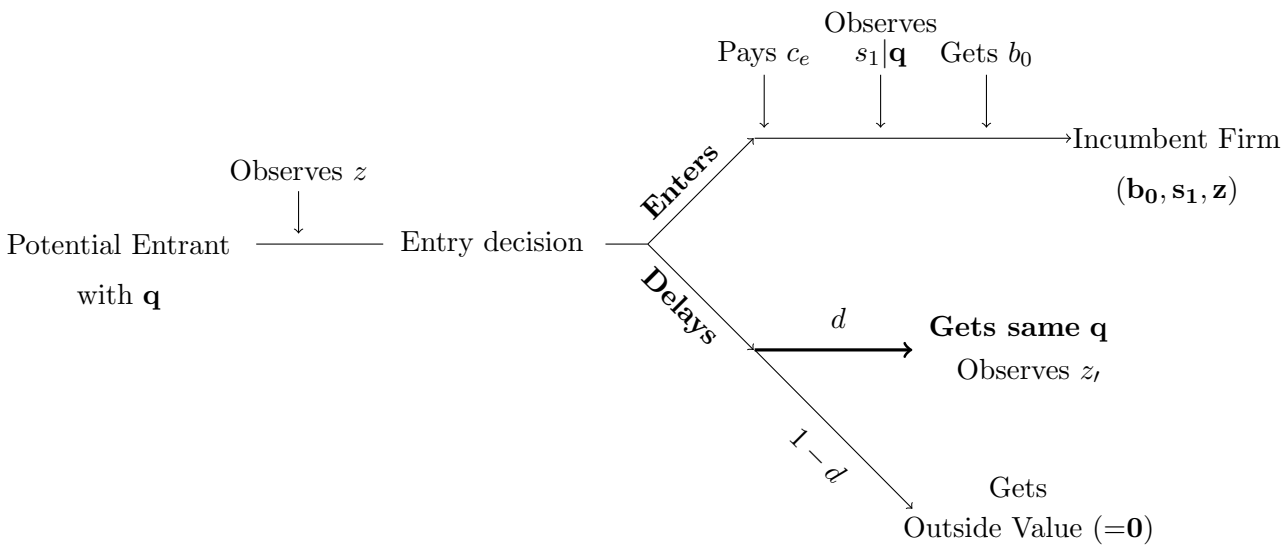
Note: The time series about the entry rate comes from the BDS and covers the period 1978-2019. The cyclical component of the log entry rate is calculated using the HP filter with smoothing parameter 100.

E.2 Model without Delay

A model w/o delay is a version of the baseline model with $d = 0$, calibrated to match the same set of facts described in Section 5.1. In this model, the entry decision follows a traditional neoclassical investment rule: enter if the NPV is non-negative. The entry cost is fixed and does not vary over the cycles (Figure 14a). Thus, the aggregate shock can affect the selection of entrants only through its direct effect on potential firms' lifetime profits. I find that producing the observed variation in the number and composition of entrants without the option-value channel requires the standard deviation and the autocorrelation of the aggregate demand shock to be 0.016 and 0.64, respectively. In the baseline model, the numbers are 0.003 and 0.75, respectively. In terms of the unconditional variance, the model w/o delay requires shocks with 5-times higher magnitudes to produce the observed variation in entry than the baseline model. To put it differently, the endogenous countercyclical variation in the cost of entry amplifies the elasticity of entrants to aggregate shocks 5-times. Table 15 summarizes parameter values, Table 16, and Table 17 summarizes how the moments targeted in the model w/o delay compares to the data counterpart and other scenarios.

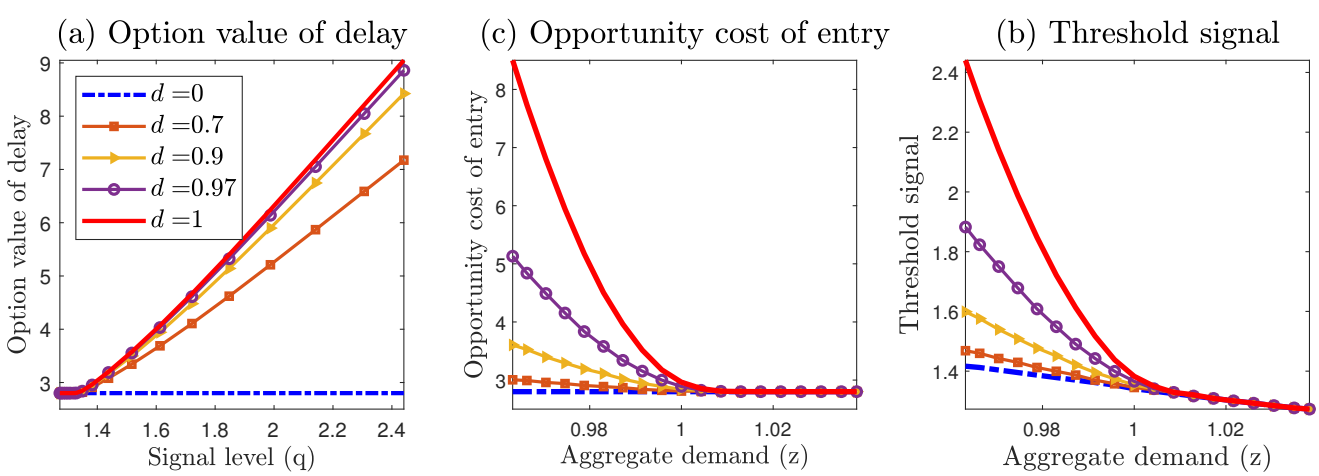
F The Probability of Keeping Signal

Figure 31: **Potential Entrants' Timing**



F.1 Comparative Statics

Figure 32: Selection of entrants across different aggregate conditions at entry



In this section, I investigate how the value of delay changes if potential entrants can only keep signals with some probability. That is, I allow $d \in [0, 1]$ in Equation (4), where d represents a probability that the potential entrants will carry signals tomorrow. Figure 32(a) displays the value of the option to wait for different values of d in the stochastic steady state. As

expected, the value of waiting decreases with d. Figures 32(b) and 32(c) show that the total opportunity cost of entry, as well as, the threshold signal level significantly increases with d.

Estimation Strategy for d The aggregate demand shock process affects incumbent firms' production and continuation decisions and potential entrants' entry decisions. On the other hand, the level of d affects only potential firms' entry decisions. Utilizing the differential effect, one can use the aggregate demand shock process and the probability of keeping signal to match the process of aggregate employment jointly with dynamics of entry and the total number of firms. Following the exercise, I find that $d = 0.976$.