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

Ia Vardishvili[1]

February 15, 2022

Click here for the latest version

Abstract

This paper studies the optimal timing of market entry decisions made by heterogeneous entrants and its implications on the selection of firms over the business cycles. I show that firms' option to delay entry, missing in existing frameworks, generates an endogenous countercyclical opportunity cost of entry, that significantly amplifies the effect of the initial aggregate economic conditions on selecting firms at entry. The mechanism works through procyclical variation in survival rates: during recessions, a higher risk of post-entry failure increases entry cost by increasing the value of delay. Firms wait until the expected survival rates are high enough to compensate for the lower expected post-entry profits. Using the Business Dynamics Statistics dataset, I document that consistent with the mechanism cohort of firms that enter the market during recessions have significantly higher survival rates compared to their expansionary counterparts. The option to delay entry enables existing models to reconcile the observed dynamics of entrants over the business cycles. Finally, I argue that not accounting for the option to delay entry may result in misleading predictions about entrants' responses to different shocks or policies.

Keywords: Option value, entry, firm dynamics, business cycles, propagation, Great Recession.

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, 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

Recent empirical findings document that aggregate economic conditions at inception have a significant and persistent effect on the US entrant establishments' life-cycle characteristics. Specifically, the number of entrants is procyclical and four times as volatile as aggregate employment.[2] Moreover, cohorts of establishments that start operating during recessions employ fewer workers at entry and over time, although they are, on average, more productive than expansionary cohorts.[3] A considerable body of theoretical and empirical microeconomics literature show that the option to wait to enter the market could significantly affect the potential firms investment decisions. However, this channel has not been studied in the existing firm dynamics literature. In this paper I explore the optimal timing of firms' market entry decisions over the business cycles and its aggregate implications.

I show potential entrants' ability to wait, missing in existing frameworks, significantly amplifies the effect of the initial aggregate conditions on selecting firms at entry. With the intertemporal choice, even a small shock that alters the relative lifetime benefits of entry today versus tomorrow has a substantial effect on firms' decisions to start a business. Without this channel, standard models cannot reconcile the observed variation in the entry margin since the lifetime profits are insensitive to aggregate shocks of reasonable magnitude.[4] This mechanism enables me to develop a model that closely accounts for the US establishments' life-cycle dynamics on average and over the business cycles. Using the framework, I quantify the role of the observed variation in the entry margin in shaping aggregate fluctuations. Finally, I argue that missing the option-to-delay channel may result in misleading predictions about the response of potential entrants to different shocks or policies.

I build a firm dynamics model with endogenous firm entry and exit and aggregate demand volatility. Heterogeneous firms operate in monopolistically competitive markets and make decisions about production and exit. Potential entrants hold heterogeneous signals about their post-entry initial productivity and make entry decisions. Upon entry, they pay the fixed entry cost and behave like incumbents. I deviate from the existing framework and

[2] Author's calculation using the establishment-level data from the Business Dynamic Statistics (BDS) dataset over 1977-2015.

[3] Moreira (2015) and Sedlacek and Sterk (2017) document that cohort-level employment is significantly and persistently procyclical. Lee and Mukoyama (2015), Moreira (2015), and Ates and Saffie (forthcoming) find that firms that are born during recessionary periods are, on average, more productive at entry and over time.

[4] The existing empirical microeconomics literature also supports the result and finds that the traditional entry decision rule does not explain much of the variation in the entry rate. For example, see O'Brien, Folta, and Johnson (2003). See Geroski (1995) for detailed discussion.

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 entry with respect to the 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 expected profits are more than the fixed entry cost.

To provide an empirical evidence of the option-to-delay channel, I study pre-entry and post-entry decisions made by firms in the US.[5] First, I use newly developed Business Formation Statistics (BFS) dataset to document that the part of the business cycle variation in the number of start-ups comes from the entrants' option to delay entry. Second, I use country- and state-level annual time series about the number of establishments/firms by age from the Business Dynamics Statistics (BDS) dataset to document that life cycle survival rates of cohorts of establishments/firms are negatively correlated with the aggregate conditions at the time of entry. In the model, the latter is a direct implication of the mechanism: firms wait until the expected survival rates are high enough to compensate the lower expected post-entry profits. Without the option to wait, the model leads to acyclical survival rates.

I parameterize the model using the establishment-level data over the period 1977-2015 from the BDS dataset. The calibrated model generates a close match to the US cohorts' average size, exit, and survival for up to 30 years of operation, and the share of cohorts' employment in aggregate employment for up to 5 years of operation. I parameterize the exogenous aggregate demand shock process to match the dynamics of the entry rate in the model and the data. Finally, I show that the calibrated model generates the documented persistent and significant differences in the life-cycle characteristics of cohorts that entered the market at different stages of the business cycle.

[5] A considerable body of existing theoretical and empirical microeconomics literature supports the finding that the option to wait profoundly affects entry decisions. For example, see Bernanke (1993), McDonald and Siegel (1986), and Pindyck (1991). See Dixit and Pindyck (1994) for a detailed review.

The option-to-delay channel is quantitatively important to account for the observed dynamics of entrants over the business cycles. The endogenous countercyclical opportunity cost of entry increases the variance of the number of entrants for a given aggregate demand shock process by seven times. The mechanism also leads to a significant variation in the composition of entrants. Specifically, due to the increased cost of entry, the group of firms that enter the market during recessions is, on average, more productive; however, the share of the high-survival, high-growth firms is lower in these cohorts due to the medium-productivity entrants who choose to postpone starting a business. The latter channel persistently decreases the recessionary cohorts' employment. I show that without the option to delay entry, the productivity composition of entrants, and the cohort-level employment vary little over the business cycles.[6]

Utilizing the good fit of the model, I quantify the role of the observed demographics of entrants in shaping aggregate fluctuations. I find that a model built to reconcile the life-cycle dynamics of U.S. establishments generates more than three-fourths of the observed persistence and variance of the aggregate employment. I show the variation in the number and the composition of firms at entry, which leads to the persistent procyclical variation in cohort-level employment, is responsible for shaping the aggregate fluctuations.[7]

The result seems surprising when compared with a small share of entrant cohorts' employment in aggregate employment. To support the finding and validate the model, I study the Great Recession, which is notorious for the historical drop in entry and the unprecedented slow recovery. Using the BDS data, I show that cohorts of firms that started operating over 2008-2016 employed persistently fewer workers, which cumulatively accounts for 45% of the drop and more than 85% of the slow recovery in aggregate employment.[8] Next, I repeat the exercise using the model simulated data.[9] I find that the baseline framework closely replicates the result and explains the drop in cohort-level employment through the variation

[6]This mechanism speaks for the empirical findings by Pugsley, Sedlaćek, and Sterk (2016), who show that ex-ante variation in the types of entrants explains most of the differences in cohorts' post-entry performance; Haltiwanger et al. (2013), Decker et al. (2014), and Haltiwanger et al. (2016) stress the importance of the share of the high-growth firms in a cohort for aggregate job creation.

[7]The implications are consistent with the empirical findings by Sedlaćek, and Sterk (2017), who show that the selection of firms at the entry stage, rather than the post-entry choices made by the firms, drive the cohorts' contribution to aggregate fluctuations.

[8]Gourio, Messer, and Siemer(2016) and Sedlaćek (2019) use data over 2008-2012 and study how the persistent drop in the number of entrants contributes to the aggregate dynamics. In my exercise, I concentrate on changes in cohort-level employment rather than the number of entrants.

[9]Specifically, I study the response of the baseline economy to a shock process that matches entrants' dynamics over 2008-2016.

in the number and composition of firms at entry.

Next, I show that firm dynamics models that employ a traditional entry decision rule are not able to account for the observed dynamics of entrants without generating excessive aggregate fluctuations. I consider a version of the baseline model without the option to delay entry, parameterized to account for the same set of facts. For the calibrated aggregate demand shock process, the model leads to the variance of the aggregate employment that is 1.7 times larger than the data counterpart. To put the number into perspective, I illustrate that a shock series that match the dynamics of entrants over 2008-2016 lead to a decline in aggregate employment that is twice as large as that observed during the Great Recession.

Finally, I argue that potential entrants' ability to postpone entry also qualitatively alters existing models' implications about the response of entrants to different shocks or policies. With the intertemporal choice, the dynamics of entrants depend on the changes in the relative benefits of entry today versus tomorrow, whereas in standard frameworks the entry decisions depend only on the expected post-entry profit today. Indicating that not accounting for the option to wait may lead to imprecise predictions about the response of potential entrants to various shocks, depending on their magnitude, timing, and duration. I illustrate the point by contrasting the response of entrants to a permanent, temporary, and future reduction in entry cost with and without the option to postpone entry.

Relation to The Literature This paper mainly contributes to three strands of literature.

First, it contributes to the firm dynamics literature that studies the significant and persistent effect of the aggregate economic conditions on the selection 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's (1993) framework is a puzzle that can be solved by introducing an entry cost that varies over the cycles in a particular way. Sedlaćek and Sterk (2019) introduce entry function, which enables the model to account for the elasticity of entrants with respect to the aggregate shocks. Others rely on exogenous entry-specific shock processes (e.g., Clementi and Palazzo (2016), Sedlaćek and Sterk (2017)). I show that these models can be reconciled with the data by allowing potential entrants to postpone starting a business. The additional selection generated through the option to delay entry also complements the literature that use "missing generation" mechanism (e.g.,Gourio, Messer and Siemer (2015), Clementi and Palazzo (2016)) and demand-side factors (Sedlacek and

Sterk (2017), and Moreira (2015)) to explain the persistent procyclical variation in cohorts' employment.

Second, the paper contributes to a large body of theoretical literature that studies the role of endogenous entry and exit in the amplification and propagation of aggregate shocks. Samaniego (2008) finds that aggregate fluctuations are insensitive to entry and exit, whereas Lee and Mukoyama (2008), Bilbiie, Ghironi, and Melitz (2012), Clementi, Khan, Palazzo, and Thomas (2014), and Clementi and Palazzo (2016) find that endogenous dynamics in the entry and exit significantly shapes the dynamics of the aggregate variables. Recent empirical literature emphasizes the importance of the life-cycle demographics of entrants in measuring and understanding the contribution of the entry margin to aggregate fluctuations. Haltiwanger et al.'s (2013) findings show that young firms exhibit distinct life-cycle dynamics compared with their mature counterparts, and emphasize the importance of accounting for not only the entry process but also the subsequent post-entry dynamics (growth, survival, job creation). In this paper, I propose a model that closely accounts for the US establishments' life-cycle dynamics on average and over the business cycles. Using the framework, I revisit and fully quantify the role of the observed variation in the entry margin in shaping aggregate fluctuations.

Third, the paper links the real options literature to the firm dynamics literature. Specifically, the paper relates to a considerable amount of theoretical and empirical microeconomic literature that finds the ability to delay entry could profoundly affect entry decision under aggregate volatility.[10] Pindyck (2009) shows that various risks to post-entry profits could magnify the cost of entry, and have a profound effect on firm dynamics. I additionally find endogenous variation in the risk of post-entry failure increases the entry threshold. This paper also relates to the theoretical macroeconomics literature that studies the role of real options in shaping aggregate dynamics (e.g., Jovanovich (1993), Veracierto (2002), Bloom (2009)). I contribute to the literature by extending the analysis on an entry margin. I find that the option to delay entry significantly amplifies and propagates aggregate shocks by affecting the number and composition of entrants. In that respect, the paper also relates to the literature that points out the weak internal propagation mechanism of standard business cycle models (e.g., Cogley and Nason (1995), King and Rebelo (1999)).

Finally, this paper is also related to the theoretical and empirical literature that studies the

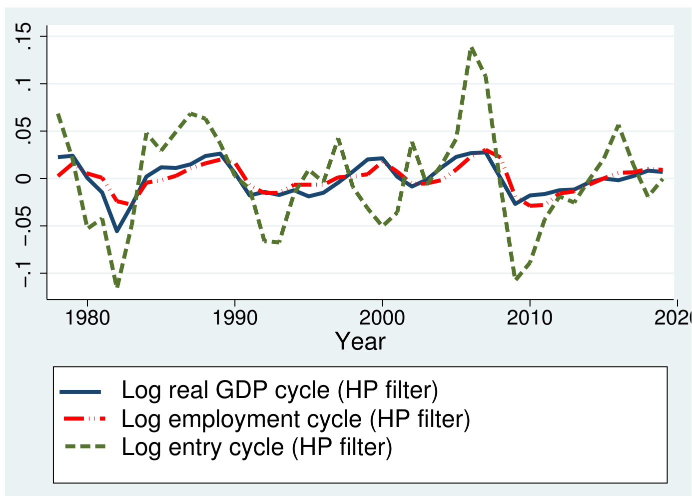
[10]For example, see Bernanke (1993), McDonald and Siegel (1986), and Pindyck (1991). See Dixit and Pindyck (1994) for a detailed review.

causal relationship between the significant and persistent drop in the entry rate and the slow recovery in aggregate employment observed after the Great Recession (e.g., Gourio, Messer and Siemer (2016), Sedlaćek and Sterk (2019), Siemer (2016), Clementi and Palazzo (2016), Khan, Senga, and Thomas (2016)).

2 Empirical Motivation

Figure 1 illustrates the business cycle dynamics of the log number of entrants together with the cycle component of log real GDP and aggregate employment. The table below describes the standard deviations of each of these time series and shows that the number of entrants is three times as volatile as real GDP and four times as volatile as aggregate employment.

Figure 1: Business cycle dynamics of the log number of entrants, log real GDP, and log Employment (HP filter, smoothing parameter 100)



	Standard deviation
Log real GDP cycle (HP filter)	.019
Log employment cycle (HP filter)	.015
Log entry cycle (HP filter)	.056

In addition to the variation in the number of entrants, recent empirical literature documents that the aggregate economic conditions at inception significantly affect the selection of firms at entry. Specifically, Moreira (2015) and Sedlacek and Sterk (2017) document that cohorts of establishments that start operating during recessions employ fewer workers at entry and over time. Lee and Mukoyama (2015), Moreira (2015), and Ates and Saffie (2021) find that firms that are born during recessionary periods are, on average, more productive over the life cycle.

The goal of the paper is to understand what accounts for the observed significant selection of entrants over the business cycles. A considerable body of theoretical and empirical microeconomics literature shows that the option to time the investment decision could significantly affect the potential firms' entry decisions.[11] At the same time, the literature points out that the conventional measure of entry decision – invest if the net present value of entry is more than zero – does not explain much of a variation in the entry rate because the variation in the expected stream of profits over time is minor.[12]

In the following section, I document that cohorts of new businesses that start operating during the recessionary periods are characterized with higher survival rates compared to their expansionary counterparts. Then, I document that part of the business cycle variation in the number of start-ups comes due to the timing of entry decisions. Later, in the paper, I use the heterogeneous firm dynamics model with endogenous entry and exit to show that firms' option to time their decision is one of the main drivers of the observed significant selection and variation of entrants over the business cycles.

2.1 Aggregate Conditions at Entry and Cohorts' Survival Rates

In this section, I document that life cycle survival rates of new businesses are negatively correlated with the aggregate conditions at the time of entry. I use country- and state-level annual time series about the number of establishments/firms by age from the Business Dynamics Statistics (BDS) dataset.[13] The dataset covers the period $1978 - 2019$.

I measure a survival rate of a cohort of age g at year t as

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

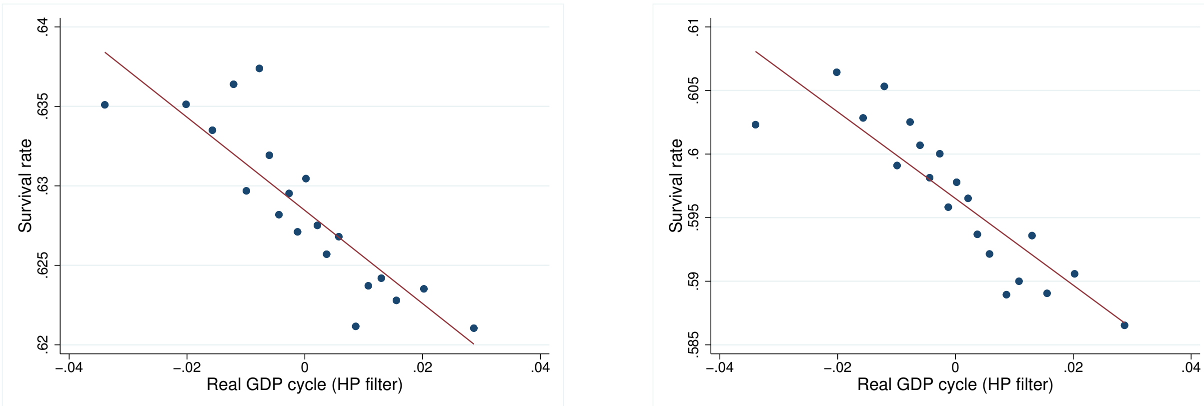
where $N_{g,t}$ measures the number of establishments (firms) in a cohort of establishments (firms) of age g at year t; $N_{0,t-g}$ measures the number of establishments (firms) in the same

[11] For example, see Bernanke (1993), McDonald and Siegel (1986), and Pindyck (1991). See Dixit and Pindyck (1994) for a detailed review.

[12] For example, see O'Brien, Folta, and Johnson (2003). See Geroski (1995) for a detailed discussion.

[13] The BDS dataset covers the universe of employer businesses in the US and provides annual measures of business dynamics for the economy 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 span multiple physical locations.

Figure 2: Cohorts' survival rates against the aggregate economic conditions at the time of entry (the US–level data)



(a) Cohort of establishments

(b) Cohort of firms

Note: Each panel plots a binned scatterplots of the survival rates up to age 5 against the aggregate conditions at the time of entry measured by the cycle component of HP-filtered log real GDP. The time series are at country-level. Bin scatter controls for year- and age-fixed effects.

cohort at the time of entry (age 0).[14] In this analysis, I consider cohorts' survival rates for up to age 5 after they enter the market.[15] To investigate how the cohorts' survival rates over the life cycle are correlated with the aggregate conditions at the time of entry $t - g$, I use the cycle component of the annual real GDP.[16] To find the cyclical component of the annual log real GDP I apply the HP filter with a smoothing parameter of *100*.

Figure 2 provides binned scatter plots of pooled life cycle survival rates of cohorts at the country-level against the business cycle indicators at the time of entry. The binned scatter plots control for 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' of establishments average survival rates over the life cycle. Panel (b) of Figure 2 shows that the negative relationship is robust if we consider cohorts of firms rather than establishments.

Next, I use the state-level variation in the life cycle dynamics of new businesses to further

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

[15]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.

[16]I annualize the quarterly real GDP data so that its consistent with BDS timing. Specifically, in the BDS dataset, 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. The source and the construction of the annual real GDP data are described in Appendix E.2. For more information see the website.

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)
State FE	✓	✓	✓	✓	✓	✓
Age FE	✓	✓	✓	✓	✓	✓
Year 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 pooled survival rates up to five years of operations of cohorts of new businesses. Panel A uses cohorts of establishments and Panel B uses cohorts of firms as a unit of analysis. Columns (1)-(3) use different indicators for characterizing the business cycle conditions at entry. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

investigate the relationship. I estimate the following regression:

$$S_{g,s,t} = \alpha + \beta Z_{t-g} + \eta_a + \theta_t + \gamma_s + \varepsilon_{g,s,t}, \tag{1}$$

where $S_{g,s,t}$ is a survival rate of a cohort at age g, in state s, at time t; Z_{t-g} represents the economic conditions at the time when the cohort first entered the market.[17] 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 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 12 reports the results of the regression equation (9) 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 the life cycle. Specifically, a 1-percentage point increase in real GDP above the trend decreases cohorts' average survival rates by 0.028 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 if the cyclical component of the log real GDP is below trend ($Y_{HP,I}$). Column (3) uses the NBER-based recession indicator for the US from the period following the peak through

[17] 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 2: The survival rates by age and aggregate economic conditions at the time of 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 pooled survival rates up to five years of operations of cohorts of new businesses. Panel A uses cohorts of establishments and Panel B uses cohorts of firms as a unit of analysis. Columns (1)-(3) use different indicators for characterizing the business cycle conditions at entry. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

the trough ($NBER$).[18] The indicator equals one 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 12 shows that the results hold if I use cohort of firms as a unit of analysis rather than establishments.

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

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

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 the change in the survival rates of a cohort at age g with the variation in the business cycle conditions at entry.

[18] The latter indicator specifies peak and the 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.

Panel A of Table 13 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 on the cohorts of establishments post-entry survival rates: cohorts of establishments that start operating during recessions are characterized with higher survival rates for up to age five. The results are robust to alternative business cycle indicators. The results also hold if we use firms as the unit of analysis rather than establishments.

To interpret the results, note that the aggregate economic conditions have two counteracting effects on new cohorts' survival rates. On the one hand, bad aggregate conditions directly decreases cohorts' survival rates by increasing firms' post-entry failure rates. On the other hand, bad aggregate conditions could increase cohorts' survival rates by selecting better firms at entry. The finding that cohorts' average survival rates are countercyclical supports the hypothesis that the initial aggregate conditions have a significant effect on the selection of firms at entry.

In Appendix A.1 I apply same analysis to 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.

2.2 The Evidence of Delays in Entry Decision

Next, I document that part of the business cycle variation in the number of start-ups comes from the entrants' 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.[19] Information provided in the EIN application is used to identify a subset of applications associated with the start of new busi-

[19]EIN is a unique number assigned to most of the business entities. An 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 new EIN.

nesses, referred to as business applications (BA).[20] The BA are matched to the set of firms in the BDS dataset identified as new employer businesses based on payroll information. The match process is straightforward because both of the datasets contain information about EINs.

The publicly available part of the BFS dataset provides country- 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 ($\boldsymbol{F}8Q$). This group of businesses covers more than 80% of the total number of entrants each year in the US.[21] In the analysis, I consider the time series of the number of applications that form businesses within the *first* four ($\boldsymbol{F}4Q$) and *second* four ($\boldsymbol{S}4Q$) 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*.[22] 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 entry. Then, the share of the applications that form businesses with one year delay should increase with deteriorating economic conditions at the time of the application.

Table 3 reports the summary statistics of the share of late start-ups at the state and country level. 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. Panel A of Table 3 summarizes the state-level variation in these variables. Out of the total applications that form businesses in the first eight quarters, 13% starts businesses with a one-year delay. The

[20] The EIN application includes 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 so on.

[21] For more details see Appendix A.2, Figure 20.

[22] Information about the raw number of EIN applications can not be used to identify delays in business formation. 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. In particular, 12% become employer businesses in the first four quarters, and an additional 2% become employer businesses after a year. 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 the significant share of the business applications ends up becoming non-employer businesses.

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
	$Dur\boldsymbol{F}4Q$	2,142	1.02	0.16	0.55	1.99
	$Dur\boldsymbol{S}4Q$	2,142	5.43	0.20	4.83	6.21
	$Dur\boldsymbol{F}8Q$	2,142	1.58	0.27	0.81	2.62
Panel B. Country-level	*Share of late start-ups*	42	0.14	0.02	0.11	0.18
	$Dur\boldsymbol{F}8Q$	42	1.66	0.17	1.37	1.96
	$Dur\boldsymbol{S}4Q$	42	5.46	0.26	4.95	5.75
	$Dur\boldsymbol{F}4Q$	42	1.06	0.09	0.88	1.21

share of late start-ups varies from 2% to 26% across time and states, with the overall standard deviation around 3 percentage point. 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 within the fifth and sixth quarters from the quarter of the application - implying that the variation in the share of late start-ups goes beyond the variation in the average duration of business formation. Panel B of Table 3 reports the same statistics for the aggregate data. Appendix A.2 provides a detailed description of the dataset.[23]

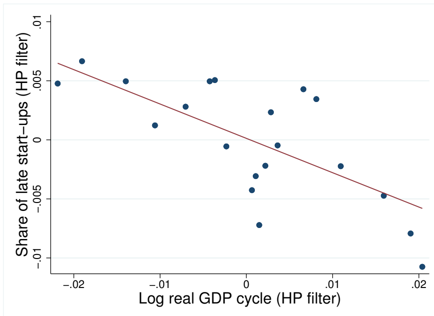
To assess economic conditions at the time of the application, I use the following business cycle indicators: (1) The cycle component of the quarterly log real GDP. To find the cyclical component of the yearly log real GDP I apply the HP filter with a smoothing parameter of *1600*. (2) The log annual real GDP growth between t and $t+1$. I measure the latter as a change in the rolling sum of consecutive four quarters starting from the quarter of the application.[24] The positive value of this variable indicates that the outlook for the aggregate economic conditions tomorrow is better than today. I construct both of the indicators at the state- and country-level;

Figure 3 plots binned scatter plots of the share of late start-ups against the business cycle indicators. Panels (a) and (b) illustrate correlations at the country level, while Panels (c) and (d) display the 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

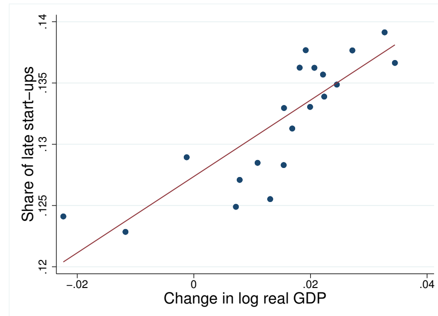
[23] In Appendix A, I also discuss in detail the information provided in the BFS dataset and its relevance for the mechanism developed in the paper.

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

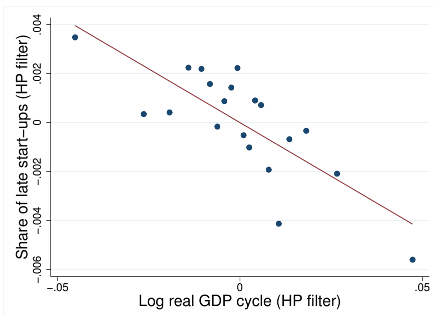
Figure 3: The share of late start-ups against the aggregate economic conditions at the date of the application



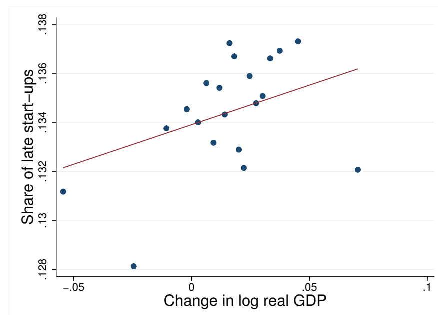
(a) Country-level



(b) Country-level



(c) State-level



(d) State-level

Note: Each panel plots 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. In the latter, I control the state-level fixed effects. Panels (b) and (d) also contain linear and quadratic time trends to only account for the business cycle variation in the share of late start-ups.

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

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	
$Z =$	$Y_{HP\ s,t}$ (1)	$\Delta Y_{s,t}$ (2)	$Y_{HP\ t}$ (3)	ΔY_t (4)	$Y_{HP\ t}$ (1)	ΔY_t (2)
β	-0.063***	0.068***	-0.161***	0.154***	-0.310***	0.165***
	(0.015)	(0.017)	(0.027)	(0.036)	(0.115)	(0.037)
$Dur\boldsymbol{F}4Q$	0.036***	0.042***	0.033***	0.027***	0.080***	0.071**
	(0.006)	(0.005)	(0.006)	(0.006)	(0.023)	(0.027)
$Dur\boldsymbol{S}4Q$	-0.062***	-0.020	-0.062***	-0.010	-0.218**	-0.086
	(0.008)	(0.016)	(0.007)	(0.016)	(0.103)	(0.125)
$\boldsymbol{F}8Q$	0.022***	-0.016***	0.025***	-0.001	0.020	-0.013
	(0.006)	(0.003)	(0.007)	(0.0007)	(0.030)	(0.009)
WBA	-0.031***	-0.011***	-0.025***	-0.014***	-0.010	-0.021***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.010)	(0.007)
State FE	✓	✓	✓	✓	✓	✓
Quarter FE	✓	✓	✓	✓	✓	✓
Observations	2,040	2,040	2 040	2,040	39	35
R-squared	0.148	0.699	0.139	0.720	0.704	0.757

Notes. 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.

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 (**F**8Q). I also include the total number of wage-based business applications (WBA) to control potential 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 or/and the variation in the total number of business applications.[25]

Table 4 reports the results of the regression equation (3). Panel A considers state-level

[25] 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.

variation in the share of late start-ups. Column (1) uses the state-level HP-filtered log real GDP as a business cycle indicator. I apply HP-filter to all other variables in this specification, too. 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 0.0063 percentage points. Column (2) considers the state-level change in the log real GDP as a business cycle indicator. I apply the linear trend to all other variables. 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 consider the same regression specification as before except uses business cycle conditions at the country-level rather than state-level. Panel B of Table 4 runs the same regression using country-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 some part of the variation in the new business formation as a response to the changes in the aggregate conditions at entry could be due to the entrants' option to delay entry.

3 The Model

The model builds on Moreira (2015), and 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 only source of business cycles. Time is discrete. Agents face an infinite horizon. 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 in the future 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

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

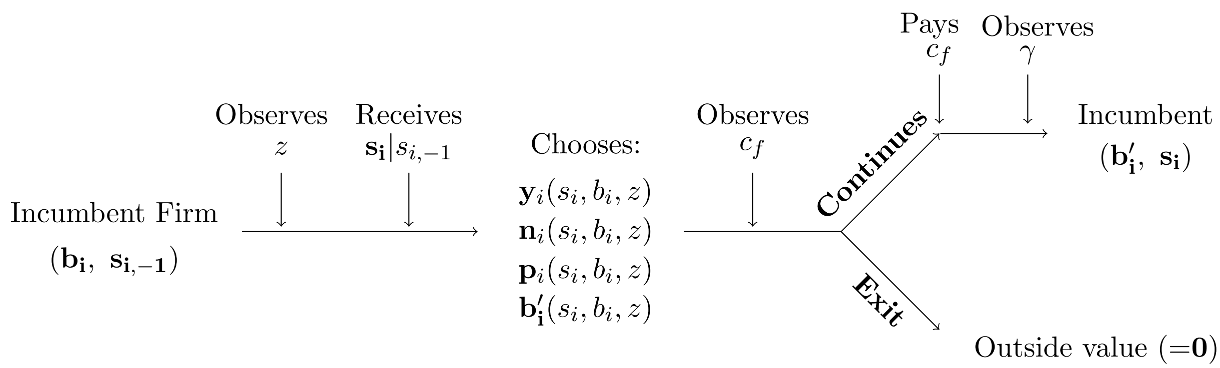
where p_i is the price set by firm i, and $\rho > 1$ is the price elasticity of demand. $\eta \in (0,1)$ measures the elasticity of demand with respect to customer capital b_i, which evolves according to:

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

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.[26] z represents a common aggregate demand

[26] Foster et al.'s (2016) findings motivate incorporating the persistent customer-capital-accumulation process in the model. Specifically, they find the differences between young and mature firms are due to individual demand dynamics rather than differences in productivity. Sedlacek and Sterk (2017), and Moreira (2015) explain the persistent procyclical variation in cohorts' employment using the demand-side factors. This framework enables me to quantify the role of the demand-side factors versus the option value of delay in explaining the post-entry cohorts' performance.

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.

Incumbent Firm's Timing 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.[27] Firms discount future profits at the time-invariant factor β.

The incumbent firm solves the following functional equation:

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

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

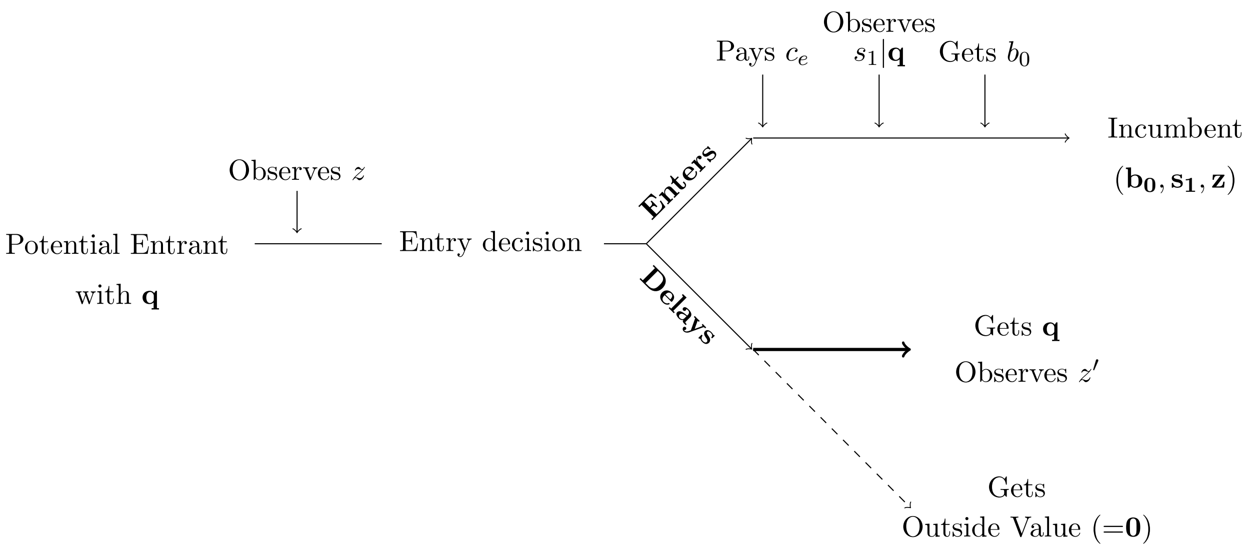
$$y = \alpha p^{-\rho} b^{\eta} z.$$

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

[27] Assume that if the incumbent firm decides to exit from the market, the probability that the firm receives an initial productivity signal and becomes a potential entrant again is zero.

Figure 5: Potential entrant's timing

‘



3.1.1 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 $\epsilon \sim N(0, 1)$.[28]

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 $\xi > 0$.[29] The potential entrant's timing is described below and is summarized in Figure 35.

[28]The ex-ante heterogeneity of potential entrants is crucial to study the role of the option value of delay. In particular, if one decides to delay entry, other entrants want to defer entry if potential entrants are ex-ante homogeneous. For the interior solution of the entry rate, the option value of delay has to equal to zero. For example, see paper Bilbiie, Ghironi, and Melitz (2012). The distribution of the potential entrants across the signal does not vary with the business cycles; thus, the feature does not contribute to the cyclical variation of entrants.

[29]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), Sterk (2019), 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 wait. 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. In fact, I find that allowing the accumulation of potential entrants amplifies the differences between the characteristics of entrants over the business cycle.

Potential Entrants' Timing 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. Entrant solves the following Bellman equation

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

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 Entry Timing and the Value of Waiting

The goal of the section is twofold. First, I study the optimal timing of market entry decisions made by heterogeneous potential entrants. Second, I investigate the implications of the option to wait on the selection of firms at entry over the business cycles. Toward the goal, consider the following modification of equation (3)

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

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, $V^w(z,q)$.

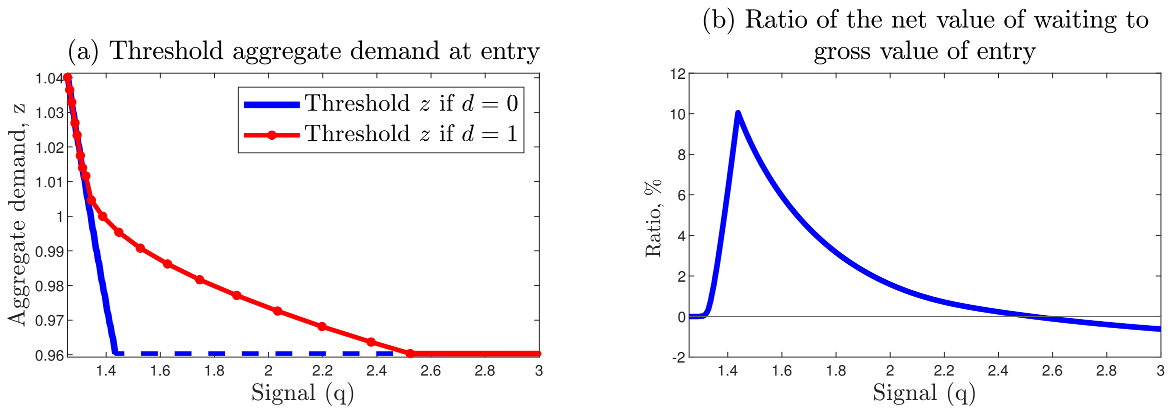
Result 4.1. *(i) $V^w(q,z)$ is non-negative for all q and z; (ii) For a given aggregate demand level z, $V^w(q,z)$ is a weakly increasing function of the signal q; and (iii) For a given signal q, $V^w(q,z)$ weakly increases with the aggregate demand level z.*

Result 4.2 indicates that the option to wait weakly increases the total cost of entry for all potential entrants and more so for firms with higher productivity signals. Solving the entry timing problem for a potential firm with signal q consists of finding a threshold aggregate demand level $\underline{z}^d(q)$ for which the firm enters the market if $z \geq \underline{z}^d(q)$, where superscript d indicates whether the entrant has the option to delay the entry decision. Result 4.2 formally characterizes the optimal entry rule and the threshold aggregate demand level across heterogeneous potential entrants using the numerical solutions.

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

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

Figure 6: Entry decision and the value of waiting



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 level for the baseline model, while the solid-blue line displays the same statistics for $d = 0$ case. The blue-dash line 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 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 profitable to wait for higher aggregate demand levels. Interestingly, the option to wait significantly increases the threshold aggregate demand level even for the potential firms whose net present value of entry is more than zero for all reasonable aggregate demand levels (the range of signals are illustrated by the blue-dashed line in Figure 6a).

To evaluate the quantitative significance of the option to wait on potential entrants' entry decisions, I evaluate the cost that a firm incurs by entering the market at suboptimal times. The *net value of waiting* equals the difference between the net present value of entering the market today and the net present value of making the decision tomorrow. That is,

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

Note that when the aggregate demand level at entry is $\underline{z}^{d=0}(q)$, the net present value of becoming an incumbent today equals zero for a potential entrant with signal q, and the net value of waiting equals to the option value of delay $V^w(q, \underline{z}^{d=0}(q))$. Figure 6(b) illustrates the net value of waiting for each signal level relative to the gross value of entry as a comparison base. The figure illustrates that the net value of waiting – the cost that entrant incurs by entering at a suboptimal time can be as high as 10 percent of the gross value of entry for entrants with medium productivity signals. Conversely, the option to wait has no value for entrants with low productivity signals and has a negative value for firms with high-productivity signals.

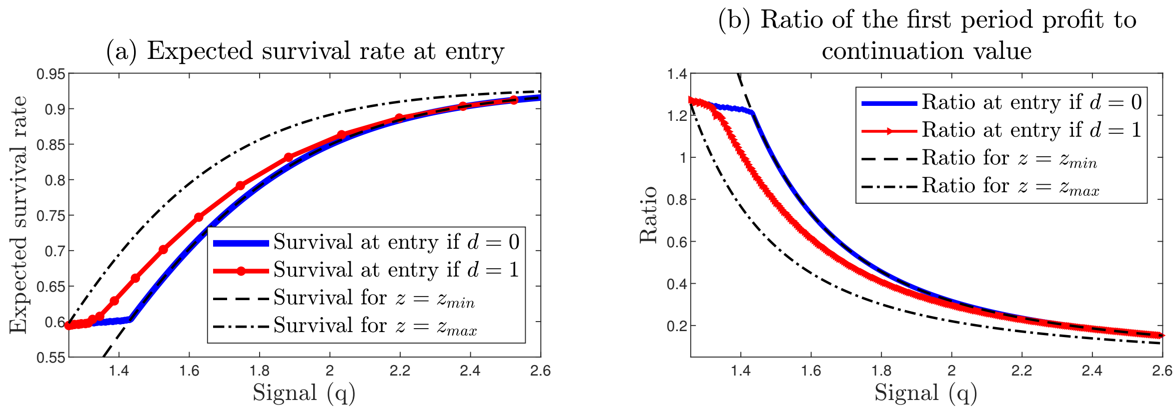
To investigate the rationale behind a firm's choice to delay entry consider Equation (5), which decomposes the gross value of entry into the expected first-period profit and the expected continuation value. The latter value depends on the probability that a potential entrant stays in the market after the first period, given by the following expression $(1-\gamma)G(c_f^*)$. Figure 7(a) illustrates the expected survival rate for the expansionary (black-dotted line) and recessionary (black-dashed line) periods. One can see that the expected survival rate is procyclical - the lower the aggregate demand level, the higher the expected risk of post-entry failure. During the recessions, firms' lose part of their long-run benefits due to the higher probability of exit. With the option to delay entry potential entrants sensitivity toward the post-entry failure risk significantly increases.

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

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

There is a cost to delay - the forgone period profits but the cost must be weighed against the benefits of increasing the continuation value of the entry. To illustrate the trade-off, consider Figure 7(a) that displays the differences between firms expected survival rates across the threshold aggregate demand levels. The blue-solid line displays the optimal expected survival of firms at entry without the option to wait, whereas the red-circle line displays

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



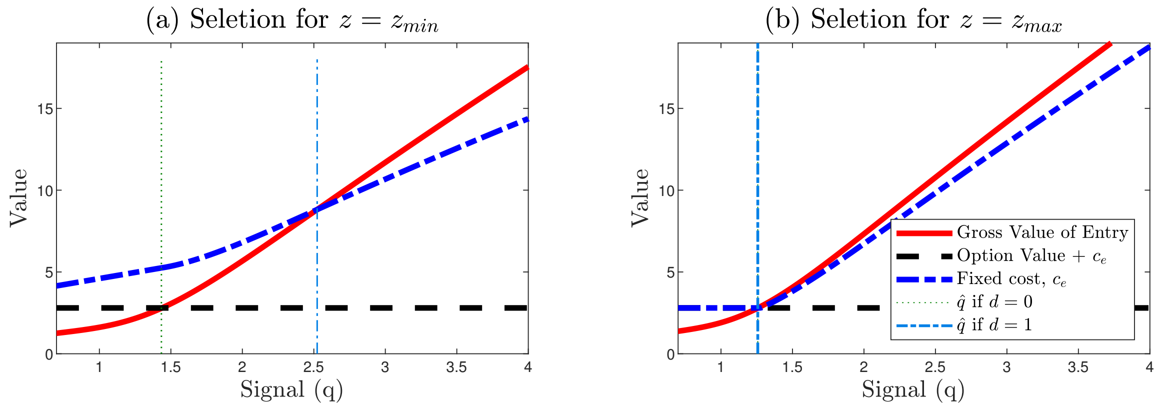
the expected survival rate if the entrants are able to time their entry decisions. The figure illustrates that by postponing entry, these group of firms are able to increase the expected survival rates. Figure 6(b) illustrates the ratio of the expected first period profit to the long run value. Entrants with low-productivity signals have, on average, a higher risk of post-entry failure, and the first period profits represent a significant share of their entry value. They enter the market when their first-period profit is positive that happens during the highest aggregate demand levels and during the time the waiting has no additional value. Note that the expected survival rate and, hence, the continuation value increase with the signal level. At the same time, the procyclical variation in the survival rates 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. Those who delay entry stay outside the market until the expected survival rate is high enough to compensate for lower demand levels in the first several years of operation.

4.1 Aggregate Selection of Entrants

Next, I investigate the implications of the option to wait on the selection of entrants across the aggregate conditions. 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. If $d = 0$, firms enter the market if the gross value of entry is greater than the fixed entry cost; these firms are the ones that hold signals $q \geq \hat{q}_{d=0}(z)$. The

Figure 8: Selection of entrants across different aggregate conditions at entry



rest stay outside the market. I refer to $q \geq \hat{q}_{d=0}(z)$ as a threshold signal for an aggregate demand level z when $d = 0$. A signal with similar characteristics exist in the case $d = 1$. In particular, $q \geq \hat{q}_{d=1}(z)$ characterizes a group of firms that enter the market because their expected post-entry profits are greater than the total opportunity cost of entry. Again, the rest stay outside the market. Comparing these two cases helps us isolate the selection through the option to delay entry. In particular, for an aggregate demand level z, the option generates an additional group of firms with $q \in [\hat{q}_{d=0}(z),\ \hat{q}_{d=1}(z)]$ that, despite the positive net expected benefits of entry, decide to stay outside the market. Figure 8(b) shows that during the "highest" aggregate demand periods, 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.

The following result formally summarize the numerical solution findings for the baseline and $d = 0$ cases:

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

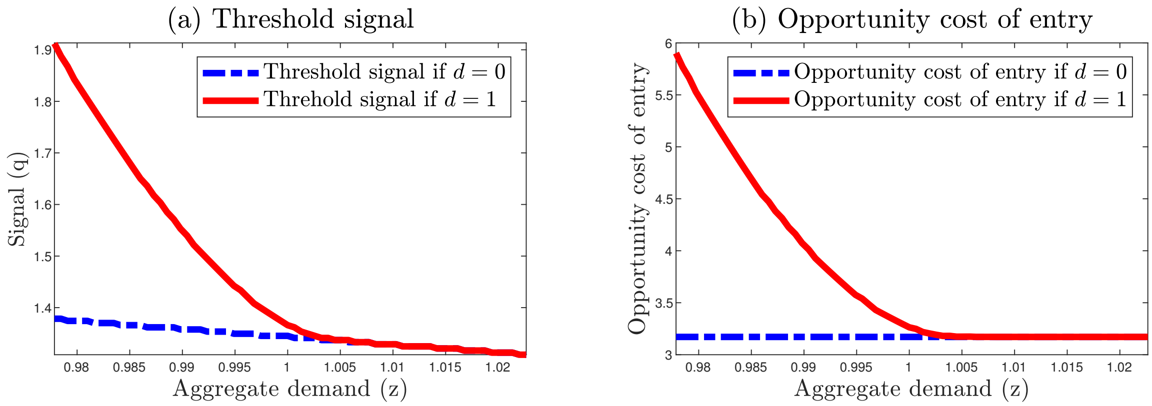
$$V^{gross}(z,\ \hat{q}_d(z)) - c_e = dV^w(z, \hat{q}_d(z));$$

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

Result 4.4. *The threshold signal $\hat{q}_d(z)$ is countercyclical.*

Figure 9(a) shows the threshold signal $\hat{q}_d(z)$ 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. The mechanism leads to an endogenous variation in the number and the productivity composition of entrants over the cycles. Specifically, during

Figure 9: Selection of entrants across different aggregate conditions at entry



recessions, an increased threshold signal leads to a fewer but higher-productivity entrants compared with expansionary cohorts. That said, reconciling the documented variation in the number and composition of entrants requires high elasticity of the threshold signal with respect to aggregate demand level.

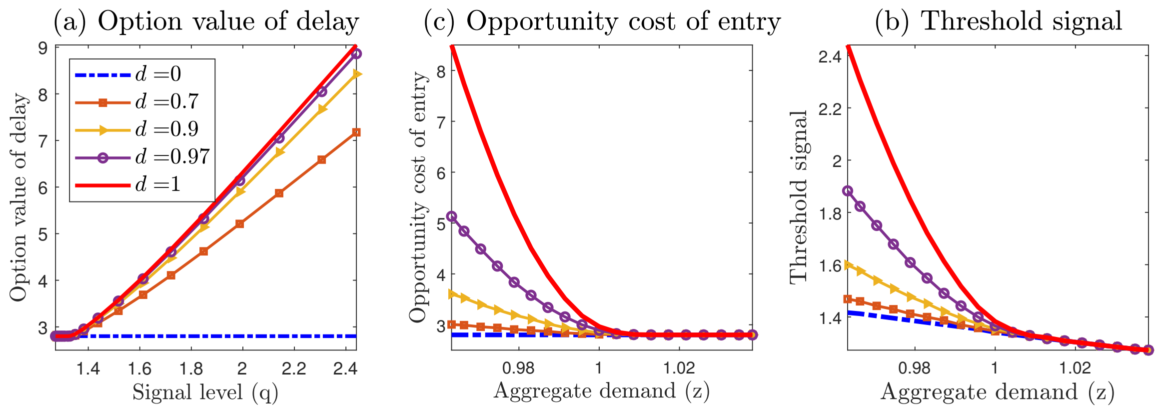
Note that without the option to delay entry, the threshold signal hardly varies with the aggregate demand level. In this case, the entry decision follows a traditional, neoclassical investment-decision rule: a firm starts a business if the net life-time benefits of entry are non-negative. The latter value is relatively insensitive to aggregate shocks of reasonable magnitudes. As a result, models that rely on conventional entry decisions could explain only a modest part of the observed variation in the entry margin.

Figure 9(a) illustrates that the option to delay entry significantly increases the elasticity of the threshold signal with respect to the aggregate demand level compared with the case $d = 0$. The latter is due to the medium-productivity firms with $q \in [\hat{q}_{d=0}(z),\ \hat{q}_{d=1}(z)]$ that choose to postpone entry despite the positive expected post-entry benefits. Note the lower the aggregate demand level, the wider the range of signals that leads to the delay decision.

To understand how the option to delay entry amplifies the effect of the aggregate conditions on the selection of entrants I compare the threshold cost of entry across these scenarios. I define the latter as follows: all potential entrants with the gross value of entry higher than the threshold cost enter the market, while the rest decide to stay outside the market. In a model with the option to delay entry, the threshold cost coincides with the threshold signal's $q^*_{d=1}(z)$ opportunity cost of entry.[30] Figure 9(b) illustrates that the latter value is

[30]Potential entrants with signal $q > q^*_{d=1}(z)$ enter the market and expect returns that are higher than the threshold signal's $q^*_{d=1}(z)$ total opportunity cost of entry. *Proof:* $V^{gross}(q, z)$ strictly increases with the

Figure 10: Selection of entrants across different aggregate conditions at entry for $d \in [0, 1]$



countercyclical: the cost of entry significantly increases above the fixed entry cost during the recessions. In fact, for reasonable parameter values, potential entrants postpone exercising the signal until the present value of entry is up to twice the fixed entry cost. Comparing the threshold cost of entry across cases elucidates the mechanism of how the option to delay entry increases the affect of the aggregate conditions on the selection of entrants.

4.2 Comparative Statics

Finally, for comparative statics I investigate how the value of waiting change if the potential entrants can only keep signals with some probability. That is I consider the Equation (4) where I allow $d \in [0, 1]$. In that case, d represents a probability that the potential entrants will carry signals tomorrow. Figure 10(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 10(b) and 10(c) show that the total opportunity cost of entry, as well as, the threshold signal level significantly increases with d.

5 Calibration and Model Performance

In this section, I calibrate the model to match the stylized facts about the average life-cycle dynamics of entrants. Then, I evaluate the model's performance in accounting for the observed dynamics of entrants over the cycles and quantify the role of the option to wait, by comparing it with an alternative scenario without the channel. Utilizing the good fit of the model, I evaluate the role of entrants' demographics in shaping the business cycle dynamics

signal. For an aggregate demand level z, firms with $q > q^*_{d=1}(z)$ enter the market. The following inequality holds: $V^{gross}(z, q) > V^{gross}(z, q^*_{d=1}(z)) = c_e + V^w(z, q^*_{d=1}(z))$.

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	0.073	Firm exit hazard by age
$\underline{q}$	Pareto location	0.700	Firm size at entry
ξ	Pareto exponent	1.349	Employment share at entry
c_e	Fixed entry cost	3.030	Entry rate – mean
ρ_z	Aggregate shock – persistence parameter	0.570	Entry rate – persistence
σ_z	Aggregate shock – SD parameter	0.002	Entry rate – SD

of the 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 of the establishments' idiosyncratic productivity process equals $\rho_s = 0.814$.[31] Foster et

[31]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

Table 6: Calibration targets and the model-implied counterparts

Statistics	**Data**	**Model**
Firm size	17.0	16.9
Firm size at entry	8.81	9.58
Firm size at age 5	13.9	14.1
Firm size at age 23	21.2	22.4
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	0.48
Firm survival rate up to age 23	0.15	0.12
Entry rate (%)	12.2	10.2

Note: The moments are calculated using the US-level cohorts of establishments from the BDS dataset covering the the period 1977-2015.

Table 7: Calibration targets for the aggregate demand shock process

Statistics	**Data**	**Model**
Autocorrelation of the cycle component of entry rate	0.25	0.25
Standard deviation of the cycle component of entry rate	0.06	0.06

Note: The time series about the entry rate comes from the BDS and covers the period 1977-2015. The cyclical component of the log entry rate is calculated using the HP filter with smoothing parameter 100.

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

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).

entry, the average size at entry relative to the size of all active establishments.[32] 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.

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.

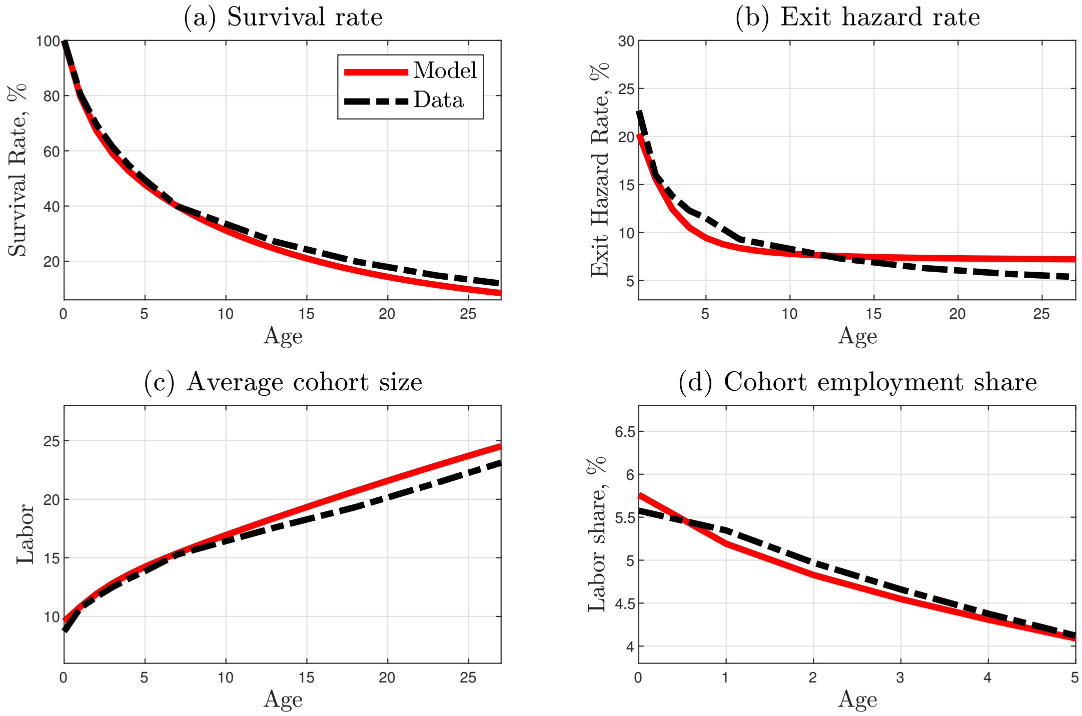
Lastly, the persistence (ρ_z) and standard deviation (σ_z) of the aggregate demand shock process are calibrated to match the autoregressive properties of the cycle component of the entry rate in the model and the data. To calculate the cycle component of the entry rate, I apply the HP filter with a smoothing parameter of *100*. To calculate the same moment in the model, I simulate the economy over many periods and apply the same detrending method to the model-simulated 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 $\rho_z = 0.57$, and $\sigma_z = 0.002$.

5.2 Cohorts' Average Life Cycle Characteristics

Table 6 lists the data moments (column 2) and their model-implied counterparts (column 3). 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

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

Figure 11: Cohorts' average characteristics: Data, Model



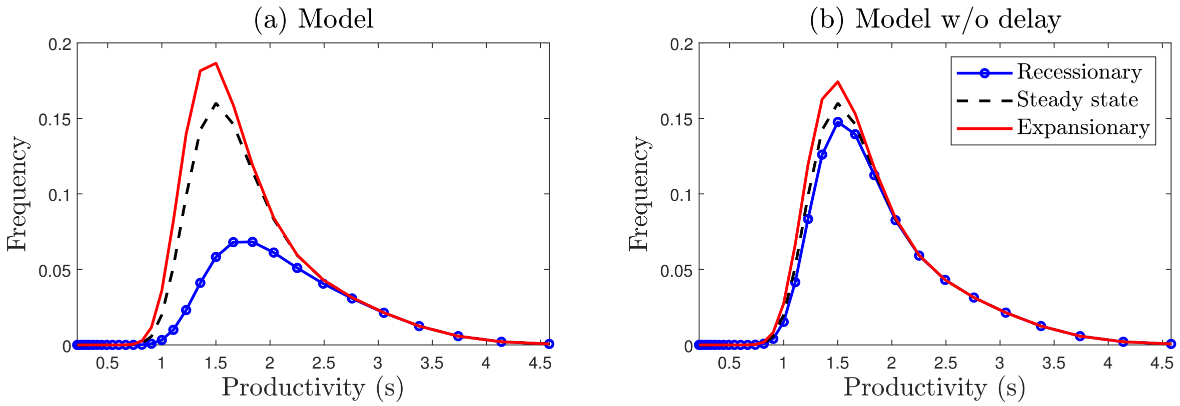
only around 5.6 percent to total employment in the model and the data. The model does a good job in reproducing the well-known 'up or out' dynamics of firms. About 50% of the entrants fail within the first five years, and by age 23, only around 10% out of original start-ups survive. At the same time, cohorts of firms grow from 9.6 workers at entry to 22 workers by age 23.

Figure 11 goes beyond the targeted 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 exit by age 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 by age. Overall, Figure 11 shows that the model quite closely reproduces average US cohorts of establishments life cycle dynamics.

5.3 Entry Conditions and Persistent Cohort Dynamics

In the section, I show that the calibrated model generates the documented persistent and significant differences in the life cycle characteristics of cohorts that start operating at different stages of business cycles. Alongside, I quantify the role of the option-to-delay channel

Figure 12: Aggregate Entry Conditions and Entrant' Productivity Distribution



in accounting for the observed dynamics of entrants.

To describe the business cycle conditions at entry, I use the aggregate demand shock process. I refer to a period as a recession (expansion) when the aggregate demand level is below (above) the stochastic steady state level $z < 1$ ($z > 1$). I define cohorts as recessionary (expansionary) if they start operating during the recessions (expansions).[33]

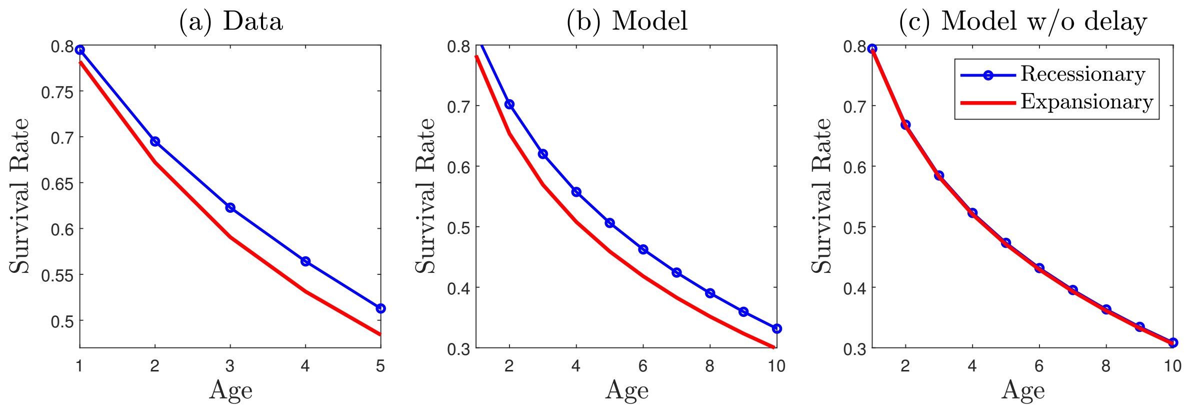
To quantify the role of the option-to-delay channel, I consider a version of the model without the option to delay entry ($d = 0$) that is re-calibrated to produce the same set of facts in the stochastic steady state as the baseline model. I refer to the case as a '*model w/o delays*'. It turns out that the *model w/o delays* is identically parameterized except for the fixed entry cost. I set the latter equal to the steady state total opportunity cost of entry in the baseline model.[34] With the choice of the fixed entry cost the alternative scenario exhibits the same dynamics in the stochastic steady state as the baseline model. The differences in the entry-cost structure outside the steady state is due to the option to delay entry and implied endogenous countercyclical entry cost.[35] Therefore, by comparing the business cycle dynamics in the baseline model against the *model w/o delays* allows me to quantify the role of the option to delay entry in accounting for the observed significant and persistent differences in the cohort post-entry characteristics.

[33]The results are robust to the definition of the business cycles within the model. In particular, results are similar 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). The results are robust because the model generates more or less symmetric business cycles.

[34]$c_{d=0} = c_{d=1} + V^w(q^*_{d=1}(z_{ss}), z_{ss})$. Equalizing the opportunity cost of entry ensures that the threshold signal coincides across these two scenarios, which in turn imply the same number and composition of entrants in these scenarios. The Column (b) of Table 18 summarizes the parameter values used in the *model w/o delays* scenario.

[35]For illustration see Figure 32.

Figure 13: Aggregate Conditions at Entry and Cohorts' Life Cycle Survival Rates



Productivity Consistent with the empirical findings, in the model the aggregate economic conditions at entry have a significant and persistent effect on the productivity composition of entrants.[36] Figure 12(a) depicts entrants' distribution over the initial productivity across different aggregate demand levels. The productivity distribution of entrants is positively skewed. The skewness decreases with the aggregate demand level, producing countercyclical average productivity. Recessionary cohorts average productivity is around 3% higher than their expansionary counterparts. The difference persists in later years due to the persistent idiosyncratic productivity process. Figure 12(b) shows the same statistics for the *model w/o delays*. Shutting down the option-to-delay channel reduces the differences in the average productivity of the recessionary and expansionary cohorts to *0.4*%.

Survival Rates Next, I investigate cohorts' life cycle survival rates over the business cycles. Figure 13(a) shows that the average survival rates of the US cohorts of establishments born during recessionary periods are persistently higher compared to their expansionary counterparts. For robustness and extensive analysis of this fact refer to Section 2. Figure 13(b) illustrates that the model accounts for the differences in the life cycle survival rates of cohorts born at different stages of the business cycles. Figure 13(b) shows that the model w/o delays leads to the acyclical average survival rates. Thus, the model implied countercyclical survival rates are a direct implication of the optimal timing of entry decisions. As discussed in Section 4, the option to delay entry allows firms to endogenous post-entry survival rates in their entry decision. That is, firms decide to wait until the expected survival rate is high enough to compensate for lower demand levels in the first several years of operation. As a result, consistent with the empirical facts due to this mechanism the cohorts

[36] Lee and Mukoyama (2015), Moreira (2015), and Ates and Saffie (2021) find that cohorts' of firms born during crisis are, on average, more productive over the life cycle than their expansionary counterparts.

Table 8: Cohort-level Employment in the Baseline and Counterfactual Scenarios

		Recessionary Cohorts			Expansionary Cohorts		
		Age 0 % dev.	Age 5 % dev.	Age 15 % dev.	Age 0 % dev.	Age 5 % dev.	Age 15 % dev.
(a)	Baseline	-5.7	-4.7	-4.8	5.0	4.0	4.1
(b)	The $\tau = 0$ case	-1.2	-1.0	-1.0	1.0	0.9	1.0
(c)	Baseline, adjust lowest s	-3.4	-1.4	-1.5	2.6	1.2	1.3
(d)	Baseline, adjust highest s	-12.5	-14.1	-13.3	10.0	11.2	10.6
(e)	Baseline, only selection	-5.3	-4.4	-4.5	5.4	4.3	4.4

Note: The numbers in the table describe percentage deviations (% dev.) of the recessionary (expansionary) cohorts' employment from the average cohort employment. Recessionary (Expansionary) cohorts refer to the group of firms that started operation when $z < 1$ ($z > 1$).

of firms that start operating during recessions has, on average, higher survival rates than their expansionary counterparts.

Employment Row (a) of Table 8 summarizes the dynamics of cohort-level employment at entry and over time for the baseline model. According to the results, the recessionary (expansionary) cohorts employ 5.7% less (5.0% more) workers than the average cohort and the differences do not disappear even after 15 years of operation. Row (b) of Table 8 summarizes the dynamics of cohort-level employment for the $\tau = 0$ case and shows that shutting down the option to delay entry reduces the difference to 1%, thus implying that the major share (80%) of the variation in cohort-level employment comes from the entrants that delay entry.[37]

I find that the persistent differences in cohort-level employment are due to variations in the composition rather than the number of firms at entry. Rows (c) and (d) of Table 9 summarize the dynamics in two counterfactual scenarios that feature the same variation in the number of entrants as the baseline model, whereas I let the composition of entrants vary systematically across these scenarios. Specifically, "Baseline, adjust lowest s", and "Baseline, adjust highest s" refer to the scenarios in which the variation in the number

[37] The model generates cohorts' with a countercyclical average size. The result is in line with Lee and Mukoyama (2015), who show that the average size of US manufacturing plants is countercyclical. However, the result is at odds with Sedlacek and Sterk's (2016) finding. Using the BDS dataset, they show entrants' average size is procyclical. I expect that extending the model to account for the procyclical average size at entry will increase the difference in cohort-level employment over the cycles. One potential extension to generate the procyclical average size of entrants is to assume the first-period level of customer capital is procyclical.

of entrants are generated by adjusting, respectively, the lowest- and highest-productivity firms from the steady state distribution of entrants.[38] Comparing the dynamics of these two scenarios shows that the variation in the number of entrants has a persistent effect on cohort-level employment if it comes from the high-productivity entrants.[39] Note that the dynamics of the baseline economy are in between these two counterfactual scenarios. The medium-productivity firms that delay entry increase the pro-cyclical variation in the high-productivity entrants and lead to higher persistence in the dynamics of cohort-level employment. The mechanism corresponds to Decker et al.'s (2014) empirical findings that the entrant cohorts' contribution to the aggregate employment comes from the small share of the high-growth firms. Pugsley, Sedlaćek, and Sterk (2016) also find that the major share of the entrant cohorts' post-entry performance is due to ex-ante differences in the types of entrants.

Finally, consider row (e) in Table 9. "Baseline, only selection" refers to the baseline scenario in which the aggregate demand shocks affect only the selection of entrants and have no effect on the firms' post-entry demand structure. Contrasting the baseline model with the counterfactual scenario shows that the persistent customer-capital-accumulation process plays a minor role (less than 7%) in generating persistence in the dynamics of cohort-level employment.

6 Aggregate Fluctuations

In this section, I use a model that closely mimics the life-cycle dynamics of the US establishments on average and over the business cycles to quantify the role of the entry margin in shaping aggregate fluctuations. To compute the business cycle moments from the data, I use the time series of the natural logarithm of aggregate employment, real GDP, and the total number of establishments that covers the period 1977-2015.[40] I apply the HP-filter with a smoothing parameter of 100 to find the cycle component of these variables. I use the same methodology to compute the moments from the model-simulated time-series.[41] The

[38]For more details refer to Appendix E.3.3.

[39]The mechanism corresponds to the "missing generation" channel initially discussed in Gourio, Messer and Siemer (2015).

[40]The time series of the aggregate employment and the real GDP are constructed to be consistent with the timing of the BDS dataset. Detailed information about the source and the construction of the aggregate variables are provided in Appendix E.2.

[41]In particular, I run the baseline economy over a large number of periods. I find the cyclical component of the natural logarithm of the simulated aggregate employment, output, and the total number of firms using the HP filter with a smoothing parameter of 100. I use the latter time series to compute the standard

Table 9: Business Cycle Moments: Data, the Baseline Model, and the Counterfactuals.

		Data (a)	Baseline (b)	Baseline, only selection (c)	The case $\tau = 0$ (d)
No. of firms	ρ	0.640	0.619	0.607	0.661
	σ	0.012	0.010	0.010	0.002
Employment	ρ	0.610	0.574	0.622	0.432
	σ	0.015	0.012	0.010	0.004
Entry Rate	ρ	0.250	**0.253**	0.252	0.222
	σ	0.062	**0.065**	0.065	0.010
No. of Entrants	ρ	0.311	0.278	0.278	0.245
	σ	0.066	0.073	0.073	0.011

Notes. All series are computed in log deviation from the HP trend. The numbers that are in bold refer to the targeted model statistics. All other values indicate untargeted model statistics and their empirical counterparts.

statistics from the data and the model are described in columns (a) and (b) of Table 9, respectively.

Table 9 shows that the variance and the autocovariance of the simulated total number of firms are very close to the data counterpart. The variation in the exogenous aggregate demand shock affects firms' life-cycle demographics in the following two ways: First, the aggregate demand condition affects the composition/number of entrants at entry. Second, aggregate demand affects incumbent firms' decisions about production and continuation. Aggregation of these two effects by adding up cohorts at different stages of their life cycle creates dynamics of the total number of firms that are very close to the data counterpart. The result can also additionally be interpreted as an external validation of the exogenous aggregate demand shock process.

Table 9 shows that the model that is built to account for the life-cycle demographics of firms (selection at entry, growth, survival) accounts for more than three fourths of the business cycle fluctuations in aggregate variables. In particular, the autocorrelation of the aggregate employment in the model is 0.57, whereas in the data, it equals 0.61. The standard deviation in the model and the data is 0.012 and 0.015, respectively.

Further investigation of the results shows that the variation in the number and the compo-

deviation and the autocorrelation of these variables.

Table 10: Impulse-Response Analyses

		Panel A: One-time shock			Panel B: Persistent Shock		
		Full Model	Fixed entry		Full Model	Fixed entry	
				z_{high}			z_{high}
		(a)	(b)	(c)	(d)	(e)	(f)
Depth (%)	Employment	-1.83	-0.72	-1.90	-2.0	-0.72	-2.1
	No. of Firms	-2.93	-0.07	-0.18	-3.04	-0.14	-0.43
50% Recovery	Employment	3	2	2	16	7	6
	No. of Firms	3	15	14	9	17	18
75% Recovery	Employment	15	2	2	28	14	15
	No. of Firms	8	23	23	17	26	27

Note: Baseline refers to a model with baseline specification. Fixed entry refers to a case in which the shock affects cohorts' post-entry performance, whereas the entry rate is fixed at the steady state level. z_{high} refers to a case in which the magnitude of the shock is chosen to produce a drop in employment as in Baseline scenario. Depth refers to the highest deviation of the time series from trend. 50% Recovery (75% Recovery) describes the number of periods (years) starting from period 1, after which economy recovers 50% (75%) from the 'depth'.

sition of firms at entry is responsible for shaping the dynamics of the aggregate variables. In particular, I consider a counterfactual scenario in which the variation in the aggregate demand affects selection but does not have an effect on firms' post-entry decisions.[42] The dynamics of the economy are summarized in column (c) of Table 9. One can see the aggregate dynamics in the counterfactual and the baseline scenarios are quite similar, which means the observed significant and persistent differences in cohorts' characteristics over the cycles build up significant persistence and variance in aggregate variables. At the same time, the result also implies that the post-entry shocks that affect firms' post-entry decisions provide a relatively minor contribution to aggregate fluctuations. The latter result corresponds to the recent empirical findings by Sedlacek and Sterk (2017), who show the selection of firms at the entry stage, rather than the post-entry choices made by the firms, drive the cohorts' contribution to aggregate fluctuations.

[42]In particular, I construct a counterfactual economy in which the aggregate demand shock has the same impact on the selection (composition/number) of entrants as in the baseline model. However, I set aggregate demand shocks equal to zero for all the firms that operate in the market.

6.1 Impulse Response Analysis

Considering the small share of entering firms in aggregate employment the result from the previous section might seem surprising. To illustrate how the variation in cohorts' characteristics can build up persistence and variance in the aggregate dynamics, I study the response of the baseline economy to a one-time negative aggregate demand shock, summarized in Panel A of Table 10. The magnitude of the shock is chosen to yield a 25% decline in the number of entrant establishments.[43] One can see that after the shock, the baseline economy takes three years to recover half-life and another *12* years to recover an additional 25% of the decline. By contrast, an economy in which the shock does not affect the entry margin takes only *2* years to recover the full, three fourths of the decline. The result is robust to the magnitude of the shock. Thus, changes in the number and composition of firms at entry that leads to a persistent decline in entrant cohorts' employment plays a significant role in the propagation of aggregate shocks. Panel B of Table 10 shows that if the change in the entry margin is persistent, the effect accumulates and has a substantial impact on the depth and the long-run recovery of the economic aggregates.[44]

6.2 The Great Recession

To additionally support the findings and validate the model, I study the Great Recession, which is notorious with the historical drop in the number of entrants and the unprecedented slow recovery of the aggregate employment that followed.[45] I use the episode to illustrate that the persistent drop in entrant cohorts' employment over the period 2008-2016 had a substantial effect on the slow recovery of the aggregate employment. Then, I use the model to investigate how much of the effect is due to the variation of entrants at the entry margin. I find that the persistently low aggregate demand shock series that leads to the persistent changes in the number and the composition of firms at entry quite closely account for the contribution of cohorts born over the period 2008-2016.

[43] The number corresponds to the decline in the number of entrants observed during the Great Recession.

[44] The mechanism is consistent with empirical findings by Gourio, Messer and Siemer (2016). Using an annual panel of US states over the period 1982-2014, they show that changes in the number of entrant firms have a persistent effect on the dynamics of the aggregate variables.

[45] Figure 38(a) plots the cyclical variation in the number of entrant establishments and the aggregate employment in the US over the period $1977-2016$.

6.2.1 The Data

Initially, I use an accounting exercise to directly quantify how much of the changes in the employment of cohorts that started operating over the period $2008-2016$ contributed to the slow recovery of aggregate employment.[46]

The aggregate employment at time t can be expressed as a sum of the total employment of the cohorts of establishments at different ages:

$$N_t = n_{0,t} + n_{1,t-1} + n_{2,t-2} + n_{3,t-3} + n_{4,t-4} + n_{5,t-5} + Res_t, \tag{6}$$

where N_t denotes aggregate employment and $n_{g,t-g}$ refers to total employment of a cohort of age g who started operating at time t, $g = 0, 1, 2, 3, 4, 5$. Due to the data limitations, I only consider cohorts up to age five.[47] Res_t combines part of the aggregate employment that belongs to establishments with ages 6+ and the segment of employment that is not part of the BDS dataset.

I consider the beginning of the recession to be year 2008.[48] Consider $\hat{N}_t$ to be the level of aggregate employment at time $t \geq 2008$ had the Great Recession not happened. $\hat{N}_t$ can be expressed as follows:

$$\hat{N}_t = \hat{n}_{0,t} + \hat{n}_{1,t-1} + \hat{n}_{2,t-2} + \hat{n}_{3,t-3} + \hat{n}_{4,t-4} + \hat{n}_{5,t-5} + \hat{Res}_t, \tag{7}$$

where $\hat{n}_{g,t-g}$ refers to the employment of a cohort of age g that entered the market at time t, had the Great Recession not happened. I define $\hat{Res}_t$ similarly. I use equation (6) and equation (7) to decompose changes in the aggregate employment as a sum of the changes in the cohort-level employment by age:

$$\Delta\hat{N}_t = \Delta\hat{n}_{0,t} + \Delta\hat{n}_{1,t-1} + \Delta\hat{n}_{2,t-2} + ... + \Delta\hat{Res}_t, \tag{8}$$

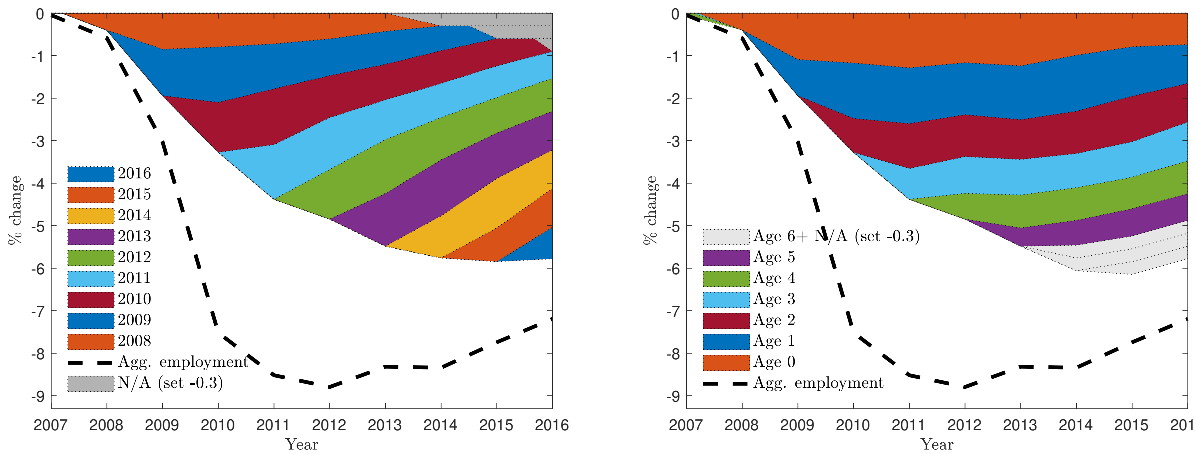
where $\Delta\hat{N}_t = \dfrac{N_t - \hat{N}_t}{\hat{N}_t}$ and $\Delta\hat{n}_{g,t-g} = \dfrac{n_{g,t-g} - \hat{n}_{g,t-g}}{\hat{N}_t}$ for $g = 0, 1, 2, 3, 4, 5$. $\Delta\hat{n}_{g,t-g}$ shows

[46] Gourio, Messer, and Siemer(2016) and Sedlaćek (2019) use data over the period $2008-2012$ and study how the persistent drop in the number of entrants contributes to the aggregate dynamics. In my exercise, I concentrate on changes in cohort-level employment, rather than the number of entrants.

[47] The publicly available part of the BDS dataset only allows me to separately track cohorts from age zero up to age five.

[48] The National Bureau of Economic Research (NBER) dates the beginning of the Great Recession as December 2007. In the BDS, the year 2007 characterizes establishment-level activity from March 2006 to March 2007. To be consistent with the NBER, I choose year 2008 as the beginning of the Great Recession.

Figure 14: New Cohorts' Contribution to the Slow Recovery of Aggregate Employment after the Great Recession



(a) By cohort

(b) By age

how much of the changes in the cohort employment of age g contributes to the changes in the aggregate employment at time t.[49]

Using the equation, I isolate the dynamics of the aggregate employment accounted for by cohorts that entered the market starting from $t \geq 2008$. Toward the end, consider the following counterfactual: for each year $t \geq 2008$, I only consider the deviations of the aggregate employment, $\Delta \hat{N}_{t,\ counter}$, that is accounted by the cohorts that entered the market from $t \geq 2008$. At year 2007, $\Delta \hat{N}_{2007,\ counter} = 0$. Starting from the year 2008,

$$\Delta N_{2008,\ counter} = \Delta \hat{n}_{0,2008},$$

$$\Delta N_{2009,\ counter} = \Delta \hat{n}_{0,2009} + \Delta \hat{n}_{1,2008},$$

$$\cdot \quad \cdot \quad \cdot$$

$$\Delta N_{2016,\ counter} = \Delta \hat{n}_{0,2016} + \Delta \hat{n}_{1,2015} + \Delta \hat{n}_{2,2014} + ... + \Delta \hat{n}_{6,2013} + \Delta \hat{n}_{7,2012} + \Delta \hat{n}_{8,2011}.$$

I apply a linear trend over the period 1979-2007 to predict the evolution of aggregate employment from the year 2008 as if the Great Recession had not happened.[50] I set $\hat{n}_{g,t-g}$ equal

[49] One can also think about it as a percentage deviation of the actual cohort-level employment from the predicted cohort-level employment, weighted by the share of the cohort employment in aggregate employment:

$$\frac{N_t - \hat{N}_t}{\hat{N}_t} = \left(\frac{n_{0,t} - \hat{n}_{0,t}}{\hat{n}_{0,t}}\right)\frac{\hat{n}_{0,t}}{\hat{N}_t} + \left(\frac{n_{1,t-1} - \hat{n}_{1,t-1}}{\hat{n}_{1,t-1}}\right)\frac{\hat{n}_{1,t-1}}{\hat{N}_t} + .. + \Delta \hat{Res}_t.$$

[50] In Appendix G.2, Figure 40 displays the evolution, pre-crisis trend and the prediction for the aggregate employment.

to the average employment of cohorts of age g over the period $2003-2007$. The latter allows me to study how the aggregate employment would have evolved during the Great Recession had the new cohorts of establishments behaved as the representative pre-crisis cohorts of establishments.[51]

Figure 14(a) illustrates the result of this exercise. The dashed black line represents the total deviation of the aggregate employment from the pre-crisis trend. The shaded areas represent the contribution of each cohort born over the period $2008-2016$ to the drop in aggregate employment. Several observations stand out. The cohorts that entered the market after the year 2008 employ persistently fewer workers, compared with their pre-crisis counterparts. These cohorts' dynamics contribute around 45% of a total 8.9% drop in aggregate employment in the year 2012. By the year 2016, the aggregate employment is 7% below the trend, and now 85% of the drop is due to cohorts that started operating over the period $2008-2016$. Thus, whereas the incumbent firms drive the depth of the recession, the dynamics of the new cohorts build up significant persistence in the dynamics of aggregate employment.[52] Figure 14(b) shows the same exercise by establishment age rather than the cohort year. The figure once again illustrates how much the persistent drop in cohort-level employment across different age groups contributed to the drop in aggregate employment.

6.2.2 The Model

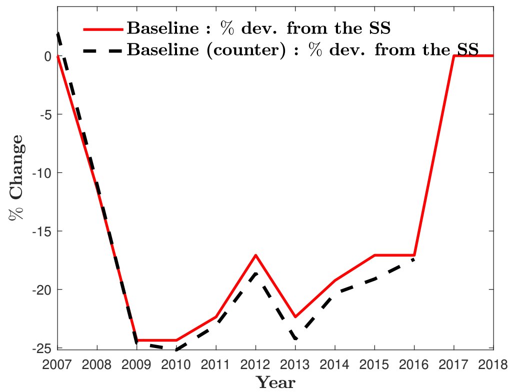
Next, I investigate how much a model that accounts for the US establishments' life-cycle demographics could explain the documented contribution of 2008-2016 cohorts. I also use the model to quantify the role of variation in the number and the composition of entrants in this contribution. Toward the goal, I construct an aggregate demand shock series that matches the changes in the simulated number of entrant establishments to the data counterpart over the period 2008-2016. Figure 46(b) illustrates the evolution of the number of entrant establishments in the model and in the data. Figure 46(a) displays the series of the aggregate demand shocks that generate the match. As in the empirical part, I used a linear trend over the period $1979-2007$ to predict the evolution of the number of entrant establishments starting from the year 2008, as if the Great Recession had not happened.[53]

[51]The cohort-level employment by age over the period $1983-2007$ shows that the times series exhibits an increasing trend; see Figure 41(a) in the Appendix G.2. The share of cohorts' employment in aggregate employment have a decreasing trend; see Figure 41(b). Thus, constructing a representative cohort based on pre-crisis average cohort-level employment captures a lower bound of the recent cohorts' contribution.

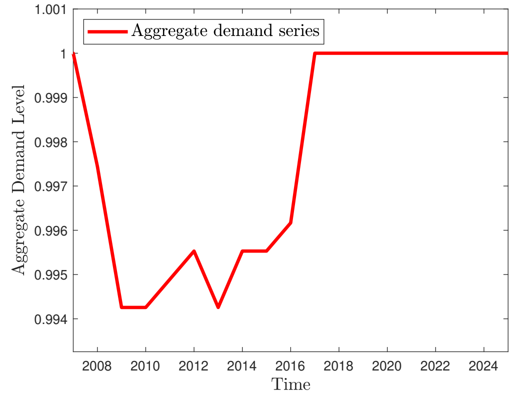
[52]In Appendix G.2, I show that the results are robust if I consider ten-year pre-crisis average of cohort-level employment.

[53]Figure 45 displays the evolutions, pre-crisis trends and predictions for these time series.

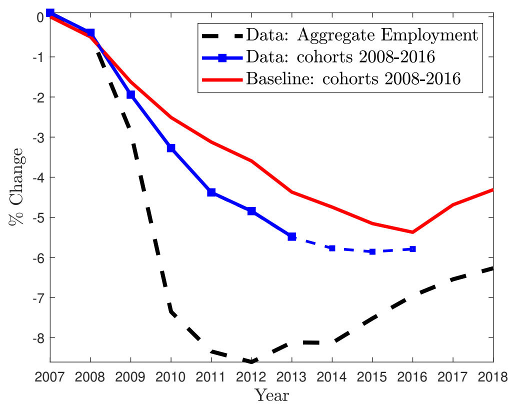
Figure 15: The Great Recession Exercise



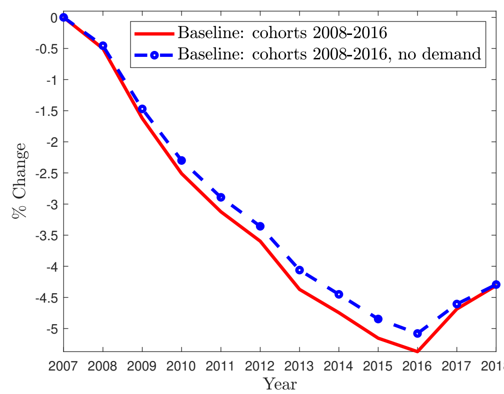
(a) Change in no. of entrants



(b) Aggregate demand process



(c) Cohorts 2008-2016



(d) Cohorts 2008-2016, no demand

The model predicts that changes in the number and the composition of entrants over the period 2008-2016 account for around 39% of the depth of the aggregate employment reached in 2012. By 2016, the persistent drop in the new cohorts' employment level accumulate, and it explains around 75% of the drop in aggregate employment. Figure 46(c) contrasts the changes in aggregate employment accounted by the cohorts born over $2008-2016$ in the model and data. The exercise shows that the combination of the aggregate demand shocks and the variation in the entry margin accounts for the major share of the documented contribution of 2008-2016 cohorts.[54] Next, to isolate the contribution of the changes in the number and composition of entrants at the entry margin, I consider a counterfactual

[54] Other economic forces, not considered in the paper, could explain the drop in 2008-2016 cohrots' employment. For example, the credit crunch that occurred during the Great Recession, significantly increased the cost of financing. The existing literature also points out that a potential structural change in the entrants during the Great Recession might have played an important role in the protracted recovery in aggregate variables. Figure 39 shows that all sectors experienced a significant and persistent drop in the number of entrants compared to the pre-crisis level (Gourio, Messer, and Siemer(2016)).

scenario in which the aggregate demand shocks only affect the selection and not the post-entry dynamics of firms. Figure 46(d) shows that post-entry demand shocks play a minor role and most of the observed contribution comes from the variation at the entry margin.

7 Other Applications

In this section, I show that existing business cycle firm dynamics models that employ a traditional neoclassical entry decision rule cannot account for the observed dynamics of entrants without generating excessive variation in the aggregate variables. The latter leads to counterfactual predictions about the role of entry. Firm dynamics models use various approaches to overcome the puzzle. For example, Lee and Mukoyama (2018) introduce entry cost that varies over the cycles in a particular way. Sedlaćek and Sterk (2019) introduce entry function, which allows choosing the elasticity of the number of entrants with respect to aggregate shocks. Others rely on exogenous entry specific shock processes (e.g., Clementi and Palazzo (2016), Sedlaćek and Sterk (2017)). In the second part of the section, I show that not accounting for the option to delay entry may lead to imprecise predictions about the response of potential entrants to different shocks.

7.1 The Standard Model

I study a model without the option to delay entry (I refer to it as the *Standard model*) that produces the same set of facts as the baseline model described in Section 5.1. Because firms' values are relatively insensitive to the aggregate state, the Standard model requires a variance of the aggregate demand shock σ_z, almost seven times higher than a model with the option to delay entry. Appendix E.3 provides a detailed description of the Standard model's calibration procedure.[55]

First, I show that the Standard model that accounts for the observed business cycle dynamics of entrants lead to excessive variation in aggregate variables and counterfactual predictions about the role of entry. Column (c) of Table 11 summarizes the business cycle properties of the economy. One can see that the model generates a variance of the aggregate employment that is *1.7* times higher than the data counterpart. Column (d) of Table 11 shows that the post-entry shock process explains a major share of the cohort performance, and hence the dynamics of the aggregate variables. This result is also at odds with the recent empirical

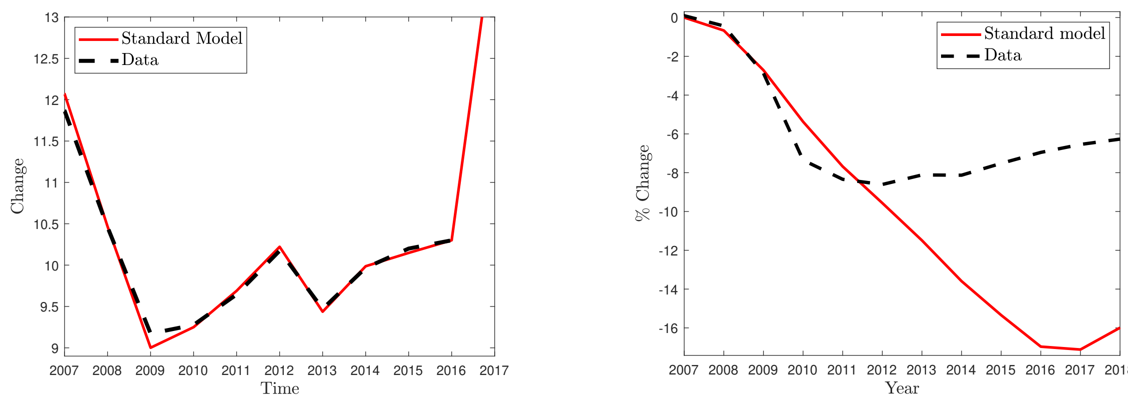
[55] In Appendix E.3, Table 18 summarizes the parameter values, and tables 19, and 20 summarize how the moments targeted in the Standard model compare with the data counterpart and the baseline model.

Table 11: Business Cycle Moments: Data and Model

		Data	Baseline	The Standard Model	
					only selection
		(a)	(b)	(c)	(d)
No. of firms	ρ	0.640	0.619	0.684	0.605
	σ	0.012	0.010	0.011	0.010
Employment	ρ	0.610	0.574	0.439	0.620
	σ	0.015	0.012	0.025	0.011
Entry Rate	ρ	0.250	**0.253**	**0.272**	0.265
	σ	0.062	**0.065**	**0.064**	0.063

Notes. The numbers that are in bold refer to the targeted model statistics. All other values indicate untargeted model statistics and their empirical counterparts.

Figure 16: The Great Recession and the Standard Model



(a) Entry rate

(b) Total employment

findings that emphasize the role of the pre-entry selection of firms in explaining cohorts' post-entry differences. Additionally, I use the Standard model to quantify the role of the entry margin in the anemic recovery observed after the Great Recession. Figure 16 shows that an aggregate demand shock series that generates the dynamics in the number of entrants observed over the period *2008 − 2016* leads to the drop in the aggregate employment that is twice as large as that observed during the Great Recession.

Next, I show that overlooking the observed variation in the entry margin undermines the role entry plays in propagating aggregate shocks. Following the existing literature, in the Standard model, I calibrate the aggregate demand shock process to match the business cycle fluctuations in aggregate employment, rather than the entry rate.[56] I find that in the Standard model, matching the observed persistence and variance of aggregate employment requires the auto correlation and the variance of the aggregate demand shock process to be, respectively, 1.40 and 25 times higher than a model that accounts for the documented variation in the entry margin.[57]

To sum up, the option to delay entry is an important mechanism that enables standard firm dynamics models to reconcile the observed business cycle demographics of entrants, and quantify the role the variation in the entry margin plays in aggregate fluctuations.[58]

7.2 Policy Implications

Potential entrants' ability to postpone entry not only quantitatively but also qualitatively alters existing firm dynamics models, predictions about the response of potential entrants to different shocks. The reason is the following. With the option to delay entry, the dynamics of entrants depends on how the changes in the aggregate environment affect 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. In this section, I illustrate the point by analyzing potential entrants' reactions to the permanent, temporary, and future reduction in the entry cost with and without the option to postpone entry.

Permanent versus temporary policy Figures 17(a) and 17(b) contrast the changes in the threshold signal level as a response to a permanent and a temporary decrease in the fixed entry cost with and without the option to delay entry.[59] 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

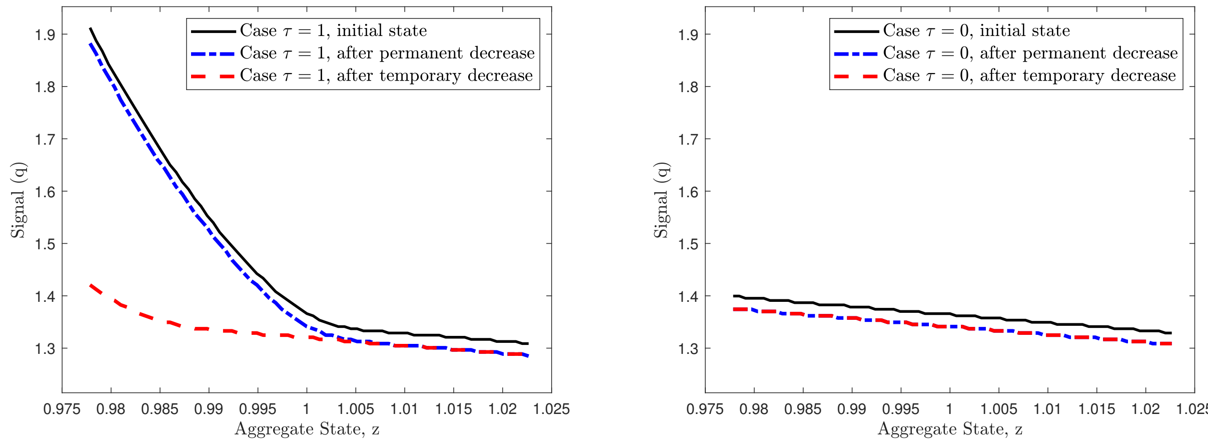
[56]For example, see Clementi and Palazzo (2016).

[57]Specifically, in the Standard model, the auto-correlation and the variance of the aggregate demand shock equal 0.80 and 0.05, respectively. In the baseline model, these values equal to 0.56 and 0.002, respectively.

[58]Even in general equilibrium settings, the model with persistent signal performs at least as good as standard firm dynamics models. The reason is as follows. The option value of delay is always non-negative, due to entrants' ability to obtain 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 a persistent signal than in the models without persistent signals. Appendix B.3 describes a general equilibrium version of the model.

[59]In Appendix G.3, Figures 48(c) and 48(d) translates the threshold signal into the number of entrants, using the assumed distribution $W(q)$ of potential entrants.

Figure 17: Permanent/Temporary Decline in the Fixed Entry Cost.



(a) Threshold signal, $\tau = 1$ (b) Threshold signal, $\tau = 0$

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 response of entrants does not vary with the duration of the policy, neither quantitatively nor qualitatively.

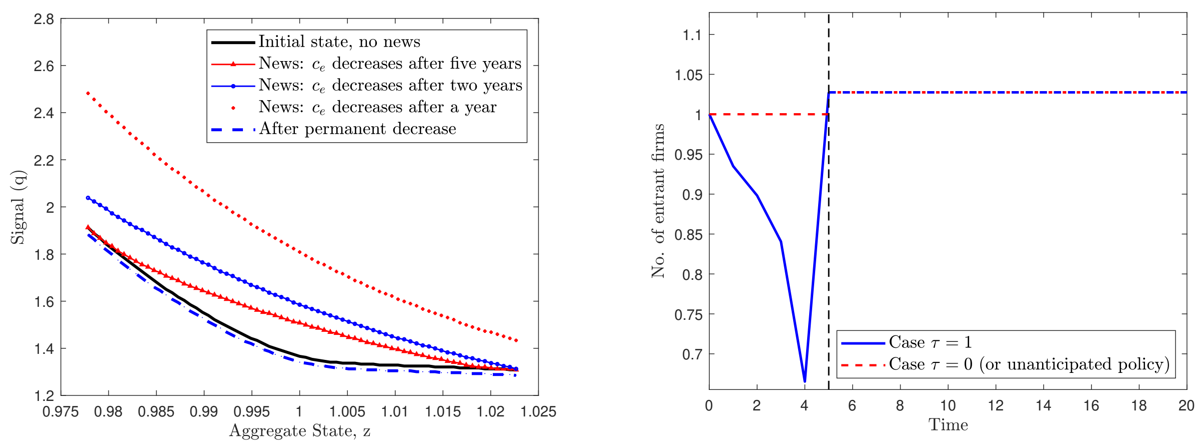
The news shock Now, consider he response of potential entrants to an anticipated decline in the fixed entry cost after T periods from today. Figure 18(a) shows that the threshold signal in the news scenario is weakly higher than in the baseline (no-news) scenario in all aggregate states and for all T. The magnitude of the change depends on the distance between today and the policy's actual time. Interestingly, 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.[60] 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; see Figure 18(b) that compares the dynamics of entrants in the steady state as a response of news about the decline in the fixed entry cost with and without the option.[61]

[60] Constantini and Melitz (2007) also show that potential entrants respond differently to the news about trade liberalization depending on the timing and the implementation of the policy.

[61] In Appendix G.3.2, I describe and illustrates the dynamics of the economy as a response to the announcement; see 49. Figure 50 illustrates the dynamics of the economy as a response to the announcement, in which I allow accumulation of potential entrants.

To conclude, after accounting for the ability to delay entry, the response of entrants to the changes in the aggregate environment depends on the relative variation in the benefits today versus tomorrow and any policy designed to affect entrants' behavior should take these channel into account.

Figure 18: News about the Anticipated Decline in the Fixed Entry Cost



(a) Threshold signal for different T and z

(b) Entrants' dynamics when $z = 1$ and $T = 5$

8 Conclusions

In the paper, I show that potential entrants' ability to delay entry leads to the countercyclical opportunity cost of entry and significantly amplifies the role initial aggregate conditions play in the selection of entrants. The feature allows existing firm dynamics models to reconcile the observed variation in the number and composition of entrants without generating the counterfactual variance of the aggregate variables. I propose a model that is able to reconcile the documented life-cycle dynamics of US establishments, on average, and over the business cycles. I find that the observed variation in the number and composition of firms at entry is responsible for around three-fourths of the business cycle fluctuations in aggregate employment. To validate these findings, I show the model accounts closely for the recent cohorts' contribution to the persistent drop in aggregate employment observed after the Great Recession. Finally, I show that not accounting for the option to delay entry may result in misleading predictions about the response of potential entrants to different shocks or policies.

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; How the heterogeneity in the ability to delay entry could explain the business cycle variation in the entry margins 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. In the paper, I study how allowing potential entrants to delay entry modifies their 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] Arrow, K., J. (1968). Optimal capital policy with irreversible investment. Institute for Mathematical Studies In the Social Sciences, Stanford, California. Technical report no. 146, 1–19.

[2] Ates, S. T., and Saffie, F. E. Fewer but Better: Sudden Stops, Firm Entry, and Financial Selection. American Economic Journal: Macroeconomics.

[3] 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. Manuscript.

[4] Bernanke, B., S. (1993). Irreversibility, Uncertainity and Cyclical Investment. The Quarterly Journal of Economics, 98 (1), 85–106,

[5] Bilbiie, F., Ghironi, F., and Melitz, M. (2012). Endogenous Entry, Product Variety, and Business Cycles. Journal of Political Economy, 120(2), 304-345.

[6] Clementi, G., Khan, A., Palazzo, B., and Thomas, J. (2014). Entry, exit and the shape of aggregate fluctuations in a general equilibrium model with capital heterogeneity. Manuscript.

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

[8] Cogley, T., and Nason, J. M. (1995). Output dynamics in real-business cycle models. The American Economic Review, 85(3), 492-511.

[9] Constantini, J., and Melitz, M. (2007). The dynamics of firm-level adjustment to trade liberalization. Manuscript.

[10] 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), 3-24.

[11] Dixit, A., and Pindyck R. (1994). Investment under uncertainty. Book. Princeton University Press.

[12] Fajgelbaum, P., Schaal, E., and Taschereau-Dumouche, M. (2017). Uncertainty traps. The Quarterly Journal of Economics, 132(4), 1641–1692.

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

[14] Fort, T., Haltiwanger, J., Jarmin, R., and Miranda, J. (2013). How firms respond to business cycles: the role of firm age and firm size. IMF Economic Review, 61(2).

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

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

[17] Geroski., P., E. (1995). What do we know about entry? International Journal of Industrial Organization, 13, 421-440.

[18] Gourio, F., Messer, T., and Siemer, M. (2016). Firm Entry and macroeconomic dynamics: a state-level analysis. The American Economic Review, 106(5), 214-218.

[19] Gourio, F., Messer, T., and Siemer, M. (2015). A missing generation of firms? Aggregate effects of the decline in new business formation. Manuscript.

[20] Haltiwanger, J., Jarmin, R., Kulick, R., and Miranda, J. (2016). High growth young firms: contribution to job, output and productivity growth. CES 16-49.

[21] Haltiwanger, J., Jarmin, R., and Miranda, J. (2013). Who creates jobs? Small versus large versus young. The Review of Economics and Statistics, 95(2), 347-261.

[22] Hong, S. (2018). Price markups for small and large firms over the business cycle. Federal Reserve Bank of St. Louis Economic Synopses, 17.

[23] Hopenhayn, H., A. (1992). Entry, exit, and firm dynamics in long run equilibrium. Econometrica, 1127-1150.

[24] Jarmin, R., and Miranda, J. (2002). The Longitudinal Business Database. CES Working Paper, US Census Bureau.

[25] Jovanovic, B. (2008). Investment options and business cycle. Journal of Economic Theory, 144, 2247-2265.

[26] Khan, A., Senga, T., and Thomas, J. (2019). Business formation and economic growth beyond the Great Recession. Manuscript.

[27] Khan, A., Senga, T., and Thomas, J. 2016. Default risk and aggregate fluctuations in an economy with production heterogeneity. Manuscript.

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

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

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

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

[32] McDonald, R., and Siegel, D. (1986). The Quarterly Journal of Economics, 101(4), 707-728.

[33] Moreira, S. (2015). Firm dynamics, persistent effects of entry conditions and business cycles. Manuscript.

[34] 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), 525-533.

[35] Pindyck, R., S. (1991). Irreversibility, uncertainty, and investment. Journal of Economic Literature, 29(3), 1110-1148.

[36] Pindyck, R., S. (2009). Sunk costs and risk-based barriers to entry. Manuscript.

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

[38] Pugsley, B., Sedlaćek, P., and Sterk, V. (2017). The Nature of Firm Growth. manuscript.

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

[40] Sedlaćek, P., and Sterk, V. (2017). The Growth Potential of Startups over the Business Cycle. American Economic Review, 107(10), 3182-3210.

[41] Sedlaćek, P. (2019). Lost generations of firms and aggregate labor market dynamics. Journal of Monetary Economics, 111, 16-31.

[42] Siemer, M. (2016). Firm entry and employment dynamics during the Great Recession. Manuscript.

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

Appendix (For Online Publication)

Table of Contents

A Empirical Findings Appendix

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

In this section, I investigate how does the cohorts average size over the life cycle varies with the aggregate conditions at entry. I use the same methodology for the survival rate analysis. I use country- and state-level annual time series about the cohorts of establishments/firms by age from the Business Dynamics Statistics (BDS) dataset. The BDS dataset covers the universe of employer businesses in the US and provides annual measures of business dynamics for the economy 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 span 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

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

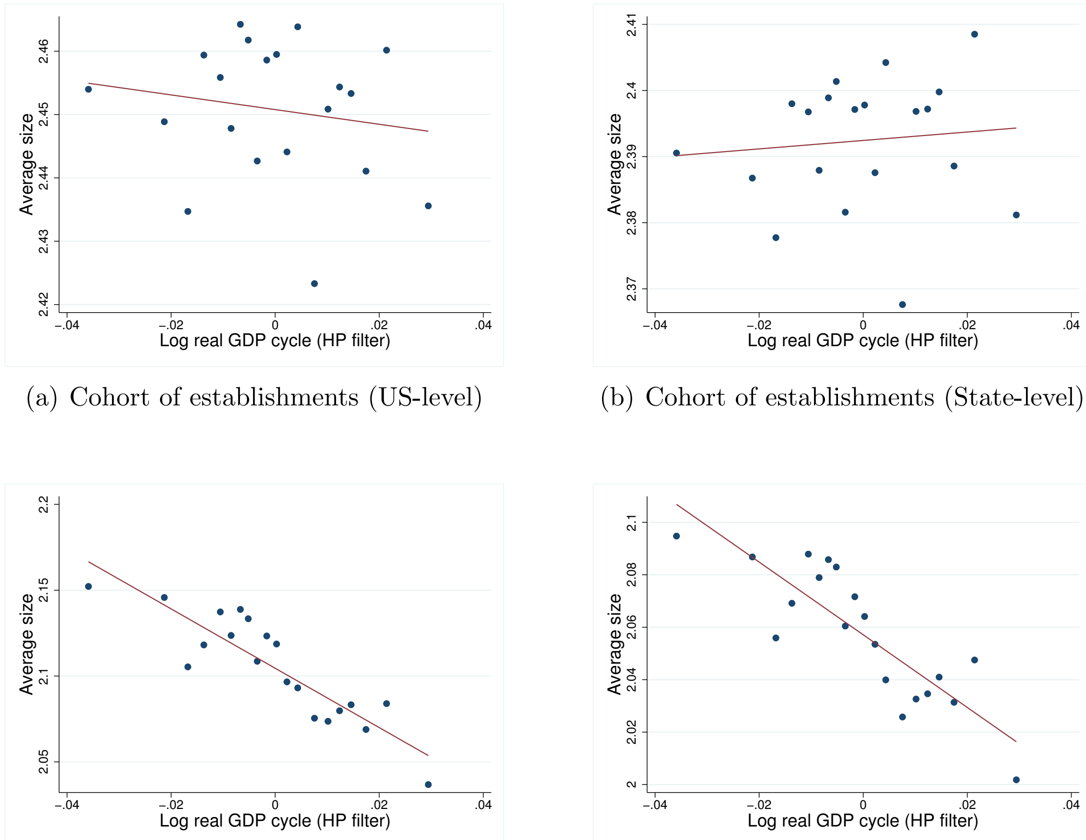
where $L_{g,t}$ and $N_{g,t}$ measure the total employment and total number of establishments (firms) in a cohort of establishments (firms) of age g at time t. $\bar{L}_{g,t}$ measures the average size of a cohort of establishments (firms) of age g at year t; In this analysis, I consider cohorts' average size for up to age 5 after they enter the market.[62] To investigate how cohorts' average size over the life cycle vary with the aggregate conditions at the time of entry $t - g$, I use the cycle component of the annual real GDP.[63] To find the cyclical component of the annual log real GDP I apply the HP filter with a smoothing parameter of *100*.

Figure 2 plots pooled average size for up to six years of operation of cohorts at the US- and state-level against the business cycle indicators at the time of entry. Panel (a) of Figure 2 shows that the business cycle conditions at entry are positively associated with cohorts' of establishments' average size at the US level. However, Panel (b) shows that the correlation becomes negative at the state level. Panel (c) and Panel (d) consider cohorts of firms at the US- and state-level. In both cases, aggregate conditions at entry are negatively associated

[62] 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.

[63] I annualize the quarterly real GDP data so that its consistent with BDS timing. Specifically, in the BDS dataset, 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. The source and the construction of the annual real GDP data are described in Appendix E.2.

Figure 19: Cohorts' average size against the aggregate economic conditions at the time of 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)

Notes. Each panel plots a binned scatterplots of the average size up to age 5 against the aggregate conditions at the time of entry measured by the cycle component of HP-filtered log real GDP with smoothing parameter of 100. Bin scatter controls for year- and age-fixed effects. Panels (b) and (d) additionally control for the state-fixed effects.

with the average size. That is, increase in aggregate conditions at entry above trend is associated with the cohorts 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:

$$log(\bar{L}_{g,t}) = \alpha + \beta Z_{t-g} + \eta_a + \theta_t + \gamma_s + \varepsilon_{g,s,t}, \tag{9}$$

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 the economic conditions at the time when the cohort first entered the market.[64] η_a, θ_t, γ_s

[64] 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 12: 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. The dependent variable represents pooled survival rates up to five years of operations of cohorts of new businesses. Panel A uses cohorts of establishments and Panel B uses cohorts of firms as a unit of analysis. Columns (1)-(3) use different indicators for characterizing the business cycle conditions at entry. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

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 the 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 the cohorts' average size after controlling for the age, year and state fixed effects.

Panel A of Table 12 reports the results of the regression equation (9) when the unit of analysis is a cohort of establishments. Column (1) indicates that cohorts born during good economic conditions are characterized by higher average size over the life cycle, but the 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 recession indicator for the US from the period following the peak through the trough ($NBER$).[65] The indicator equals one if the year is indicated as recession, 0 otherwise. Columns (2) and (3) of Panel A show that cohorts born during recessions, on average, have lower average size compared to their expansionary counterparts. Thus, there is no robust statistically significant relationship between cohorts of establishments average size and aggregate conditions at entry. Panel B of Table 12 shows that there is statistically significant negative relationship between aggregate conditions at entry and cohorts of firms average size over the life cycle. establishments.

[65] The latter indicator specifies peak and the 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.

Table 13: Cohorts average size by age and aggregate economic conditions at the time of 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. The dependent variable represents pooled average size from age zero to age five of cohorts of new businesses. Panel A uses cohorts of establishments and Panel B uses cohorts of firms as a unit of analysis. Columns (1)-(3) use different indicators for characterizing the business cycle conditions at entry. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

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

$$log(\bar{L}_{g,t}) = \alpha + \sum_{g=1}^{5} \beta_g D_g Z_{t-g} + \eta_g + \theta_t + \gamma_s + \varepsilon_{g,s,t}, \tag{10}$$

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 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 13 reports the regression results. $1_{\{age=g\}} \times Z$ describes the interaction of the business cycle indicators with the cohort of age g. Panel A describes the relationship between the average size of cohorts of establishments and aggregate conditions at entry. Analyzing Columns (1)-(3) show no statistically robust relationship between average size and aggregate conditions at entry for the cohorts of establishments. Panel B considers the same regression

All other years are defined as expansionary.

when the unit of analysis is cohorts of firms rather than the cohorts of establishments. Panel B shows 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 compared to their expansionary counterparts. However, one can see that the effect dissipates over time when the cohort age.

A.2 Cyclicality of Business Formation

A.2.1 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.[66] 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. The business applications are matched to the set of firms in Business Dynamics Statistics Dataset (BDS) identified as new employer businesses based on payroll information.[67] Match process is straightforward since both of the 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 (BF4Q)* - the number of employer businesses that originate from the business applications within four quarters from the quarter of application. Time period: 2004Q3-2015Q4. In the analysis, I refer to this time series as First 4Q.

2. *Business formations within 8 quarters (BF8Q)* - the number of employer businesses that originate from the business applications (BA) within eight quarters from the quarter of application. Time period: 2004Q3-2014Q4.

3. *Average duration (in quarters) from business application to formation within 4 Quarters (DUR4Q)* - a measure of delay between business application and formation, conditional on business formation within four quarters. Time period: 2004Q3-2015Q4. I refer to this time series as Av. duration, first 4Q.

4. *Average duration (in quarters) from business application for formation within eight quarters (DUR8Q* - a measure of delay between business application and formation, conditional on business formation within eight quarters. Time period: 2004Q3-2014Q4.

[66] EIN is a unique number assigned to most of the business entities. 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 new EIN.

[67] The BDS dataset covers the universe of employer businesses in the U.S. and provides annual measures of business dynamics for the economy aggregated by the establishment and firm characteristics. Employer businesses are identified as start-ups based on their first payroll information by Longitudinal Business Database.The

Additionally, I construct the following two-time series:

5. *Business formations within the second half of eight quarters (Second 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:

$$\textit{Share of late start-ups} = \frac{BF8Q - BF4Q}{BF8Q}$$

7. *Average Duration (in Quarters) from Business Application to Formation from 5 to 8 Quarters:* average duration it takes for the group of applications that need to form business for more than 4 quarters. To re-construct this information using the following formula:

$$DUR(second4Q) = \frac{DUR8Q\ BF8Q - DUR4Q\ BF4Q_t}{BF8Q - BF4Q}$$

Summary Statistics The summary statistics of the considered time series are given in Table 14 and Table 15. Several facts stand out. (1) The average share of the applications that start business in the second four quarters equals 13.68%. The time series varies from 10.96% to 17.73%; see Table 16. (2) The business applications that form businesses within the first four quarters do so in the first two quarters. Specifically, it takes, on average, from five to six months to form an employer business from the date of the application. (3) The group of the business applications that form employer business with four quarter delay, do so, on average, in sixth quarters.

Table 14: Summary Statistics for Quarterly Business Formation (Thousands)

	Mean	St. Deviation	Min	Max
BF8Q	97.5208	18.1831	80.3434	134.1869
First 4Q	83.4509	17.0360	68.3442	119.4842
Second 4Q	13.0668	1.1480	11.3703	15.2153
Share	0.1368	0.0191	0.1096	0.1773

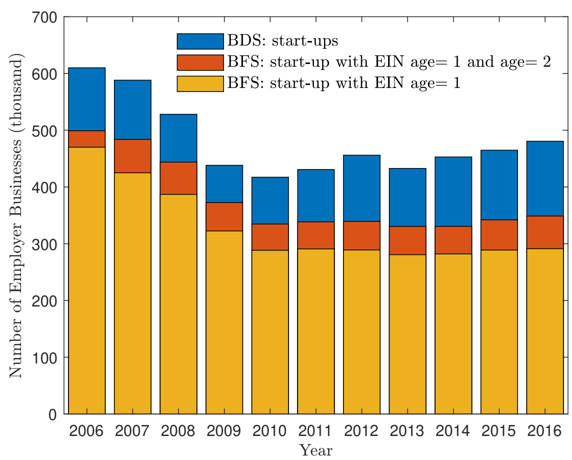
Table 15: Summary Statistics for Average Delays (in Quarters)

	Mean	St. Deviation	Min	Max
DUR8Q	1.66	0.16	1.39	1.93
DUR4Q	1.66	0.16	1.39	1.93
DUR(second 4Q)	5.46	0.11	5.22	5.78

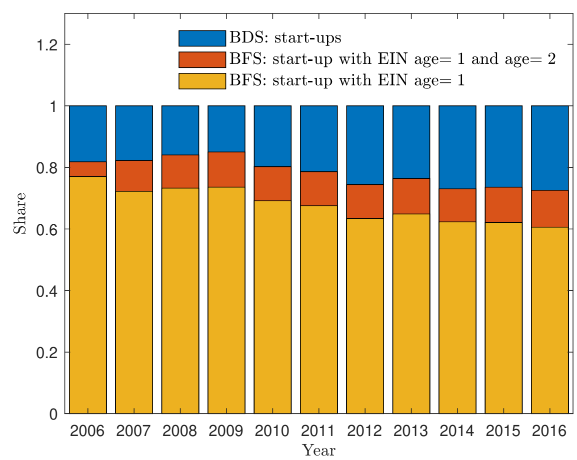
A.2.2 Comparability of the Publicly Available BFS dataset with the BDS

All firms that show up in the BDS have EINs. Thus, they show up in the BFS dataset before entry.[68] 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.

Figure 20: BDS and BFS



(a) The number of business formation



(b) The share of business formation

Notes. The figure shows the annual total number of employer business start-ups from 2005 to 2016 from the BDS and from the BFS. The number of employer birth from the BDS is constructed from the number of employer birth within eight-quarters window.

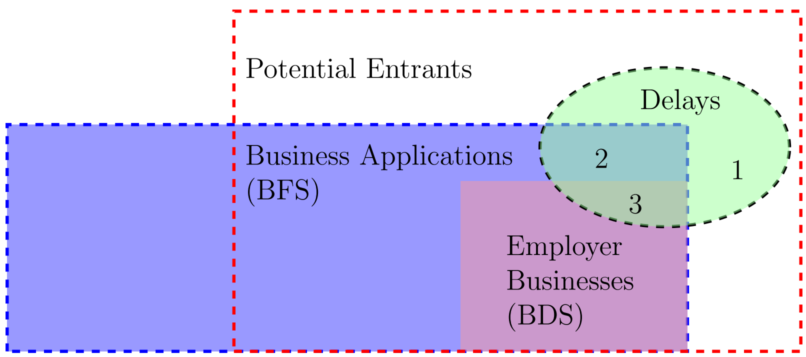
I compare the information in the BFS dataset to the one provided by the BDS dataset. Toward that end, I convert the quarterly data from the BFS into yearly time series. I defined business formation for a year t as the total number of businesses generated from the cohort of applications applied within the first quarter of year t to the fourth quarter of year t. The average duration of the business formation within four quarters happen within 1.5 quarters. In that case, the applications from the fourth quarter of year t are going to become employer business before March 12 and show up in the BDS dataset. Figure 20 shows that these employer businesses comprise more than 80% of the total start-ups in the BDS.

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

A.2.3 Discussion

The goal of the empirical section is to identify how the aggregate conditions at entry affect business formation through the "wait-and-see" channel of the entry decision. To explain the identification strategy, Figure 21 illustrates the potential links between the Business Formation Statistics Dataset (BFS), the Business Dynamics Statistics Dataset (BDS), and potential entrants in the model. I use the diagram to discuss also the relevance of different parts of the BFS dataset in answering the question.

Figure 21: The potential links between the Business Formation Statistics Dataset (BFS), the Business Dynamics Statistics Dataset (BDS), and potential entrants in the model.



Notes. The figure illustrates potential links between the BDS, the BFS datasets, and the potential entrants that could potentially choose to delay entry. Segment 1: Potential entrants who decide to delay entry and do not apply for the EIN. Segment 2: Potential entrants who apply for the EIN, decide to delay entry, and never start a business. Segment 3: The potential entrants that applied for the EIN, decide to delay entry and come back in the market after a year.

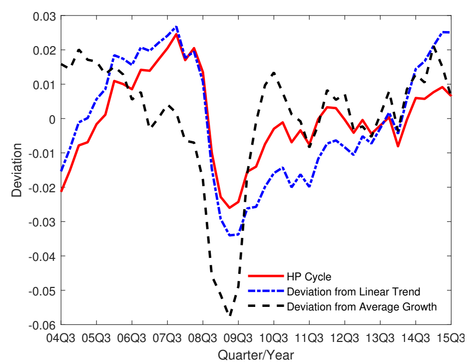
The potential entrants that delay entry could belong to the following three groups. First, the group of potential entrants that delay entry and also delay applying for the EIN; Second, the group of potential entrants that apply for the EIN, delay starting a business at least first eight quarters from 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.

Initially, I argue that the first and the second group of entrants can not be identified 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 part 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. In particular, 12% become employer businesses in the first four quarters, and an additional 2% become employer businesses after a year. 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) claim that the lower transition rates is due to the fact that a major 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 time it takes for the third group of entrants to become employer businesses to identify delays in potential entrants entry decisions.

A.2.4 Time Series for the Business Cycle Conditions

Figure 22: The time series of the business cycle indicators



A.2.5 Robustness: annual data

In this section, I repeat the analyses for the annualized business formation time series. I construct the following time series: 1. The annual number of applications that form a business in a year ($BF1Y$); The annual number of applications that form businesses within two years ($BF2Y$); The annual number of applications that form businesses after a year from the date of the applications ($BF2Y$); The share of the business applications that form business after a year in the total number of business applications that form business within two years ($Share$);

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

covers the period $2006-2016$, and the time series for $BF2Y$ covers the period $2006-2015$.[69]

Summary statistics for the annual time series is given in Table 16. For comparison, the table also reports summary statistics for the employer business start-ups from the BDS dataset.

Table 16: Summary statistics for yearly business formation (thousands)

	Mean	St. Deviation	Min	Max
Firms (BDS)	491.4534	70.8420	417.2020	610.006
BF in 2 years	376.0336	62.5343	330.7949	505.902
First year	326.2789	59.6975	281.5538	462.239
Second year	51.2315	4.8027	43.6623	59.377

Table 17: Correlations between the cyclical variation in the business application time series with the business cycle indicators

	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 in 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)

Cyclicality of the business formation at annual frequency In this section I study the cycle properties of the annual business formation data. Table 17 reports the results. The results implies that during the recessionary periods the number of applications that form business within a year decreases. The subset of the applications that take more than a year to form a business also decreases if the initial state in the year of entry is recession. On the other hand, the share of the applications that from business in one year delay increases in the total number of applications that form businesses in two years. Since we saw that the variation in the average duration of delays does not exceeds to two month, at least the part of the countercyclical variation in share supports to the "wait and see" channel of business formation.

[69] $BF4Q$ and $BF8Q$ data starts from the year $2004Q4$ which means that for the year 2005 only three-quarters of application data is available ($2004Q3 + 2004Q4 + 2005Q1$). Since I do not have the complete number of applications for the year 2005 I had to drop from the analyzes

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 $q \sim W(q)$.[70]

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 which decomposes entrants between aspiring start ups, those that want to start a business and potential entrants that actual hold business ideas and enter the market.

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. If a potential entrant with a signal q postpones entry to the next period, the potential entrant gets the same signal tomorrow with a probability $\tau \in [0, 1]$. With a probability $(1 - \tau)$, the potential entrant loses the signal and drops out from the pool of potential entrants.

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

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.[71] The total mass of available business opportunities equal to the total number of business opportunities within each signal category $W(q)$ minus the mass of business opportunities that is held by the group of potential entrants that delayed entry in the previous periods. $n_t(q)$ refers to a number of aspiring startups competing for the business opportunities with signal q. The ratio in the equation represents a probability by which an individual receives a signal q and becomes a potential entrant.[72] $V^e(q,z_t)$ 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 $\dfrac{B_t(q)}{n_t(q)} V^o(q,z_t) = c_q$.

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 opportunities equals to

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

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 variation of $N_{t,\text{aspiring startups}}$ over the cycles depends 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 3.1.1.

[71] $0 < B_t(q) < W(q)$

[72] $0 \leq \dfrac{B_t(q)}{n_t(q)} \leq 1.$

Parametrization of the entry phase To parametrize the entry phase I use information from the newly developed Business Formation Statistics dataset that collects information about business applications and formation. Business application data is based on applications for Employer Identification Number (EINs) filed in the United States. From the business applications only around 12% transitions into employer businesses within the first year, and 14% in the second year from the date of the application.

In the entry phase described above the number of applications can be considered 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 indicates that only additional 2% of the applications transitions into employer businesses in the following year. In terms of the model setup the fact implies that $B(q)$ is close to $W(q)$; only few potential entrants that decide to delay entry enters 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 the future research.

Note that the entry phase also can 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 wants to start business and actually applies for the EIN, from the potential entrants, those that actually hold business ideas and make entry decisions. According to the model the restricted number of actual business ideas does not allow most of the aspiring start-ups to enter the market.

Interestingly, the simple modification of entry phase developed in Lee and Mukoyama (2008) could also address an additional challenge faced by the firm dynamics models developed in general equilibrium: the pro-cyclical variation in the wages and the free entry condition mitigate the effect of the aggregate conditions on the selection of entrants. However, in the set up above, allowing aspiring start ups to compete in a specific signal categories endogenously restricts entry margin to have an effect on aggregate prices. The business cycle variation in the willingness to start a business is absorbed at the free entry stage due to changes in the probability of becoming a potential entrant.

To conclude, the restriction on the number of available business opportunities implies that the aggregate distribution of potential entrants are constant over time, and accumulation of the entrants happens at aspiring start up level.

B.2 Accumulation of Potential Entrants

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 entrants, that decide to delay entry, modifies entrants' characteristics over the cycles and affects the dynamics of aggregate variables. I find that cohorts that enter during different aggregate economic conditions have significantly and persistently different characteristics, even after allowing the accumulation of potential entrants over time.

Figure 23: New potential entrants, $W(q)$

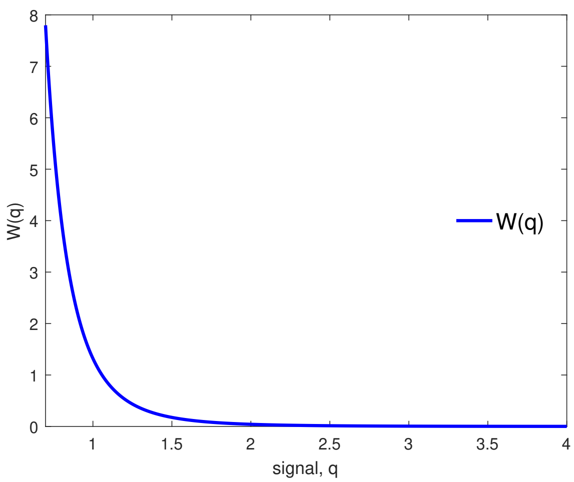
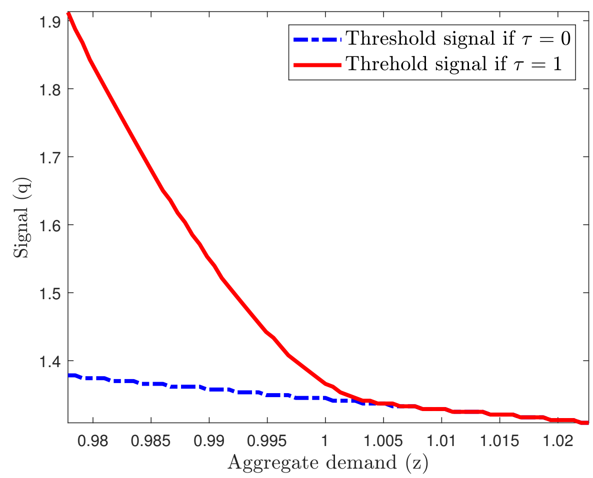


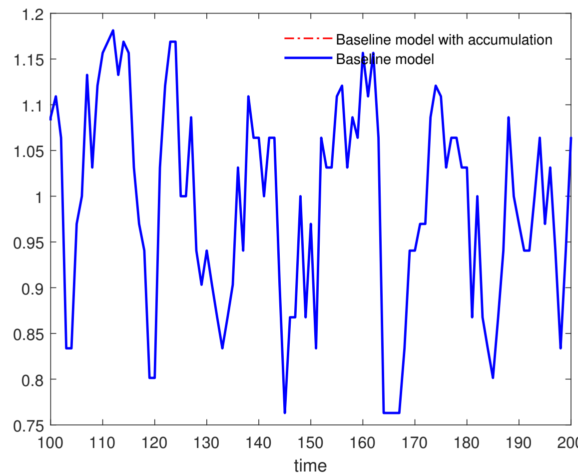
Figure 24: Threshold signal



In the baseline model, in every period, the distribution of new potential entrants, which make entry decisions for the first time, is equal to the distribution of potential entrants entering the market in the previous period. The assumption ensures that the number of potential entrants is constant over time. 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 23. In addition to the new potential entrants, the aggregate distribution of potential entrants consists of old potential entrants. Old potential entrants who chose to delay entry in the previous periods, while their expected value of being an incumbent was more than zero.[73] Figure 24 displays the threshold signal, $q^*_\tau(z)$ for each

[73] Consistent to the baseline model I keep $\tau = 1$, which means that potential entrants that delay entry can

Figure 25: Entrant firms



aggregate state when $\tau = 0$ (blue-dashed line) and $\tau = 1$ (solid red line). For given z, potential entrants that decide to delay entry hold signals in between $[q^*_{\tau=0}(z) \quad q^*_{\tau=1}(z)]$.

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 $\Lambda_t^{\text{old entrants}}(q)$.

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

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 the entrants in the baseline model to a model that allows the accumulation of potential entrants. Note that when the aggregate demand decreases from z_{t-1} to z_t Then, the distribution/number of entrants in the baseline model and the model with signal accumulation coincide. If potential entrants delayed entry when the aggregate state was z_{t-1}, nobody from these old potential entrants is going to enter in an aggregate state z_t' ($< z_{t-1}$). As a result, selecting potential entrants at entry happens only from the distribution of new potential entrants, like in the baseline model. However, if the aggregate

keep the signal forever.

Figure 26: Average productivity

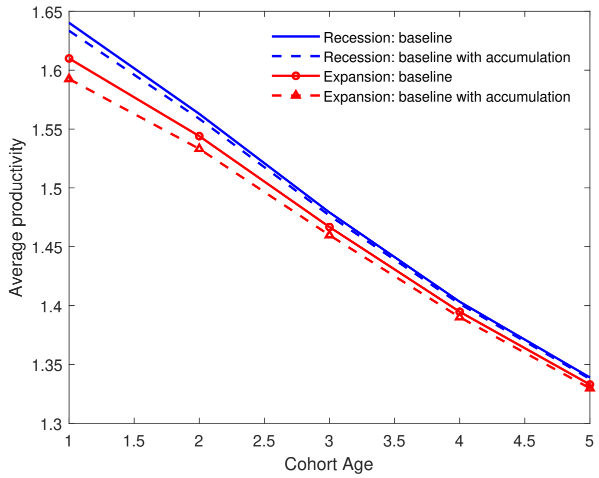


Figure 27: Average survival rate

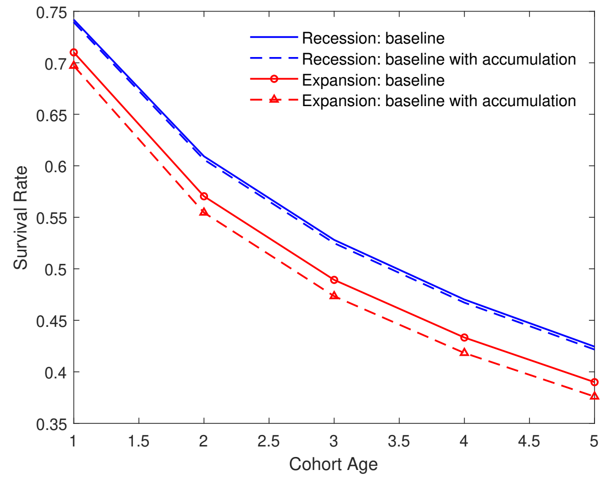
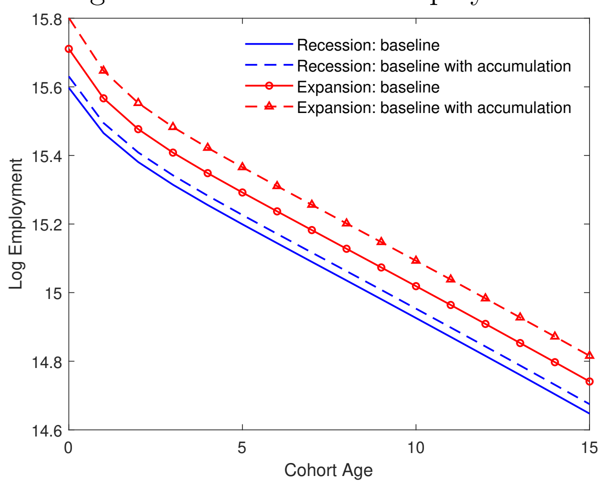


Figure 28: Cohort-level employment



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.

It turns out that the increase in the number of entrants during expansions outweighs the increase in the number of entrants during recessions, and extending the baseline model to account for the accumulation of potential entrants increase procyclical variation in the entry rate. Moreover, the differences in cohorts' characteristics that start operating during different aggregate economic conditions increase after allowing potential entrants to accumulate. The latter feature modifies the distribution of the entrants over the cycles in the following way. During recessions, defined as periods when $log(z) < 0$, potential entrants that decide to start a business hold lower signals on average compared to the baseline scenario. Consequently, as shown in Figure 26 and Figure 27 the average productivity and the average survival rates of the cohorts that enter the market during recessions decrease compared to the baseline scenario. The accumulated groups of old potential entrants, on average, hold less productive signals, and most of them end up low-productivity firms after entering the market.

Consequently, average productivity and survival rate decreases significantly during expansionary periods compared to the baseline scenario. Altogether, the extension produces countercyclical average productivity and survival rates. Moreover, compared to the baseline model, the differences between the cohorts' post-entry characteristics that start operating at different aggregate conditions increases.

Allowing accumulation of potential entrants over time increases recessionary as well as expansionary cohorts employment compared to the baseline model. However, since the number of entrants significantly increases during expansionary periods the difference between recessionary and expansionary cohorts employment increases compared to the baseline scenario.

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 Set up

B.3.2 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:

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

such that

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

B.3.3 The Mutual Fund

The household owns shares in the mutual fund. The mutual fund consists of the heterogeneous of 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 factor. 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.

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

s.t.

$$y_t^s = s_t n_t;$$

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

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

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

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

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

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:

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

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

$$\Lambda_t = \int\limits_s \int\limits_b V(s, b, z)\Omega(b, s, z)dsdb + \int\limits_{q_*}^{\infty} \int\limits_s V(b_o, s, z)H(s|q)W(q)dq.$$

Denote $N_{e,t}$ be the number of entrants in period t, than: $N_{e,t} = \int\limits_{q_*}^{\infty} W(q)dq$. 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.3.4 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.