The London School of Economics and Political Science

Essays on Firms Heterogeneity and Business Cycles

Andrea Alati

A thesis submitted to the Department of Economics for the degree of Doctor of Philosophy, London, September 2020

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is also the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 34,000 words, excluding figures and tables' captions.

Statement of conjoint work

I confirm that Chapter 3 is joint work with Edoardo Maria Acabbi and Luca Mazzone, and I contributed one-third of this work.

Abstract

This thesis consists of three chapters in macroeconomics. They study the effects of business cycles on heterogeneous firms and workers and their consequences for the aggregate economy.

The first chapter provides the empirical evidence used to motivate and calibrate the theoretical framework developed in the second chapter, discusses the empirical methodology and provides robustness checks to address its limitations. The paper documents two facts. First, aggregate conditions close to the listing year negatively affect firm-level markups. Second, these effects are long-lasting but not permanent.

The second chapter explores the role of aggregate fluctuations as a persistent determinant of heterogeneity in firm-level markups. Informed by firm-level estimates of markups, the paper builds a general equilibrium model that features heterogeneous product markets, customer base accumulation, and firm dynamics. In the model, firms' demands are constrained by the size of their customer bases. Firms can accumulate customers using two complementary channels: i) increasing sales by lowering prices and ii) making direct investments in customer acquisition. As the value of operating in each product market fluctuates endogenously with business cycles, aggregate conditions generate a selection of the product market composition of the cohorts of active firms that induces time-varying heterogeneity in their cross-section. This heterogeneity is persistent and significantly affects the transmission of aggregate shocks to the economy and the co-movements of aggregate markups with business cycles.

The third chapter develops a structural model of the labor market that features both worker and firm heterogeneity and where workers accumulate human capital and search on the job. The paper analyzes the optimal provision of insurance within the firm through an optimal dynamic contract that, paired with limited liability on the firm side, implies downward wage rigidity. In this framework, insurance incentives and contractual rigidities are crucial in determining the pattern of job matches and separations along the business cycle. In particular, we show that aggregate fluctuations can alter the sorting between workers and firms by affecting workers' search strategies and, as a consequence, distort their human capital accumulation.

Acknowledgements

I am deeply indebted to my supervisor, Wouter den Haan, for his constant support and for the many hours spent in discussing my ideas and his patience in making sense of my preliminary writings. I have learned a lot from his guidance and approach to macroeconomics and research in general. I am also very thankful for having benefitted from many interactions with Ricardo Reis that have deeply influenced my growth as a researcher.

For the many helpful discussions and comments about my projects I would like to thank Francesco Caselli, Ethan Ilzetzki, Andrea Lanteri, Ben Moll, John Moore, Rachel Ngai, Silvana Tenreyro, Gregory Thwaites, Shengxing Zhang and the many participants to the CfM Work in Progress seminars.

A large part of this thesis has been written during the COVID-19 lockdown. During this time, I received a tangible proof of how strongly the PhD experience relies on the frequent contact with fellow students. In my years at the LSE I have been lucky to have benefitted of the constant help of brilliant researchers. In particular, I would like to thank Thomas Drechsel and Laura Castillo-Martinez for having "funneled" their tricks of the trade to our younger cohort many times; Adrien Bussy, Miguel Bandeira and Friedrich Geiecke for the many coffees, office banter and great discussions; Fabio Bertolotti, Luca Citino, Martina Fazio, Nicola Fontana, Vincenzo Scrutinio, Alessandro Sforza, Tommaso Sonno and Martina Zanella for the everyday chats, Friday beers and constant help.

I would also like to thank Edoardo Acabbi and Luca Mazzone for constantly sharing the many up and downs of research, and Chiara Fumagalli and Giovanni Pica for having introduced us to research in Economics and initiating this journey.

A special thank goes to my friends outside the program, Luca, Stefi and Max, Alex, Vicky and Matías, Carli, Leen, Nico, Cri and Davi, for helping me put the PhD in perspective and offering many needed escapes from research. Without them, my years in London would have had a lot less laughters, delicious dinners and Catan games.

My parents and my brother have been a never-ending source of encouragement and during these years they have always helped me more than they probably realize. I thank them for all they have always done for me.

Finally, my most special thought is for Euge and the constant joy, love and unwavering support that she brings in my life.

This thesis is dedicated to her.

Contents

List of Figures

List of Tables

Chapter 1

Persistent effects of business cycles: an empirical analysis of firm-level markups

Abstract

Is it possible for business cycles to persistently affect firm-level markups? In this paper, I use an age-period-cohort model to estimate cohort effects on markups for a sample of US listed firms. I proxy cohort fixed effects with aggregate conditions close to the time of listing and I document two new facts. First, negative aggregate conditions close to the time of listing are associated with higher firm-level markups. Second, these effects, contrary to what the firm dynamics literature documents on firm size, are long-lived but not permanent fading away after approximately fifteen years. To ensure that the difficulties linked to the measurement of markups are not confounding the results I complement the main analysis with a battery of robustness checks on the measurement of markups.

1.1 Introduction

Aggregate conditions can have persistent effects on many aspects of the economy and deeply influence agents' choices. An extensive labor literature, for example, documents that graduating or losing a job in a recession leads to significant and long-lasting effects on career earnings.[1] Similarly, recent studies find that business cycles can select the types of firms that are active in the economy, causing persistent size differences across firms belonging to different cohorts.[2]

In this paper, I focus the attention to the interaction between firm-level markups and aggregate conditions at key junctures of firms' lives. In particular, can business cycles have a persistent impact also on how firms decide to price their products? Is it possible for aggregate conditions, at particular times of firm lives, to permanently affect their markups? The relevance of these questions hinges on the fact that markups are a central variable in macroeconomic models.[3] Nevertheless, their behavior along business cycles and the possible effects that aggregate fluctuations have on them is still debated in the literature. Therefore, understanding how business cycles interact with markups at the firm level can foster our knowledge of how aggregate fluctuations can persistently influence firms' behavior and what are the relevant features of firms' incentive structures that have to be considered when modeling firm life-cycles.[4]

I dedicate this chapter to a comprehensive discussion of the effects of business cycles on firm-level markups and I present the empirical evidence that I then use to rationalize and calibrate the theoretical framework developed in Chapter 2. In addition, I provide details on the empirical methods used to recover markups from balance sheet data, discuss their limitations and provide a battery of robustness checks to test their solidity.

[1]See for example Schwandt and Von Wachter [2019], Oreopoulos *et al.* [2012], Kahn [2010].

[2]See in particular Sedláček and Sterk [2017], Moreira [2015], Pugsley *et al.* [2019].

[3]For example, the response of markups is a key channel in the transmission of aggregate demand and monetary policy shocks in the New-Keynesian models.

[4]As an example, in Chapter 2, I discuss a theoretical framework in which aggregate conditions persistently influence the level of markups a firm can charge by affecting its listing decision.

Empirically assessing the behavior of markups is challenging, particularly at the firm level.[5] Theoretically, a researcher interested in measuring the markup charged by a firm would need to observe both prices and marginal costs. Typically, however, it is difficult to obtain data on both of these quantities. To overcome these issues, I follow empirical methods developed in the industrial organization literature that allow the estimation of markups directly from firm-level data. Despite coming with substantial limitations, these approaches are particularly helpful as they allow to exploit firms' optimal behavior to recover a measure of markups directly from balance sheet data without requiring the specification of an explicit demand system.

To quantify the persistence of initial aggregate conditions, I adapt a cohort-age-period model to estimate the effects of aggregate conditions close to the time of listing on firm-level markups. This particular econometric framework has been used in the labor literature to estimate the persistence of aggregate conditions at the time of graduation on workers' career earnings (Schwandt and Von Wachter [2019], Altonji *et al.* [2016], Oreopoulos *et al.* [2012], Kahn [2010]). With few modifications, I adopt this empirical strategy to estimate the persistence of aggregate conditions close to the time of listing on firm-level markups.

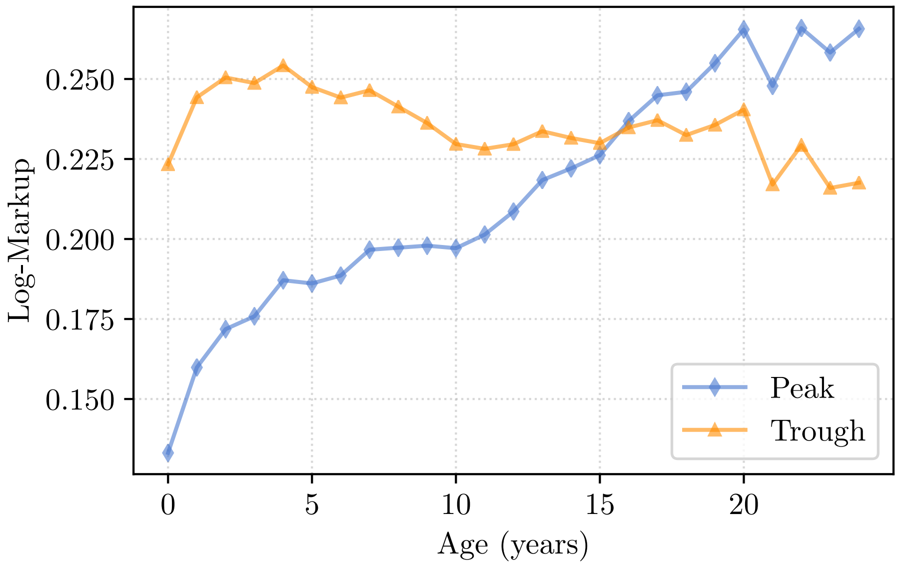
I estimate markups at the firm level for an extended panel of US-listed firms and I use these estimates to document two main facts.

First, I show that, at the firm level, markups are characterized by an increasing age profile that reaches a plateau after approximately 20 years.

Second, I document the effects of aggregate conditions close to the time of listing on the age profile of firm-level markups. As shown by Figure 1.1, I find that firms that start their listing process in periods of low aggregate activity exhibit higher initial markups and a flatter age profile compared to similar firms that face better aggregate conditions close to their listing time. The magnitude of the estimated effect is statistically and economically relevant, albeit smaller than the effects on firms' sizes reported by the literature. In particular, the estimated effects imply that firms that were close to listing at the height of the Great Recession, in 2009, are associated with average

[5] See Nekarda and Ramey [2013] for a review of the difficulties in measuring markups and their correlations with business cycles.

Figure 1.1: Cohort effects on markups' age profile



Note: The figure plots the age profile for markups estimated from equation (1.9). Specifically, at each age a, I am plotting $\hat{\mu}_0 + \hat{\phi}_a + (\hat{\beta}_0 + \hat{\beta}_1 a)Z$, where $\hat{\mu}_0$ is the average log-markup in the first available year, $\hat{\phi}_a$ are the estimated age fixed effects, $(\hat{\beta}_0, \hat{\beta}_1)$ capture the estimated persistence of aggregate conditions in the first available year. Z is a measure of business cycle outcomes that takes value 2σ for *Peak* and -2σ for *Trough*, with σ being the standard deviation of quadratically detrended real GDP. The coefficients on which this plot is based are reported in Table 1.2, Column (2).

markups approximately between 1.4% and 3% higher than the ones of similar firms that were close to listing in 2007.

It is worth noticing, however, that these effects are consistent with two alternative interpretations that could persistently link markups to business cycles at key moments of firms' lives. On one hand, it is possible that aggregate conditions close to the time of listing directly influence firms' business decisions in a way that is reflected in their markups. For example, they influence their product development and marketing mixes. On the other hand, it is also possible that different types of firms, producing different goods or competing in different markets (e.g. niche versus mass products), assess the profitability of going public differently at different stages of the business cycle. While the former interpretation implies that firms actively change their behavior after listing, and favors a causal interpretation of the effect of business cycles on markups, the latter suggests that business cycles reflect ex-ante differences in firm types that induce a selection into listing at specific times. Given the difficulty in disentangling these two alternatives, throughout the chapter I do not take a stance on which of the two prevails. However, despite the difficulties in attributing a causal relationship to the effects estimated in this paper, they are still useful as indicators of relevant firm-characteristics that are helpful in advancing our understanding of the incentive structure that firms face when managing their markups over the course of their lives.

I supplement the empirical analysis with a battery of robustness checks to address two significant issues inherent to the nature of the empirical exercise carried out in the paper. These issues are linked to the difficulties in measuring the output elasticity to variable inputs and the reliance of markup measures on the availability of variable costs data. First, I address the difficulty of measuring marginal costs by estimating markups using two alternative measures of costs and I check how different estimates of the output elasticity influence markups by estimating three different production functions - a sector varying Translog specification, a sector and sector-time varying Cobb-Douglas product functions as well as a simple cost-share approach. Second, as an additional robustness check, I complement the primary analysis by considering a sub-sample of firms for which I am able to retrieve founding dates. For these sub-sample, therefore I am able to assign firms to cohorts using their founding year rather than the first year of available accounting data as in the baseline sample.[6] For this sample, the average effect is qualitatively similar to the one estimated using the listing year but the effect on the age profile is slightly different. While the effects of business cycles vanish over time when cohorts are defined on their listing time, the effect of business cycles is stronger and more persistent for firms that I can follow since their initial incorporation.

Relation to the literature. This paper contributes mainly to the growing literature in firm dynamics and macroeconomics that extends the extensive literature on cohort effects in labor market context to firms.[7] The analysis on the persistence of business cycles on markups developed in this chapter, in fact, complements a growing literature in macroeconomics that studies the persistent effects of aggregate fluctuations on firm sizes (Moreira [2015], Sedláček and Sterk [2017], Pugsley *et al.* [2019]). In this literature, the paper closest in spirit to this one is Moreira [2015]. In this study, she analyzes the effects of business cycles at the time of firms' inception on their average size. She

[6]This robustness check is closer in spirit to the labor literature that analyzes the effect of business cycles on workers' earnings at the time of their entry in the labor market or even at the time of their graduation.

[7]See, among others, Schwandt and Von Wachter [2019], Altonji *et al.* [2016] and Oreopoulos *et al.* [2012] that estimate the effects on earnings of entering the labor market during a recession.

finds that firms that start their operations during adverse aggregate conditions remain permanently smaller than firms that start during periods of high aggregate demand. Using a more structural approach than the one followed by Moreira [2015], Sedláček and Sterk [2017] documents a first-order role of business cycles in permanently affecting the size distribution of firms directly affecting the distribution of potential entrants.

Another growing literature that this paper contributes to is the one that links developments in empirical industrial organization, particularly for the estimation of markups at the firm-level, like De Loecker and Warzynski [2012], De Loecker and Eeckhout [2017], to the macroeconomic consequences of firm-level behavior. Notable examples in this literature are, among others, De Loecker *et al.* [2020], Autor *et al.* [2020] and Van Reenen [2018]. Contrary to this paper, however, at the core of these studies there is the analysis of long-run trends in aggregate markups that have spurred the debate on the rise in market power documented for the US economy and the ensuing macroeconomic implications. However, by exploring the interaction between markups and the aggregate state of the economy at a frequency higher than the one currently dominating this literature, it is possible to uncover new channels for business cycles in determining firm-level behavior and the fluctuations of our economies. Consequently, I consider the analysis developed in this chapter as a complement to this growing literature focused on the long-run trends in firm-level and aggregate markups.

The inherent difficulties in estimating markups have spurred a lively debate in this literature with notable contributions from Traina [2018], Syverson [2019a,b] and Bond *et al.* [2020], among others. In particular, Traina [2018] and Syverson [2019a,b] have shown how the estimates of the long-run trends in markups are sensible to the weighting used in the aggregation (costs versus sales) and on the balance sheet items used to proxy firms' variable costs. For the empirical exercise at the core of this chapter, however, the choice of the proxy variable does not affect the main results of the paper, despite having a large effect on the long-run trends. Bond *et al.* [2020], instead, highlight some deep issues relative to the identification of firm-level markups using the ratio between the output elasticity of a variable input and that variable input cost share in firms' revenues, which is the baseline estimator of markups at the firm-level. They show that, for this estimator, the use of revenue data instead of quantities induces a first-order issue in identifying markups that at best delivers a biased estimate of markups. In the most severe case, this implies that the only information content about markups is contained in the bias rather than the estimator. I address the concerns raised by their work as much as possible in the discussion about the estimation of the production function and throughout the paper. Given data availability, however, I often have to rely on second-best solutions to the issues raised in their study.

Structure of the paper. The paper is structured as follows. Section 1.2 describes the procedure used to estimate markups at the firm level and discusses its limitations while Section 1.3 describes the empirical methodology used to compute the correlation between aggregate conditions at listing and markups. Section 1.4 briefly describes the data used for the analysis and Section 1.5 presents the main results. Section 1.6, instead, reports various robustness checks on markup measurements. Section 1.7 finally concludes.

1.2 Measuring markups at the firm-level

The estimation strategy I use in this paper exploits firms' optimal behavior to back-out an estimate of markups at the firm level without the need to specify an explicit demand system. This method has been developed and popularized by De Loecker *et al.* [2020] and is based on the production function approach pioneered by Hall [1986, 1988] on industry-level data.

Consider a firm j that produces using the following technology, $Q_{j,t} = Q(\mathbf{X}_{j,t}, K_{j,t}, \omega_{j,t})$, where $\mathbf{X}$ is a vector of variable inputs, ω is firm specific productivity and K are predetermined inputs. The cost minimization problem for each producer therefore is the following:

$$\min_{\mathbf{X}_{j,t}, K_{i,t}} \left\{ \mathbf{X}'_{j,t}\mathbf{P}_{j,t} + R_{j,t}K_{j,t} + \lambda_{j,t}(Q_{j,t} - Q(\cdot)) \right\}.$$

The first-order condition, for a generic variable input $X^{\nu} \in \boldsymbol{X}$, is

$$\frac{\partial \mathcal{L}(\cdot)}{\partial X^{\nu}_{j,t}} = P^{\nu}_{j,t} - \lambda_{j,t}\frac{\partial Q(\cdot)}{\partial X^{\nu}_{j,t}} = 0, \tag{1.1}$$

where $\lambda_{j,t}$ can be interpreted as the marginal cost of producing at a given level of output. Equation (1.1) can be rearranged as

$$\frac{\partial Q(\cdot)}{\partial X^{\nu}_{j,t}} \frac{X^{\nu}_{j,t}}{Q_{j,t}} = \frac{1}{\lambda_{j,t}} \frac{P^{\nu}_{j,t} X^{\nu}_{j,t}}{Q_{j,t}}. \tag{1.2}$$

Defining the markup as price over marginal costs, $\mu_{j,t} \equiv \frac{P_{j,t}}{\lambda_{j,t}}$, it is possible to rewrite equation (1.2) so that

$$\mu_{j,t} = \theta^{\nu}_{j,t} \frac{P_{j,t} Q_{j,t}}{P^{\nu}_{j,t} X^{\nu}_{j,t}}, \tag{1.3}$$

where $\theta^{\nu}_{j,t}$ is the elasticity of output with respect to the variable input X^{ν}.

Obtaining consistent estimates of markups in this setting requires assuming that firms are free to adjust their prices without incurring any cost. The use of yearly data to estimate the production function mitigates the potential drawbacks of abstracting from price rigidities at the firm level.[8]

1.2.1 Estimation of production function elasticities

The measurement of markups developed by Decker *et al.* [2017] relies on the estimation of the output elasticity to variable inputs.

For the baseline results presented in the paper, I estimate the following revenue Translog production function for each two-digits NAICS code:

$$y_{j,t} = \theta_k k_{j,t} + \theta_v v_{j,t} + \theta_{kk} k^2_{j,t} + \theta_{vv} v^2_{j,t} + \theta_{kv} k_{j,t} v_{j,t} + \omega_{j,t} + u_{j,t}, \tag{1.4}$$

where $y_{j,t}$ are firm j's log-revenue at time t; $k_{j,t}$ the logarithm of its capital stock and $v_{j,t}$ the log-value of a bundle of variable inputs.

As usual, the main identification challenge in production function estimation is the simultaneity bias induced by the unobserved time-varying firm-level productivity, $\omega_{j,t}$.

I follow the proxy variable literature,[9] and in particular De Loecker and Eeckhout [2017], to estimate the production function in (1.4) using a two-step ap-

[8]The implicit estimation of markups as wedges between output elasticities of variable inputs and their expenditure shares arise naturally in any market structure different than perfect competition. However, these wedges can reflect other distortions if the structural assumption of no adjustment cost is violated.

[9]The proxy variable literature, pioneered by Pakes [1994] relies on adopting a control function to estimate the production function. An alternative approach is given by the literature on Dynamic Panel data, for more detais see Blundell and Bond [2000].

proach based on the use of a control function for the productivity process. The identification relies on the observation that the optimal choice of firms about a variable input follows a policy function like $v_{j,t} = v(k_{j,t}, \omega_{j,t})$. Then, providing that the policy function is invertible[10], the productivity process can be proxied by a control function so that $\omega_{j,t} = \omega(k_{j,t}, v_{j,t})$ with $\omega(\cdot) = v^{-1}(\cdot)$.

First step. In the first step I clean the output value from measurement errors and unanticipated productivity shocks using a second order polynomial of capital and variable inputs

$$y_{j,t} = P(k_{j,t}, v_{j,t}; \phi) + u_{j,t},$$

where $P(\cdot)$ is a composite function of the productivity and the unknown control function.[11]

Second step. Using the estimates of ϕ from the first step, I can construct a measure of productivity that does not depend on the measurement error $u_{j,t}$. That is,

$$\omega_{j,t} = P(k_{j,t}, v_{j,t}; \hat{\phi}) - [\theta_k k_{j,t} + \theta_v v_{j,t} + \theta_{kk} k_{j,t}^2 + \theta_{vv} v_{j,t}^2 + \theta_{kv} k_{j,t} v_{j,t}].$$

Then, exploiting the assumption that productivity follows an AR(1) process, is it is possible to construct a measure of productivity innovations, $\xi_{j,t}(\Theta)$, projecting $\omega_{j,t}$ on $\omega_{j,t-1}$. Under the assumption that firms react to unanticipated productivity shocks contemporaneously so that the lagged values of variable inputs can be used as valid instruments, the production function coefficients in Θ can be identified using the following moment conditions:

$$\mathbb{E}\left[\xi_{j,t}(\Theta)\begin{pmatrix} k_{j,t} \\ v_{j,t-1} \\ k_{j,t}^2 \\ v_{j,t-1}^2 \\ k_{j,t} v_{j,t-1} \end{pmatrix}\right] = 0. \tag{1.5}$$

[10] Pakes [1994] proves the invertibility of policy functions associated to a wide class of production functions.

[11] While it is possible to estimate the coefficient on variable inputs directly at this step, as noted by Ackerberg *et al.* [2015], it is more efficient to use the first stage only to clean the output variable from potential measurement errors and estimate all the production function coefficients in the second stage.

The output elasticity to variable inputs that is relevant for the measure of markups in equation (1.3) therefore is:

$$\theta^v_{j,t} = \theta_v + 2\theta_{vv} v_{j,t} + \theta_{kv} k_{j,t}. \tag{1.6}$$

Note that, with a Cobb-Douglas specification of the production function the relevant measure for the output elasticity to variable inputs, $\theta^v_{j,t}$, is given by equation (1.6) without the cross-products.

1.2.2 Discussion on limitations

Due to data availability I am not able to separate between prices and quantities for my measure of firm-level output. As also noted by Bond *et al.* [2020], this is problematic as variations in output and input prices could bias the estimates of the output elasticities and hence the measurement of markups.

To see how lack of price data can make the estimation problematic, consider the Cobb-Douglas version of (1.4). As discussed by De Loecker *et al.* [2020], the value of production can be expressed in logs as $y_{j,t} = q_{j,t} + p_{j,t}$ so that

$$q_{j,t} + p_{j,t} = \theta_v \tilde{v}_{j,t} + \theta_l \tilde{k}_{j,t} + p_{j,t} - \sum_{i \in \{k,v\}} \theta_i p^i_{j,t} + \omega_{j,t} + u_{j,t}$$

with $\tilde{x}_{j,t} = x_{j,t} + p^x_{j,t}$ being the deflated values of input x, and $p^x_{j,t}$ the input price paid by firm j. Note that if we consider the standard specification when coefficients do not change over time, then the bias caused by not observing prices would affect the level of markups, but not their time-series behavior.

In addition, there is the concern that firm-specific shocks could be reflected in input and output prices. This pass-through is not an issue in the unlikely case that variations in output prices are completely offset by variations in input prices. In more realistic settings, in which the pass-trough of shocks between input and output prices is not perfect, firms will be able to create a wedge between the input price bundle and the output price. Yet, assuming a constant returns to scale production function we can link the size of the bias induced by the incomplete pass-through exactly to marginal costs, $\lambda_{j,t} = \sum_{i \in \{k,v\}} \theta_i p^i_{j,t} - \omega_{j,t}$.

Given that the optimal pricing strategy implies a markup over marginal costs we can rewrite the price as, $p_{j,t} = \lambda_{j,t} + \mu_{j,t}$ and substitute it back in measurement equation for the production function to get

$$q_{j,t} + p_{j,t} = \theta_v \tilde{v}_{j,t} + \theta_l \tilde{k}_{j,t} + \mu_{j,t} + u_{j,t}.$$

Hence, as long as it possible to control for the markup that a firm is allowed to charge we can correctly estimate the output elasticities following the procedure outlined in the previous section. As markups are unobserved, a viable solution is to approximate them with a function whose arguments are relevant determinants of markups.

In my empirical application I follow De Loecker *et al.* [2020] and I include a sector-year linear function in firms' market shares and productivities as an approximation for markups. This is clearly a second-best solution to an important issue in the estimation of the production function and therefore of markups' levels. I verify that the Θ estimates are robust to various specification of the production function, but without more detailed data on output quantities, it is not possible to go beyond these second-best solutions to the problem.

1.3 Empirical methodology

In this section, I describe the empirical framework that I am relying on to estimate, at the firm-level, the magnitude and the persistence of aggregate conditions close to listing on markups.

1.3.1 A reduced-form approach: autocorrelation of cohort-level markups

Before describing how I adapt age-cohort-period models to estimate cohort effects on firm-level markups, it is useful to discuss two coarser approaches that, however, can be informative of the persistence of cohort-level factors in markups. The first of the two is based on measuring the correlation of markups across different cohorts with their initial levels. In fact, absent any cohort-specific component, markups at the cohort-level should not have a significant correlation with their initial levels. The second one, instead, looks at differences between cohort-level autocorrelations and aggregate autocorrelations as

Table 1.1: Correlation of cohort-level markups to their initial levels

Dep.Variable: $\log(\mu_{a,t})$	(1)	(2)	(3)
$\log(\mu_{0,t-a})$	0.366***	0.503***	0.403***
	(0.105)	(0.072)	(0.097)
$\log(\mu_{0,t-a})\times$ Age	0.071	-0.022*	-0.010
	(0.044)	(0.012)	(0.009)
R^2	0.63	0.34	0.14
N	250	370	340

Note: The table reports the elasticity of cohort-level markups with the markup charged in the first year a cohort is observed. Specifically I estimate equation (1.7). I report results for three different horizons, 5, 10 and 20 years. Column (1) reports the coefficient of interest for cohorts of firms from 1961 to 2011 followed for up to 5 years (50 cohorts followed for up to 5 years). Column (2) is based on firms starting from 1970 to 2006 followed up to 10 years of age (37 cohorts followed for 10 years). Column (3) follows cohorts from 1980 to 1996 up to 20 years (17 cohorts for 20 years).

symptoms of the existence of persistent factors that influence markups at the cohort level.

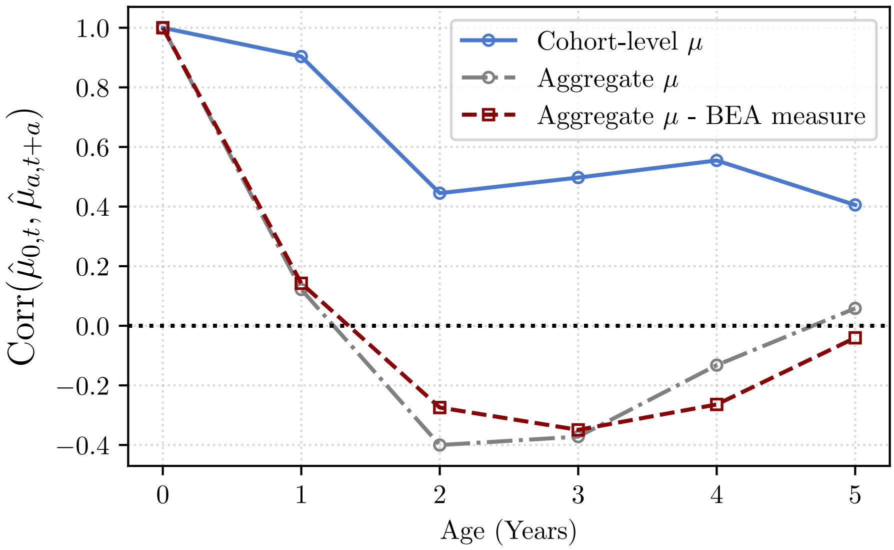
To estimate the correlation of cohort-level markups with their initial levels I am exploiting the following specification:

$$\log(\mu_{a,t}) = \alpha+\beta_0 \log(\mu_{0,t-a})+\beta_1 \log(\mu_{0,t-a})\times a+\beta_2 \log(\mu_{0,t})+\beta_3 a+\beta_4 a^2+u_{a,t}, \tag{1.7}$$

where $\mu_{a,t}$ is the average markup of cohort a in year t; $\mu_{0,t-a}$ is the average markup of cohort a in the year of birth; $\mu_{0,t}$ is the average markup of entering firms in year t; and a is age. The elasticity of each cohort markups to their initial conditions is therefore given by β_0 and the elasticity at each subsequent age is $\beta_0 + \beta_1 \times a$.

The comparison of the coefficients of interest for equation (1.7) estimated in the main cohort-level sample and is reported in Table 1.1. Even following cohorts for up to twenty years, as in column (3), the correlation with the initial level remains high and significant, indicating that there are cohort-level factors in average markups that do not vanish over time. A similar result can be shown by comparing the autocorrellation between aggregate markups and cohort-level ones. Figure 1.2, plots the autocorrelation of aggregate markups together with the correlation of the average markup by cohort with the average markup of the same cohort a years in the future. The cohort-level measure of markups exhibits an higher persistence than the aggregate ones, remaining higher than zero up to age five. This indicates that cyclical variations in markups across

Figure 1.2: Markup autocorrelation



Note: The figure plots the cohort-level and aggregate autocorrelation of markups in the model and in the data. *Cohort-level* refers to correlations of average markup by cohorts of firms with the average markups of the same cohort a years into the future, i.e. $Corr(\hat{\mu}_{0,t}, \hat{\mu}_{a,t+a})$, where $\hat{}$ indicates deviations from an Hodrick-Prescott trend taken *across* cohorts of the same age. *Aggregate* refers to autocorrelations of cost-weighted average markup in the economy between t and $t+a$, i.e. $Corr(\hat{\mu}_t, \hat{\mu}_{t+a})$ where $\hat{}$ indicates deviations from an Hodrick-Prescott trend. The BEA autocorrelation is constructed on a measure of aggregate markup that does not require the estimation of the production function, as in Kaplan and Zoch [2020]. Specifically, it is obtained from the ratio of the final producer price index (BEA code: `WPSFD49207`) to the intermediate producer price index (BEA code: `WPSID61`).

cohorts persist into later years without mean-reverting. This particular feature is not shared by the measure of aggregate markups that does not show any persistent autocorrelation beyond one year. Importantly, the lack of strong-autocorrelation in aggregate measures of markups is a characteristic of both the cost-weighted measure of aggregate markups based on Compustat data and an alternative one constructed following Kaplan and Zoch [2020] and based on the Final and Intermediate Demand PPI indexes published by the Bureau of Economic Analysis. The advantage of the latter is that does not require the estimation of the production function and therefore does not suffer from the limitations discussed in the previous section, providing a useful robustness check on the measure obtained from aggregating firm-level estimates.

1.3.2 Measuring cohort effects for firm-level markups

In order to estimate the effects of belonging to a particular cohort, c, on an individual outcome, $y_{j,c,t}$, ideally we would want to control for both the effect of aging and for the effect of contemporaneous aggregate conditions. To avoid imposing too much structure on these effects, a natural specification could be the following:

$$y_{j,c,t} = \alpha + \phi_a + \phi_t + \phi_c + u_{j,c,t},$$

where ϕ_a, ϕ_t and ϕ_c are respectively age, time and cohort fixed effects.

Unfortunately this model suffers from a well-known identification problem as the set of age, time and cohort fixed effects are perfectly collinear. As noted by Heckman and Robb [1985], a simple fix to this problem is to proxy one of the fixed effects with a variable that is not collinear with the remaining two.[12]

I follow the empirical specification proposed by Moreira [2015] and, for each firm, I proxy cohort effects with a measure of aggregate conditions in the first year of available accounting data.[13] Therefore, the main firm-level specification that I bring to the data to estimate the correlations of aggregate conditions on firm-level markups is the following:

$$\log(\mu_{j,c,t}) = \alpha + \phi_a + \phi_t + \beta Y_c + \mathbf{X}_{j,t}\gamma + u_{j,c,t}, \tag{1.8}$$

where $\mu_{j,c,t}$ is the markup charged by firm j belonging to cohort c at time t; ϕ_a and ϕ_t are respectively age and time fixed effects; Y_c is a measure of initial aggregate conditions for firms belonging to cohort c and $\mathbf{X}_{j,t}$ is a vector of controls that includes sector fixed effects and a second-order polynomial in firm j's sale share in her three digits sector.

The main coefficient of interest is β, that, under the standard exogeneity restrictions, pins down the percent change in average markups resulting from a one-percent variation in the initial business cycle conditions, after controlling for aggregate conditions faced by firms throughout their lives and the aging process. The age fixed effects, instead, are estimates of an age profile of

[12]The estimation of cohort effects has been mostly exploited in the labor literature to quantify, for example, the effects of recessions on workers career paths. On this, see among others Oreopoulos *et al.* [2012] or, more recently, Altonji *et al.* [2016].

[13]In Compustat firms report their balance sheets also a few years before the first trading date. I use this date as the main measure of cohort effects for two reasons: the first one is that this is the time in which companies are committing to a new business model and are starting the process for the IPO; the second one is that this variable is much more populated than the IPO date. For the companies that report the IPO date (less than 40% of firms in the sample), the average time between the first available year and the IPO is 1.97 years while the median is 1 year. Restricting the sample to firms that have the first year of available accounting data at exactly the IPO year does not significantly affect the results. Appendix 1.B reports some additional evidences on the relationships between founding dates, first available years and IPO dates.

markups that reflects only the dynamics that can be attributed to the aging process, controlling for cohort and time effects.

1.3.3 Persistence of cohort effects for firm-level markups

The specification in equation (1.8) allows one to measure the average effect on yearly markups but does not allow to verify if the effects of business cycles are persistent or tend to vanish as firms age. To assess the persistence of aggregate conditions around the time of listing and their dynamics over firms' life cycles, I rely on the following specification:

$$\log(\mu_{j,c,t}) = \alpha + \phi_a + \phi_t + \beta_0 Y_c + \beta_1 Y_c \times a_{j,t} + \mathbf{X}_{j,t}\boldsymbol{\gamma} + \nu_{j,c,t}. \tag{1.9}$$

The coefficient of interest in this case are β_0, that captures the effect on aggregate conditions in the first year, and β_1, that estimates the supplementary effect of business cycle realizations for each additional year firm j is observed.

The specification in equation (1.9) assumes a monotonic effect of age on the effects of aggregate conditions at birth. As a robustness check, I also consider a less demanding specification that allows for a non-monotonic effect of age by interacting the realizations of aggregate GDP with a set of dummy variables for each age.

1.4 Data overview

The main data source for the analysis performed in the paper is the annual segment of Compustat. In this section, I discuss briefly the strengths and limitations of the data and I provide more details on the data cleaning process and the construction of the sample in Appendix 1.A.

The dataset includes detailed balance sheet information on US listed firms. It is widely used in the macro-finance literature but it has been recently used by De Loecker *et al.* [2020] to estimate the long-run trends in markups and to study the macroeconomic effects of market power. For the study performed in this paper, the data present three main limitations: i) the difficulty of recovering a consistent measure of variable costs from *Cost of Goods Sold* (COGS) due to some freedom in reporting standards for the firms included in the dataset;[14], ii) the fact that it is impossible to distinguish quantity and prices and iii) the selection issues that arise from using only listed firms to measure the effects of aggregate conditions on firm-level markups.

To address the concerns linked to the possible mis-measurement of variable costs I estimate markups using both measures of marginal costs debated in the literature, *Cost of Goods Sold* and *Operating Expenditure*. The long-run trends are significantly affected by the choice of this variable[15], however, this choice does not affect the qualitative results of the cohort effects analyzed in this paper.

The fact that the main sample of the analysis is based on listed firms makes it harder to pin down the possible mechanisms that are inducing these permanent effects, as the estimated effect on markups can be confounded by other firm choices that are directly linked to the decision of listing rather than their markup management. As mentioned, given the data and the empirical setting it is not possible to distinguish between two alternative mechanisms: either that business cycles close to the time of listing induce a firms' decisions that result in different estimated markups caused by *ex-post* different product market choices, or, equally likely, that business cycles select *ex-ante* different types of firms. Nonetheless, this dataset, unlike other firm-level databases, allows to follow a large number of cohorts for many years giving the possibility to use the full time-series variation in the proxy variables to approximate cohort effects and allowing to trace longer age profiles for firm-level markups that are useful in the calibration of the model developed in chapter 2. Therefore, either if listing is a structural choice that permanently changes the way a firm conducts its operations or is revealing of its product type, the combination of these two advantages dampens the concerns that stem from sample selection.

1.5 Main results

In this section I discuss the main results relative to the estimation of the age profiles and the cohort-effects on firm-level markups.

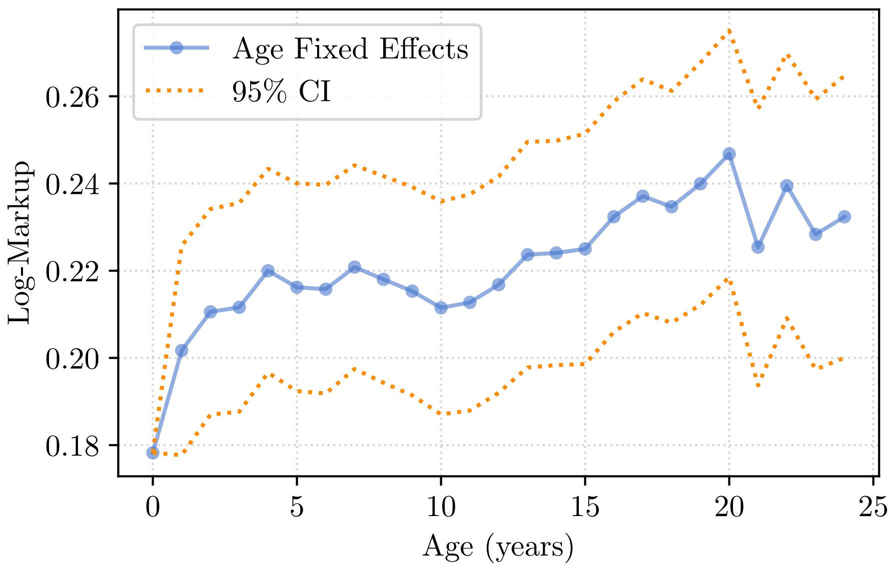
[14]See Traina [2018] for an analysis of the long-run trends in markups and how these are affected by using *Operating Expenditure* as a measure for variable cost measure.

[15]See Appendix 1.B for aggregate measures of markups under different specifications.

1.5.1 Age profile of markups.

The age fixed effects estimated in equation (1.8) show the dynamics of markups along firms' life cycles that can be attributed only to aging, taking into account aggregate conditions and cohort effects. Figure 1.3 plots the fixed effect coefficients together with a 95% confidence interval for the first 20 years of firms' lives.[16] The figure shows how the average markup steadily increases up to 15 years after the firsts available data. Using a structural estimation, Argente *et al.* [2018], estimate the price and marginal cost profiles at the product level and they show that firms increase markups on their products over time and constantly introduce new products that command an higher share of overall firm revenues. In their structural estimation markups increase, almost linearly at the product level for the first 16 quarters from their introduction with newer products having markups close to 0 at the time of their introduction.

Figure 1.3: Markups' age profile



Note: The figure plots the estimated age profile for markups from equation (1.8) together with the 95% confidence interval. Specifically, at each age a, I am plotting $\hat{\mu}_0 + \hat{\phi}_a$, where $\hat{\mu}_0$ is the average log-markup in the first available year and $\hat{\phi}_a$ are the estimated age fixed effects.

1.5.2 Cohort-effects

The main correlations between aggregate conditions and firm-level markups are reported in Table 1.2. The main result is that the correlation between

[16] As the estimates of the age fixed effects stabilize after approximately 20 years but the number of firms decline substantially do to attrition, the point estimates after 20 years are more volatile but do not add any additional information to the estimates.

markups and aggregate conditions at the time of the first available balance sheet data are significantly negative, indicating that firms that start their listing process in periods of below-trend output tend to charge higher markups on average.

Considering a two-standard deviation negative change in the cyclical component of log-real GDP, approximately -6%, the average effect reported in Column (1) implies that firms charge a markup $-0.06 \cdot -0.249 \approx 1.4\%$ higher every year, due to the aggregate conditions they faced when starting the listing process. To put it in perspective, this effect implies that firms that were close to listing at the height of the Great Recession, in 2009, charged an average markup approximately 2% higher than similar firms that were close to listing in 2007.

Trying to decompose this effect to control for its persistence, Columns (2) and (3) report the coefficients for two versions of equation (1.9): i) a parametric one where I restrict the contribution of age on the estimated cohort effect to follow a linear process, and ii) a semi-parametric one in which I estimate the effect of the initial aggregate conditions for each age group by means of fixed effects. In both cases, the coefficients reveal that the relevance of aggregate conditions is stronger in the initial years of firms' lives and then progressively vanishing within approximately ten to fifteen years.[17]

When the age effect is restricted to be linear, as in Column (2), the effects on markups in the initial years, is larger in magnitude than the average effect. A two-standard deviations drop of the cycle component of GDP, implies a markup in the initial year close to 4% higher than a similar firm that gets listed when GDP is on trend. Similarly, this implies that, for firms that were close to listing in 2009, the Great Recession is associated with a 5.1% increase in the first-year markup compared to similar firms that were close to listing in 2007.

Column (3), instead, allows more freedom in the measurement of the persistence of aggregate markups and reveals that the markup in the initial year of firm operations is approximately 5% higher for a firm experiencing a two-standard deviation negative change in the cycle component of GDP and still

[17]This result is at odds with what the firm dynamics literature has found regarding the effects on business cycles on firm sizes that tends to be permanent, see Moreira [2015], Sedláček and Sterk [2017] and more in line to the labor literature documenting the scarring effects of recessions.

significantly higher five years into firms' lives.[18] The age profiles estimated from equation (1.9) are plotted in Figure 1.1. The figure shows how for firms that experience positive aggregate conditions, defined as periods when the cyclical component of GDP is two-standard deviations above trend, close to their listing date tend to charge an initial markup that is approximately 8% lower that similar firms that instead face a negative realization of the business cycle.

In addition, besides the magnitude of the initial effect, it is worth noticing that firms that are first observed during booms exhibit a steeper age profile of markups. This particular shape of the age profile is not only a byproduct of the monotonicity of the age effect imposed by equation (1.9). Figure 1.4 shows a very similar pattern for the age profiles of firms even when the effects of initial aggregate conditions at each age are estimated using a more flexible specification than the one discussed in equation (1.9), where the effects of initial aggregate conditions at each are estimated using the interaction between the business cycle measure and dummy variables for each age.

As robustness checks, I estimate the cohort effects using also different proxy variables for aggregate conditions. The results are reported in Table 1.3. The negative effect of aggregate conditions on markups is robust for most of the proxy variables considered. The only one that completely fails is the simple indicator of recessions. This results should not be particularly worrying as the recession indicator is a poor proxy variable for aggregate conditions as it pools firms of very different recessions together. In addition, as I use yearly data, the variation in the recession indicator is very low as it is assigns a disproportionate amount of firms to booms.

As the main sample of analysis is based on firms that are already mature, I include fixed effects for deciles of initial firm sizes and deciles of initial sales to account for differences on markups that could affect markups through sizes and not only aggregate conditions close to the time of listing.

Table 1.4 reports the coefficients of interest for the baseline specifications in equation (1.8) and (1.9) with either initial size or initial sales controls. As expected, not including controls for firms sizes the average effect becomes insignificant indicating that some of the effects of business cycles on markups

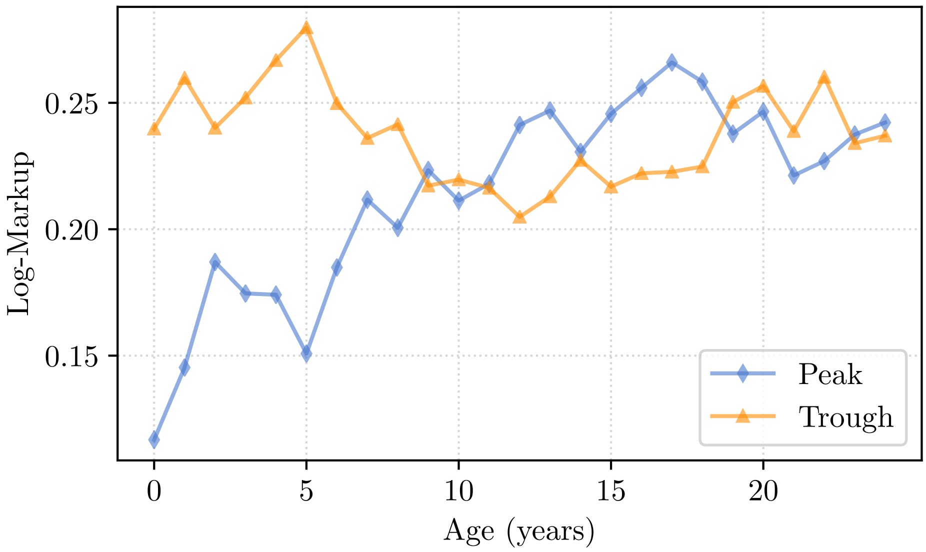
[18] It is important to note that the coefficients in this specification are not monotone and are significant only up to age five with the exception of the coefficient on age two.

Table 1.2: Cohort effects on firm-level markups

Dep. Variable: Log-Markup	(1)	(2)	(3)
Cycle measure	-0.249***	-0.672***	
	(0.064)	(0.135)	
Cycle measure × Age		0.043***	
		(0.010)	
Cycle measure × Age_0			-0.908**
			(0.383)
Cycle measure × Age_1			-0.856**
			(0.346)
Cycle measure × Age_2			-0.402
			(0.293)
Cycle measure × Age_3			-0.574*
			(0.312)
Cycle measure × Age_4			-0.677**
			(0.302)
Cycle measure × Age_5			-0.937***
			(0.322)
Cycle measure × Age_{6-10}			-0.207
			(0.133)
Cycle measure × Age_{11-15}			0.136
			(0.137)
Cycle measure × Age_{16-20}			0.149
			(0.137)
Cycle measure × Age_{21-25}			-0.051
			(0.140)
Age FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Initial Size Decile FE	Yes	Yes	Yes
Initial Sale Decile FE	Yes	Yes	Yes
R^2	0.14	0.14	0.14
N	91,317	91,317	91,317

Notes: Robust standard errors in parentheses, p :$^{***} < 0.01,^{**} < 0.05,^{*} < 0.1$. The table reports the main estimates of the elasticities between firm-level markups and aggregate conditions at the time the firm is first observed. Cohort effects are proxied using quadratically detrended log real GDP. Columns (1) and (2) report estimates using the specification in equation (1.8) and (1.9). Column (3), instead, shows estimates of cohort effects persistence using age-group fixed effects. In this case, the reported interaction coefficients estimate the effects of initial conditions for the relative age group. The measure of business cycle is quadratic detrended real log-GDP.

Figure 1.4: Cohort effects on markups' age profile



Note: The figure plots the persistence of initial aggregate conditions on the age profile of markups using a version of equation (1.9) that allows for non monotonic effects of age. Specifically, at each age, I am plotting $\hat{\mu}_0 + \hat{\phi}_a + \hat{\psi}_a Z$ where $\hat{\mu}_0$ is the average log-markup in the first available year, $\hat{\phi}_a$ are the estimated age fixed effects and the $\hat{\psi}_a$ are the estimates of the interaction terms of age dummies and the measure of business cycle in the first available year. Each $\hat{\psi}_a$, then, estimates the average effect of initial aggregate conditions a years into firms' lives. Z is a measure of business cycle outcomes that takes value 2σ for *Peak* and -2σ for *Trough*, with σ being the standard deviation of quadratically detrended real GDP.

are mediated from an effect on size. Decomposing the effect of aggregate condition close to the time of listing along the age profile of firms, in both cases we recover the baseline result that these effects are stronger at the beginning of firms' listed lives, albeit of a smaller magnitude compared with the baseline specification highlighting once again the importance of controlling for initial firm sizes.

Table 1.3: Cohort effects estimates with different proxy variables

	HP Filtered GDP		Hamilton-Filtered GDP		Unemployment rate		Recession Indicator	
Dep.Variable: Log-Markup	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cycle Measure	-0.092	-0.480**	0.015	-0.292***	0.006***	0.008***	-0.022***	0.011
	(0.108)	(0.195)	(0.049)	(0.092)	(0.001)	(0.003)	(0.004)	(0.008)
Cycle Measure × Age		0.040***		0.032***		-0.0002		-0.003***
		(0.015)		(0.007)		(0.0002)		(0.001)
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial Size Decile FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial Sale Decile FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
N	91,317	91,317	91,317	91,317	91,317	91,317	91,317	91,317

Notes: Robust standard errors in parentheses, p :$^{***} < 0.01,^{**} < 0.05,^{*} < 0.1$. The table reports estimates from equation (1.8) and (1.9) using different proxy variables for initial aggregate conditions. In particular, I report the main coefficient of interest for Hodrick-Prescott filtered log real GDP (smoothing equal to 6.25); Hamilton filtered log real GDP (one lag and two leads); the unemployment rate (in percent) and a dummy indicator for when the first available year in the data is indicated as an NBER recession.

Table 1.4: Cohort effects with either size or sale controls

Dep.Variable: Log-Markup	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cycle measure	-0.206***	-0.653***		0.002	-0.275**		0.083	-0.275**	
	(0.065)	(0.137)		(0.055)	(0.109)		(0.055)	(0.112)	
Cycle measure × Age		0.045***			0.028***			0.037***	
		(0.010)			(0.008)			(0.008)	
Cycle measure × Age_0			-0.882**			-0.592**			-0.531*
			(0.393)			(0.284)			(0.297)
Cycle measure × Age_1			-0.827**			-0.399			-0.345
			(0.353)			(0.266)			(0.275)
Cycle measure × Age_2			-0.375			-0.066			-0.013
			(0.298)			(0.242)			(0.249)
Cycle measure × Age_3			-0.549*			-0.275			-0.231
			(0.315)			(0.262)			(0.267)
Cycle measure × Age_4			-0.679**			-0.184			-0.147
			(0.303)			(0.257)			(0.262)
Cycle measure × Age_5			-0.945***			-0.180			-0.189
			(0.324)			(0.266)			(0.270)
Cycle measure × Age_{6-10}			-0.180			-0.077			-0.088
			(0.134)			(0.115)			(0.116)
Cycle measure × Age_{11-15}			0.224			0.360***			0.484***
			(0.137)			(0.114)			(0.115)
Cycle measure × Age_{16-20}			0.220			0.437***			0.642***
			(0.136)			(0.123)			(0.122)
Cycle measure × Age_{21-25}			-0.009			-0.086			0.084
			(0.140)			(0.125)			(0.126)
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial Size Decile FE	Yes	Yes	Yes	No	No	No	No	No	No
Initial Sale Decile FE	No	No	No	Yes	Yes	Yes	No	No	No
R^2	0.12	0.13	0.13	0.12	0.12	0.12	0.10	0.10	0.10
N	91,317	91,317	91,317	153,255	153,255	153,255	153,255	153,255	153,255

Notes: Robust standard errors in parentheses, p :$^{***}< 0.01,^{**} < 0.05,^{*} < 0.1$. The table reports the main estimates of the elasticities between firm-level markups and aggregate conditions at the time the firm is first observed. Cohort effects are proxied using quadratically detrended log real GDP. Columns (1-2), (4-5) and (7-8) report estimates using the specification in equation (1.8) and (1.9) with either initial size or initial sale controls. Columns (3), (6) and (9), instead, report the result of the same exercise for different age groups. Increasing the number of bins to control for initial firm sizes and sales does not affect the qualitative results reported in this table. The measure of business cycle is quadratic detrended real log-GDP.

1.6 Robustness checks

The quality of the data, the difficulty in measuring firm age and the intrinsic difficulties in the estimation of markups at the firm level are all possible causes of concern for the results discussed so far. In this section, I analyze a battery of robustness checks that address the most important issues affecting the measurement of markups and firm ages.

1.6.1 Alternative production function: Cobb-Douglas

The ratio estimator of firm-level markups in equation (1.3) is highly dependent on correctly estimating the elasticity of firm-level output to variable inputs. An incorrect measure of this elasticity would introduce a bias in the scale of markups. While this bias may not be too problematic for the analysis of changes in markups and their long-run trends, this parameter is particularly relevant in the context of the exercise developed in this paper as I estimate the level effect of business cycles on the average markup charged by different cohorts of firms.

Besides the common issues related to the identification of production functions' parameters, the estimation of a production function in a long panel, like the one I use in this paper, poses an additional conceptual problem linked to the time-frame used for the estimation of the technology adopted by firms. Pooling all years together and estimating the production function by sector implicitly assumes that the technology available to firms at the beginning of the sample is exactly the same as the one available to more modern enterprises. This is obviously an unrealistic strong assumption that would have relevant consequences for the measurement of markups if not only the technology, but also the output elasticities with respect to inputs were constant over time.

The choice of the Translog as the baseline specification for the production function is in fact dictated by a compromise between stability of the estimation across sectors and allowing a time-varying elasticity of output to variable inputs. In fact, even if the baseline production function is constant over time, the relevant elasticity for the measurement of markups is allowed to vary, as shown in equation (1.6). Nonetheless, the estimated production function is still assumed constant throughout the sample.

Therefore, in order to to address this issue thoroughly, in this subsection I describe the effects of changing the assumption on the nature of the technology on the main results of the paper. In particular, I re-estimate output elasticities for a Cobb-Douglas production function in two cases, one in which the technology available to firms is allowed to vary across sectors and time - *Sector-time varying elasticity* - and another one, more basic, where the technology is allowed to vary only across sectors - *Constant elasticity.*[19]

Sector-time varying elasticity. To allow for a sector-time varying technology, I estimate a revenue production function such as:

$$y_{j,t} = \theta_v v_{j,t} + \theta_k k_{j,t} + \omega_{j,t} + u_{j,t}$$

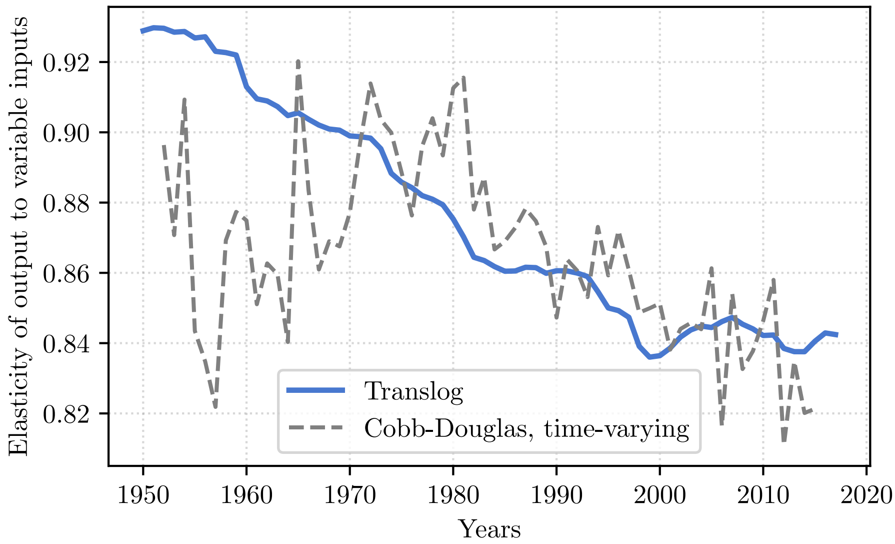
for each two-digits NAICS on a five-year rolling window. This allows to construct a series of sector-time varying elasticities to variable inputs $\{\hat{\theta}_{s,t}\}$ from 1952 to 2015.

Constant elasticity. As a useful benchmark to compare the results to, I estimate the Cobb-Douglas production function across two-digits sectors, pooling all years.

Figure 1.5 plots the sector average of this estimated elasticities together with the average elasticity estimated using the baseline Translog production function in equation (1.6). The figure shows a significant time volatility of output elasticities and a downward trend in both measures of θ_v, especially from 1980. Both estimates show a significant change in the output elasticity over time. The Cobb-Douglas estimate moves from approximately an average of 0.88 between 1950 and 1970 to 0.83 after 2000. A similar decline is captured in the evolution of the elasticity to variable inputs derived from the Translog production function. Therefore, even if the baseline estimation using the Translog production function is assuming a constant technology over time, the time series behavior of the implied elasticity to variable inputs is in line with a more flexible specification of the production function that takes into account the time-varying nature of the technology available to firms.

[19]While the former is a more coherent robustness check as it allows for a time-varying elasticity, I find it useful to report also the constant elasticity case as it provides a useful benchmark, also against other estimates in the literature, for the estimation of the production function.

Figure 1.5: Sector-time varying output elasticities to variable inputs



Note: The figure plots the averages across 2 digits sectors of the elasticities of output to variable inputs estimated using a Translog and a Cobb-Douglas production function. The Cobb-Douglas production function is estimated separately for each two-digits NAICS on a five year rolling window to allow changes in the estimated parameters from 1952 to 2015. The elasticity for the Translog function, instead, is estimated for each two-digits sector pooling all years from 1950 to 2017 and then constructed averaging the firm-level elasticities constructed following equation (1.6).

Table 1.5 reports the estimates of the cohort effects for markups measured using both the constant elasticity and the time-varying specification. The estimated effect of business cycles on markups is larger when time variation in technology are allowed in the estimation. In particular, allowing for time variation in θ_v implies an effect of business cycles on markups 25% higher (the coefficient moves from -0.302 to -0.385). For the time-varying specification, this implies that firms starting the listing process when aggregate output is two standard deviations below trend report markups that are approximately 2.6% higher than similar firms that instead start the process with aggregate output being on trend. When I use the estimate of the production function with constant elasticity, the same change in output would result in a 2% difference in markups, which is close to the baseline effect estimated using the Translog specification. These estimates mean that firms listed at the height of the Great Recession, in 2009, were able to charge a markup respectively 3% and 2.3% higher than if they decided to go public in 2007. The estimated persistence, instead, is very similar between the two specifications and the baseline estimates.

Table 1.5: Cohort effects with Cobb-Douglas production function

	Sector varying θ_v			Sector-Time varying θ_v		
Dep. Variable: Log-Markup	(1)	(2)	(3)	(4)	(5)	(6)
Cycle measure	-0.302***	-0.749***		-0.385***	-0.910***	
	(0.065)	(0.137)		(0.065)	(0.137)	
Cycle measure × Age		0.045***			0.053***	
		(0.010)			(0.010)	
Cycle measure × Age_0			-1.038***			-1.194***
			(0.384)			(0.389)
Cycle measure × Age_1			-0.829**			-0.928***
			(0.342)			(0.346)
Cycle measure × Age_2			-0.489			-0.683**
			(0.301)			(0.299)
Cycle measure × Age_3			-0.656**			-0.941***
			(0.326)			(0.321)
Cycle measure × Age_4			-0.798**			-0.901***
			(0.326)			(0.326)
Cycle measure × Age_5			-1.067***			-1.166***
			(0.335)			(0.339)
Cycle measure × Age_{6-10}			-0.263*			-0.311**
			(0.137)			(0.136)
Cycle measure × Age_{11-15}			0.111			0.085
			(0.136)			(0.135)
Cycle measure × Age_{16-20}			0.134			-0.001
			(0.134)			(0.136)
Cycle measure × Age_{21-25}			-0.100			-0.023
			(0.138)			(0.139)
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Initial Size Decile FE	Yes	Yes	Yes	Yes	Yes	Yes
Initial Sale Decile FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.08	0.08	0.08	0.08	0.08	0.08
N	92,336	92,336	92,336	90,458	90,458	90,458

Note: Robust standard errors in parentheses, p :$^{***} < 0.01,^{**} < 0.05,^{*} < 0.1$. The table reports the coefficient of interest for the specifications in equation (1.8) and (1.9) plus age group specific age effects. Markups are measured using the elasticity of output to variable inputs, θ_v, estimated from a Cobb-Douglas production function in two ways: i) pooling all years in the sample - Columns (1) to (3); and ii) estimating the production function coefficients on five-year rolling windows for each two-digits sector - Columns (4) to (6). The measure of business cycle is quadratic detrended real log-GDP.

Figure 1.6 shows the estimated cohort effects and the age profiles of markups for firms that are first observed in booms and in recessions for the two Cobb-Douglas estimations of the production function. The resulting age profiles are remarkably similar, indicating a second-order role for the output elasticity of variable inputs in the determination of the cohort effects for markups. Compared to the baseline profiles in Figure 1.1, both Cobb-Douglas specifications deliver a higher effect of business cycles on the age profiles. A negative two-standard deviation of the cycle component of GDP from its trend results in a 12% higher initial markup when I consider the sector-time variation in output elasticities, as shown in Figure 1.6a. The estimates of the cohort effects with the constant Cobb-Douglas production function instead, shown in Figure 1.6b, deliver an initial markup approximately 9% higher for firms that experience the same change in the cycle component of GDP. This last result is closer in magnitude to the baseline results obtained with the Translog specification.

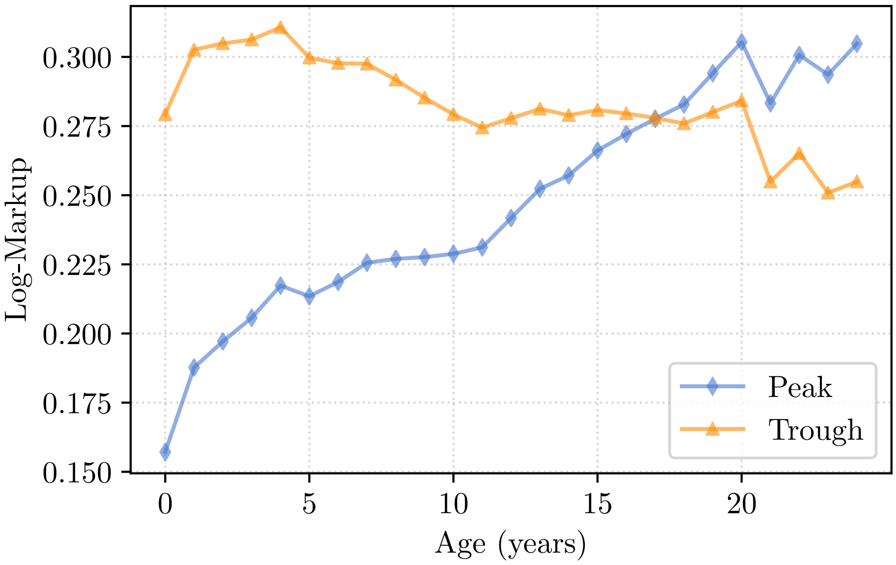
In terms of persistence, the cohort effects estimated with both Cobb-Douglas production functions and the ones estimated using the Translog show the same pattern. Firms that are listed during bad economic times tend to have higher markups for approximately 15 years of their lives as listed firms.

It is important to note that, despite delivering very similar average elasticities, the time varying estimation of the Cobb-Douglas production function, measures a larger effect of initial aggregate conditions on markups. The reason for this discrepancy can be due to the changing nature of technology for different cohorts of firms. The estimation of the production function for each sector across a relatively short horizon, allows to incorporate in the cohort effects on markups the consequences of business cycles on the technology used by firms that get listed at different moments of the cycle. These effects, that are tightly linked to the changing nature of the production process are not fully captured by the Translog specification despite featuring a firm specific elasticity of output to variable inputs.[20]

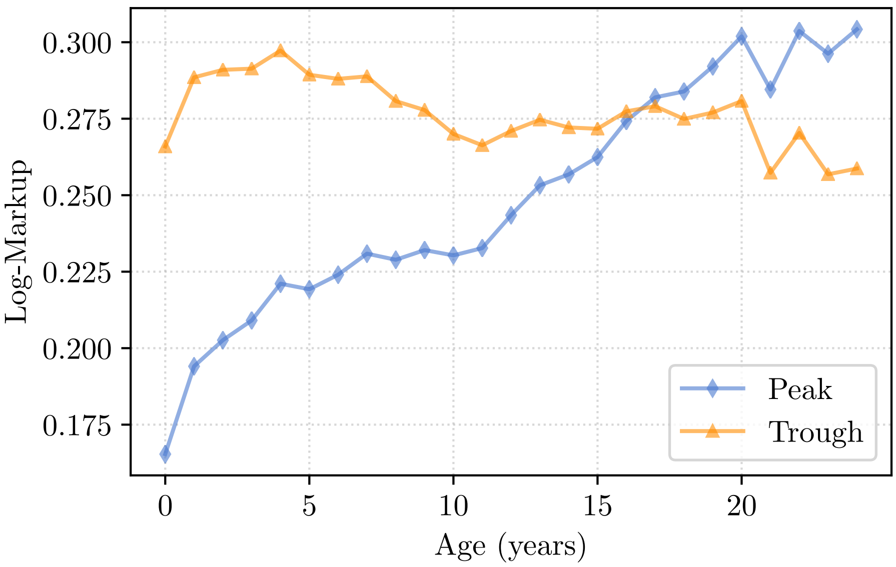
[20] Despite being a relevant aspect for the estimation of markups the investigation of different techniques to estimate production functions are beyond the scope of this project.

Figure 1.6: Cohort effects and age profiles for Cobb-Douglas production function

(a) Time-varying specification



(b) Constant elasticity



Note: The figure plots the age profile for markups estimated from equation (1.9) for markups measured using the output elasticity to variable inputs estimated from a Cobb-Douglas production function. Both the sector-time varying specification and the sector varying specification are reported. Specifically, at each age a, I am plotting $\hat{\mu}_0 + \hat{\phi}_a + (\hat{\beta}_0 + \hat{\beta}_1 a)Z$, where $\hat{\mu}_0$ is the average log-markup in the first available year, $\hat{\phi}_a$ are the estimated age fixed effects, $(\hat{\beta}_0, \hat{\beta}_1)$ capture the estimated persistence of aggregate conditions in the first available year. Z is a measure of business cycle outcomes that takes value 2σ for *Peak* and -2σ for *Trough*, with σ being the standard deviation of quadratically detrended real GDP. Table 1.5, Column (2) reports the coefficients used to construct these graphs.

1.6.2 Alternative measure of output elasticity to variable inputs: cost shares

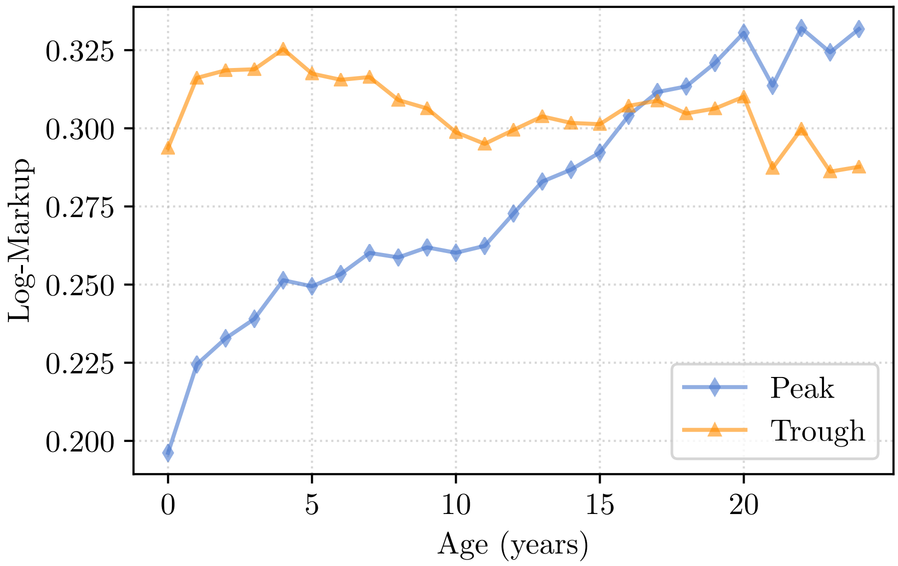
The measure of markups at the firm-level relies heavily on the estimation of the output elasticity to variable inputs. As the estimation of the production function without price and quantity data risks to be biased, I report the main empirical analysis of the paper for a measure of markups where the elasticity of output to variable inputs is not coming from a direct estimation of the production function but rather from the direct measure of the cost-share in variable inputs.[21] In order to compute total costs I assume a common user cost of capital of 12% that includes an exogenous depreciation rate, the federal fund rate and risk-premia, as in De Loecker *et al.* [2020] and Gutiérrez and Philippon [2019]. I then average the share of expenditure on variable inputs over total costs within each year and 2-digits NAICS to compute a new sector-time varying $\theta_{s,t}$ to measure markups following equation (1.3).

Also in this case, the qualitative effects of aggregate conditions at the time of first available balance sheet data are very similar to the baseline results shown in the main results of the paper. When markups are estimated using cost-shares, the effects, both on the average effect and its persistence, are very similar to the ones estimated using the baseline model, as shown in Table 1.6 and Figure 1.7.

As for the baseline results, even when the output elasticity is estimated with the cost-share of variable inputs a positive two-standard deviation in output (approximately 6%) translates to a 2% drop in the average markup charged by firms. Decomposing this effect along the age profile of firms shows that the effect on the initial markups is more than twice as large as the average one - a positive two-standard deviation change in output generates a 4.9% drop in the markup charged in the first year a firm is observed compared to a similar firm that instead does not face any change in the cyclical component of GDP. The effect of the initial business cycles, as in the baseline case, vanishes approximately after fifteen years of firms lives as public companies. Using the Great Recession as a benchmark, these estimates imply that firms starting the listing process in 2009 have an average markup 2.25% higher than if they went public in 2007. The estimated impact on markups in the first year as

[21] Assuming that the production function follows a Cobb-Douglas and markets for inputs are competitive then the cost-share in variable inputs is a valid measure for the output elasticity.

Figure 1.7: Cohort effects on markups' age profiles - Cost shares



Note: The figure plots the age profile for markups estimated from equation (1.9). Specifically, at each age a, I am plotting $\hat{\mu}_0 + \hat{\phi}_a + (\hat{\beta}_0 + \hat{\beta}_1 a)Z$, where $\hat{\mu}_0$ is the average log-markup in the first available year, $\hat{\phi}_a$ are the estimated age fixed effects, $(\hat{\beta}_0, \hat{\beta}_1)$ capture the estimated persistence of aggregate conditions in the first available year. Z is a measure of business cycle outcomes that takes value 2σ for *Peak* and -2σ for *Trough*, with σ being the standard deviation of quadratically detrended real GDP. The persistence coefficients underlying this plot are reported in Table 1.6, Column (2).

listed, instead, implies that markups are 5.6% higher for firms going public at the height of the Great Recession relative to similar firms going public in 2007.

1.6.3 Alternative measure of markups: Operating Expenditure

One of the main issues with Compustat data is linked to the fact that firms that follow different accounting standards are allowed to assign different expenditures to either *Cost of Goods Sold* (COGS) or to *Selling, General and Administrative Expenditures* (XSGA). As firms are allowed to choose which reporting standard to abide there is the risk that the variation in COGS does not reflect true differences in variable costs but rather differences in reporting standards.[22]

If the choice of the reporting standard was correlated with aggregate conditions at the time of first available data the estimate of the effect could be biased. To ensure that the negative effects of initial aggregate conditions are not due

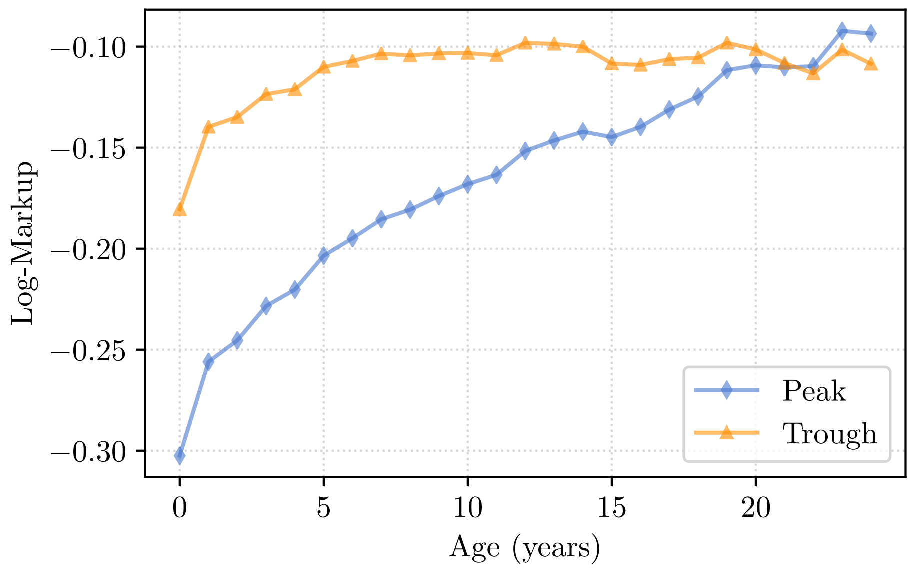
[22] See Traina [2018] and De Loecker *et al.* [2020] for an in-depth discussion of this issue.

Table 1.6: Cohort effects for markups measured using cost shares

Dep. Variable: Log-Markup	(1)	(2)	(3)
Cycle measure	-0.294***	-0.727***	
	(0.065)	(0.137)	
Cycle measure $\times$ Age		0.044***	
		(0.010)	
Cycle measure $\times$ Age$_0$			-1.041***
			(0.384)
Cycle measure $\times$ Age$_1$			-0.819**
			(0.342)
Cycle measure $\times$ Age$_2$			-0.453
			(0.301)
Cycle measure $\times$ Age$_3$			-0.626*
			(0.326)
Cycle measure $\times$ Age$_4$			-0.781**
			(0.327)
Cycle measure $\times$ Age$_5$			-1.054***
			(0.336)
Cycle measure $\times$ Age$_{6-10}$			-0.247*
			(0.137)
Cycle measure $\times$ Age$_{11-15}$			0.113
			(0.136)
Cycle measure $\times$ Age$_{16-20}$			0.145
			(0.134)
Cycle measure $\times$ Age$_{21-25}$			-0.125
			(0.138)
Age FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Initial Size Decile FE	Yes	Yes	Yes
Initial Sale Decile FE	Yes	Yes	Yes
R^2	0.10	0.10	0.10
N	92,336	92,336	92,336

Notes: Robust standard errors in parentheses, p :$^{***} < 0.01,^{**} < 0.05,^{*} < 0.1$. The table reports the coefficient of interest for the specifications in equation (1.8) and (1.9) plus age group specific age effects. Markups are measure using the elasticity of output to variable inputs measured using the cost share of variable inputs in total costs. Capital costs are included with gross user cost of capital of 12%. The measure of business cycle is quadratically detrended real log-GDP.

Figure 1.8: Cohort effects on markups' age profiles - Operating Expenditure



Note: The figure plots the age profile for markups estimated from equation (1.9). Specifically, at each age a, I am plotting $\hat{\mu}_0+\hat{\phi}_a+(\hat{\beta}_0+\hat{\beta}_1 a)Z$, where $\hat{\mu}_0$ is the average log-markup in the first available year, $\hat{\phi}_a$ are the estimated age fixed effects, $(\hat{\beta}_0, \hat{\beta}_1)$ capture the estimated persistence of aggregate conditions in the first available year. Z is a measure of business cycle outcomes that takes value 2σ for *Peak* and -2σ for *Trough*, with σ being the standard deviation of quadratically detrended real GDP. The persistence coefficients underlying this plot are reported in Table 1.7, Column (2).

from heterogeneity in reporting standards I re-estimate equation (1.8) using *Operating Expenditures* (OPEXP) as the main measure of variable costs for the construction of variable costs.[23] Traina [2018] shows that this choice has a profound impact on the long-run trend of aggregate markups, however, for the purpose of this paper the two measures deliver qualitatively very similar results, as shown in Table 1.7 and Figure 1.8.

In fact, when *Operating Expenditures* is used to measure the the sales to variable cost ratio in equation (1.3), the effects of business cycles increase in magnitude compared to the baseline estimates that uses *Cost of Goods Sold* as measure of variable costs. As XSGA includes the fixed costs a firm has to sustain in order to carry-on production, the higher business cycle effects on markups when this measure is considered could be indication of the fact that the selection operated by aggregate conditions reflects the whole structure of firms costs, affecting also the fixed component of its cost structure.

This implies that a positive two-standard deviation in output (approximately 6%) translates to approximately -3.2% change in the average markups charged

[23]In Compustat, *Operating Expenditures* = *Cost of Goods Sold* + *Selling, General and Administrative Expenditure*.

Table 1.7: Cohort effects from markups measured using *Operating Expenditures*

Dep. Variable: Log-Markup	(1)	(2)	(3)
Cycle measure	-0.487***	-0.909***	
	(0.051)	(0.111)	
Cycle measure × Age		0.043***	
		(0.008)	
Cycle measure × Age_0			-0.971***
			(0.342)
Cycle measure × Age_1			-1.056***
			(0.266)
Cycle measure × Age_2			-1.005***
			(0.264)
Cycle measure × Age_3			-1.080***
			(0.268)
Cycle measure × Age_4			-0.940***
			(0.265)
Cycle measure × Age_5			-0.677***
			(0.230)
Cycle measure × Age_{6-10}			-0.511***
			(0.107)
Cycle measure × Age_{11-15}			0.062
			(0.095)
Cycle measure × Age_{16-20}			-0.225*
			(0.116)
Cycle measure × Age_{21-25}			-0.306***
			(0.106)
Age FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Initial Size Decile FE	Yes	Yes	Yes
Initial Sale Decile FE	Yes	Yes	Yes
R^2	0.32	0.32	0.32
N	81,298	81,298	81,298

Notes: Robust standard errors in parentheses, p :$^{***} < 0.01,^{**} < 0.05,^{*} < 0.1$. The table reports the main coefficient of interest for the specifications in equation (1.8) and (1.9) plus age group specific age effects. Markups are constructed using *Operating Expenditures* instead of *Cost of Goods Sold* to mitigate misreporting of variable costs due to different accounting standards. The measure of business cycle is quadratically detrended real log-GDP.

by firms. Therefore, when measured with operating expenditures, the Great Recession translates to an average increase in markups of approximately 3.7% (firms that start listing in 2009 versus 2007). As for other robustness checks, decomposing the effect along the life cycle of firms highlights how the initial effects are larger (approximately -6.1% in the initial year for a positive two-standard deviation change in the cycle component of GDP) and slowly fading away in twelve to fifteen years.

1.6.4 Non-parametric cohort effects

An alternative solution to the identification problem of cohort effects in age-cohort-time models is to choose a normalization for either the cohort effects or the time effects.[24] I follow Moreira [2015] and normalize the cohort effects so that they sum to zero and are orthogonal to a time trend. The implicit identification assumption that this normalization choice implies is that long-run trends are fully captured by the combination of age and time fixed effects while cohort effects identify cyclical components.

Therefore, I estimate the effects of business cycle for cohorts of listing firms non-parametrically using the following specification at the firm-level:

$$\log(\mu_{j,t}) = \alpha + \phi_a + \phi_t + \sum_{c=1}^{C} \beta_c \mathbb{1}\{j \in c\} + \mathbf{X}_{j,t}\gamma + u_{j,t} \tag{1.10}$$

subject to

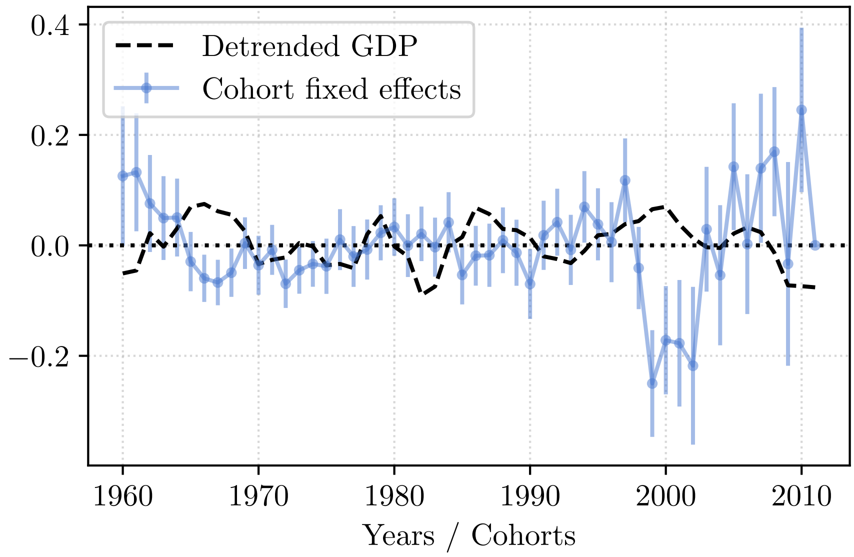
$$\sum_{c=1}^{C} \beta_c = 0,$$
$$\sum_{c=1}^{C} c\beta_c = 0.$$

where ϕ_a and ϕ_t are respectively age and time fixed effects; $\mathbb{1}\{j \in c\}$ is an indicator function that takes value one if firm j belongs to cohort c and $\mathbf{X}_{j,t}$ is a vector of firm-level controls that include size, a second order polynomial in firm j's sector shares and two-digits sectors fixed effects.

The constraints on equation (1.10) force the cohort-effects to sum to 0. The identification assumption behind this normalization choice is that any long-

[24] Deaton [1997] proposes this normalization in the context of consumers life-cycle problems. See Schulhofer-Wohl [2018] for a discussion of the *time* versus *cohort* normalization in life-cycle models.

Figure 1.9: Markups and non-parametric cohort effects



Note: The figure plots the cohort effects estimated using equation (1.10) and the Hamilton filtered log-real GDP (one lag, two leads) for 51 cohorts of firms (1960-2011) that are observed for at least 5 years. The error bars plot the 95% confidence interval for each coefficient.

run trend in firm-level markups can be controlled for by the age and time effects while the cohort effects, by being orthogonal to a time trend, capture the cyclical effects of markups for each cohort.

As a consequence, to check the robustness of the main estimates I compare the series of cohort-effects coefficients estimated in equation (1.10) with the measures of business cycles used as proxy variables in the main specifications of the paper.

Results. Figure 1.9 plots the series of normalized coefficients in equation (1.10) and the detrended GDP in the US. Visually, the time series of estimated cohort effects is very volatile, but it indicates that periods where aggregate GDP is above trend are also associated with negative cohort effects on markup. In fact, the two time series show a correlation coefficient of -0.37, significant at the 1% level, as reported in Table 1.8. In addition, as shown in the table, this negative correlation is preserved also with other proxy variables, hence I interpret these results as supplementary evidence for cohort effects on markups and their negative correlation with aggregate conditions.

Table 1.8: Correlation between non-parametric cohort effects and measures of aggregate conditions

	Correlation coefficient with cohort effects
Quadratically detrended log-GDP	-0.376
	(0.006)
Hodrick-Prescott filtered log-GDP	-0.075
	(0.596)
Hamilton filtered log-GDP	-0.362
	(0.008)
Unemployment rate	0.277
	(0.047)

Note: p-value in parenthesis. The table reports the Pearson correlation coefficients, and the associated p-values, between cohort effects estimated using equation 1.10 and plotted in Figure 1.9, with aggregate measures of aggregate conditions. I restrict the sample to firms that are observed for at least 5 years.

1.6.5 Sub-sample with founding dates

Besides listing, another point in firms' lives when entrepreneurs face possibly irreversible product market choices that have long-lasting repercussions on their business activity is the period in which they found their companies. An obvious limitations of using Compustat data is that firms balance sheets are observed only since few years before the time or their first filing, however, using CRSP it is possible to merge Compustat data to the Field-Ritter dataset and assign founding dates to approximately 34% of firms in my baseline sample. This allows to perform a similar analysis to the one developed in the previous sections with the important difference that I can define cohorts using firms actual founding dates rather than from the first observed date of their balance sheets.

Age profile of markups. Figure 1.10 plots the age profiles for this sub-sample. As expected, the estimated age profile is steeper and implies a 50% increase in markups when firms are observed since their founding date.

Cohort effects. The effects of initial aggregate conditions are much stronger in this sub-sample and increasing over time. This is a major difference with respect to the age profiles computed in the baseline sample. In addition, the reason why the interaction of business cycle measures and age group fixed effects are not significant at very early ages can be due to the fact that the sample in this case is extremely selected as these cells are populated by firms that go public almost immediately after being founded.

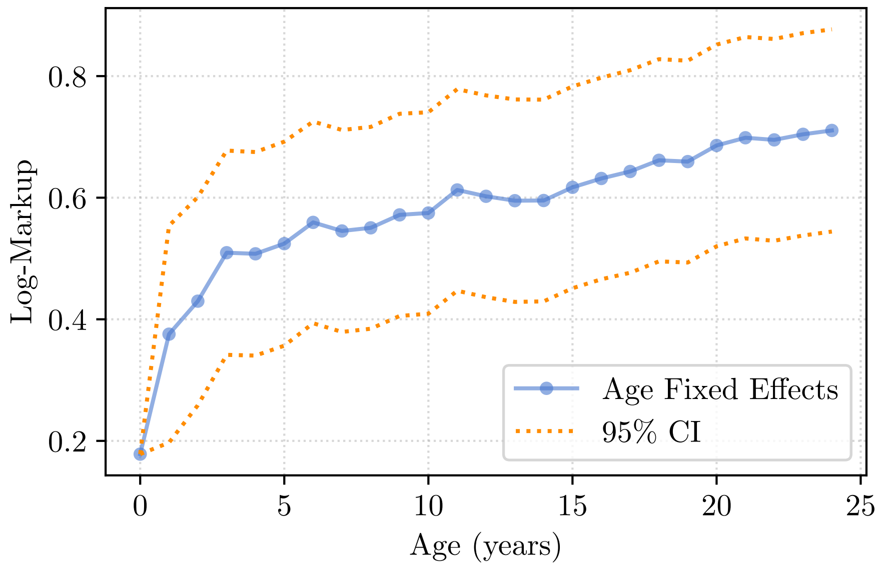
With all the caveats that this very selected dataset has, it confirms the baseline thesis of this paper that aggregate conditions can have persistent effects on the way firms manage their markups. Figure 1.11 plots the effects of aggregate conditions at birth. In this case the effect of business cycle does not vanish over time but it is actually persistent leading to a more than 40% difference between markups of firms born at opposite moments of the business cycle, i.e. firms founded contemporaneously to positive two-standard deviation change in cyclical GDP versus a negative two-standard deviation change.

Table 1.9: Cohort effects on markups - from founding dates

Dep. Variable: Log-Markup	(1)	(2)	(3)
Cycle Measure	-1.988***	-1.785***	
	(0.152)	(0.441)	
Cycle Measure × Age		-0.014	
		(0.026)	
Cycle Measure × Age_1			-3.912
			(3.073)
Cycle Measure × Age_2			-0.454
			(1.303)
Cycle Measure × Age_3			-1.312
			(1.104)
Cycle Measure × Age_4			-1.123
			(0.865)
Cycle Measure × Age_5			-1.867**
			(0.905)
Cycle Measure × Age_{6-10}			-2.326***
			(0.435)
Cycle Measure × Age_{11-15}			-1.136***
			(0.332)
Cycle Measure × Age_{16-20}			-1.558***
			(0.295)
Cycle Measure × Age_{21-25}			-2.079***
			(0.272)
Age FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Initial Size Decile FE	Yes	Yes	Yes
Initial Sale Decile FE	No	No	No
R^2	0.15	0.15	0.15
N	30,577	30,577	30,577

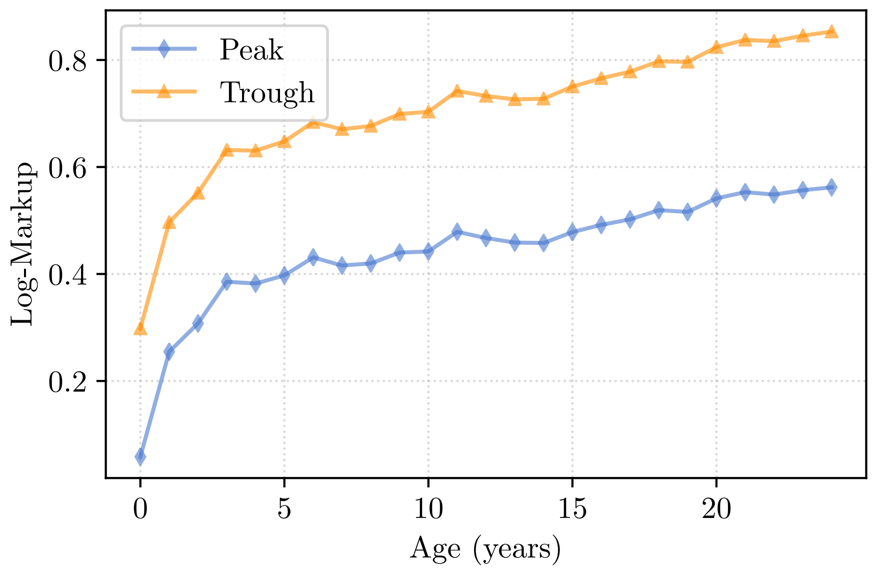
Note: Robust standard errors in parentheses, p :$^{***} < 0.01,^{**} < 0.05,^{*} < 0.1$. The table reports the main coefficient of interest for the specifications in equation (1.8) and (1.9) plus age group specific age effects. The sample is constructed retrieving founding dates for a subsample of Compustat firms using the *Field-Ritter Dataset of Fouding Dates*. Cohort effects are proxied with aggregate conditions at the *founding* date rather than the first available year, which is closer to the start of the IPO process. The measure of business cycle is quadratic detrended real log-GDP.

Figure 1.10: Markups' age profile - sub-sample from founding dates



Note: The figure plots the estimated age profile for markups from equation (1.8) together with the 95% confidence interval. Specifically, at each age a, I am plotting $\hat{\mu}_0 + \hat{\phi}_a$, where $\hat{\mu}_0$ is the average log-markup in the first available year and $\hat{\phi}_a$ are the estimated age fixed effects.

Figure 1.11: Cohort effects on markups' age profile - sub-sample from founding dates



Note: The figure plots the age profile for markups estimated from equation (1.9). Specifically, at each age a, I am plotting $\hat{\mu}_0+\hat{\phi}_a+(\hat{\beta}_0+\hat{\beta}_1 a)Z$, where $\hat{\mu}_0$ is the average log-markup in the first available year, $\hat{\phi}_a$ are the estimated age fixed effects, $(\hat{\beta}_0, \hat{\beta}_1)$ capture the estimated persistence of aggregate conditions in the first available year. Z is a measure of business cycle outcomes that takes value 2σ for *Peak* and -2σ for *Trough*, with σ being the standard deviation of quadratically detrended real GDP. The persistence coefficients underlying this plot are reported in Table 1.9, Column (2).

1.7 Summary and concluding remarks

In this paper I estimate the persistence of aggregate conditions at time of listing for a sample of US listed firms. I adapt an age-period-cohort model to firm-level data using a proxy variable approach to solve the well-known identification issues in these types of models. Specifically, I proxy cohort fixed effects with aggregate conditions close to the time of listing. I find that firms that are first observed in periods of bad aggregate conditions end up charging an higher markup throughout their lives. The average effect on markups is between -0.29% and -0.38% for a 1% change in the cyclical component of GDP. To put it in perspective, these effects imply that firms that were close to list at the height of the Great Recession, in 2009, were able to charge an average markup up to 3% higher than similar firms that instead started their listing process in 2007.

Decomposing this average effects along firms' life cycles uncovers profound differences in the impact of initial aggregate conditions of the age profile of markups. In particular, the effect of aggregate conditions are stronger in the first years of firms lives - with effects ranging approximately between -0.62% to -0.91% for a 1% change in the cyclical component of GDP - and then fading approximately after twelve to fifteen years. This has profound effects on the estimated age profiles for firms that start their listing process in booms compared to similar firms that instead decide to go public in recession, with the latter exhibiting higher markups on average but a flatter age profile.

Given the data used in this paper and the inherent nature of markups, that makes them very hard to measure, I run a battery of robustness checks aimed at minimizing concerns on the relevance of markup measurements for the results uncovered in the paper. In particular, I consider different measures for the elasticity of output to variable inputs by using different specifications for the production function and by estimating it directly using cost-shares. In all cases the results are not significantly affected by this choices. An interesting results is linked to the time horizon in the estimation of the production function. Given the length of the sample used, allowing for a sector-time varying technology proves to be important to correctly capture the changing nature of production technologies over time and fully capture the effects of business cycles. In addition, I address the concerns relative to difficulties of consistently measuring variable costs by estimating cohort effects also for markups constructed using two different measures of costs. In both cases, the results for cohort effects are both qualitatively and quantitatively very similar.

Despite the range of robustness checks, the core of the analysis is centered around the sales to variable cost ratio as the main estimator of markups. In a recent work, Bond *et al.* [2020] highlight how this particular estimator suffers from first-order identification issues linked to both the way in which firm output is measured, revenue instead of quantities, and the type of input bundle I need to use to estimate both the production function and markups. They show that when revenues instead of quantities are used in the estimation of the production function, the ratio estimator does not identify markups. To correct for this issue as much as possible, in the estimation of the production function I follow De Loecker *et al.* [2020] and I include controls to mitigate the bias in the estimated revenue elasticity. This is a clear second-best solution and more work is definitely needed to extend the analysis carried out in this paper to a setting with better data where the identification of firm-level markups can be improved.

The effects reported in this paper indicate a long-lasting, albeit not persistent, effect of business cycles on firm-level markups. This finding complements recent developments in the firm dynamics literature that find significant effects of business cycles on firm sizes. In this literature, rather than firm-level productivity, whereby more productive firms are born during booms, aggregate conditions at the time of inception and the resulting easiness of expanding their customer bases appear to drive these size differences. As recessions are periods in which it is hard to find customers, firms that start their business in these periods suffer from a persistent size gap that never recovers. The fact that I find an effect on markups is suggestive of an additional channel through which business cycles can influence firms' behavior, one in which markups and therefore prices, can be used by firms to expand their customer base.

Appendices

1.A Details on data sources

The main data source used in the paper is the annual version of North-America COMPUSTAT accessed through WRDS on April, 2020. For the sub-sample of firms with founding dates, I rely on the *Field-Ritter Database of Founding Dates*. I merge it to the main sample using the links between firm identifiers in COMPUSTAT and firm identifiers in the CRSP Database (accessed using WRDS) and the Field-Ritter dataset. The aggregate time series at annual frequency for US real GDP (series ID: `GDPCA`) and unemployment rate (series ID: `UNRATE`) are from FRED.

Sample selection and data cleaning

- I consider only domestic, consolidated statements of industrial firms incorporated in the United States and that report their balance sheet in U.S. dollars;

- I exclude financial firms (SIC codes: 6900-6999) and utilities (SIC codes: 4900-4999);

- The capital stock is computed using the perpetual inventory method (PIM), iterating forward on the capital accumulation equation, $k_{j,t} = (1-\delta_{j,t})k_{j,t-1} + i_{j,t}$.

  I initialize the value of the capital stock, $k_{j,0}$, with the first non-missing observation for the gross value of property, plant and equipment (`PPEGT`). For each firm, I then construct a series of net investments, $i_{j,t} - \delta_{j,t}k_{j,t-1}$, by taking the first difference of the reported net value of property plant and equipment (`PPENT`), i.e $i_{j,t} - \delta_{j,t}k_{j,t-1} = \texttt{PPENT}_{j,t} - \texttt{PPENT}_{j,t-1}$. In the few cases in which the value of `PPENT` is missing, I take a liner interpolation within each reporting spell using neighboring values before constructing the investment series. I deflate the series of net investment using the deflator for private, non-residential fixed investments (BEA code: `B008RG`);

- I drop observations with negative sales, capital or employment;

- For the main empirical results presented in the paper, I restrict the analysis to firms that report their first accounting data between 1960 and 2017. The main reason of focusing on this restricted sample is to avoid considering the set of very old firms that, by construction, report their first balance sheet in 1950. The results for the whole sample of firms are qualitatively very similar.

Table 1.A.1: Summary statistics

	Sales	Cost of Goods Sold	Employment	Capital Stock *(book value)*	Capital Stock *(PIM)*	Age
Mean	1,896.29	1,309.84	7.41	1,660.52	1,217.25	13.16
Standard Dev.	10,706.65	8,162.24	35.50	11,423.76	7,865.71	10.62
25^{th} Percentile	26.814	16.47	0.15	9.46	7.04	5.00
Median	139.10	87.62	0.80	53.87	39.63	10.00
75^{th} Percentile	659.27	430.44	3.54	318.22	227.79	19.00
N	181,173	181,173	165,657	179,910	178,603	183,381

Note: Main variables used for the estimation of the production function and the measurement of markups. Financial variables are deflated using GDP deflator (base year 2010). COMPUSTAT sample 1960-2017, firms observed from initial year of available accounting data.

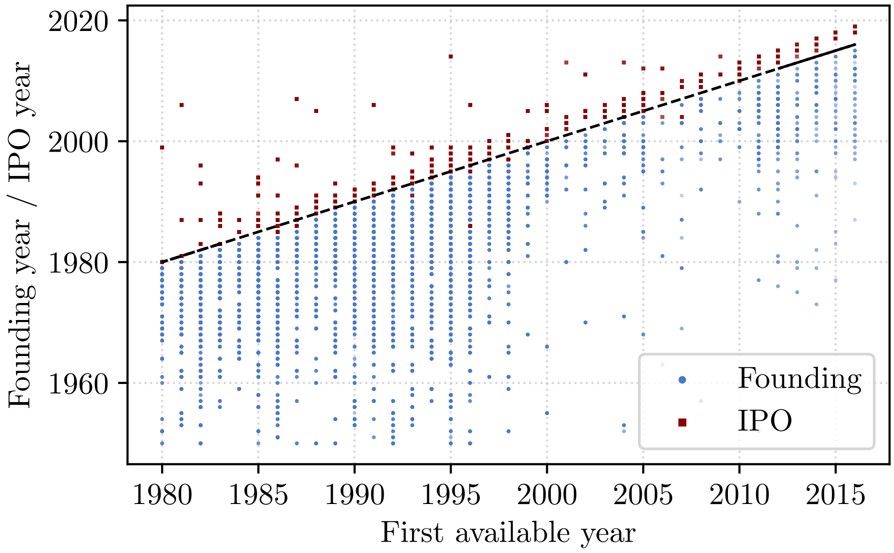
Table 1.A.2: Summary statistics, markups

	Markup			
	Cobb-Douglas *(time varying)*	Cobb-Douglas	TransLog	Cost Share
Mean	1.52	1.53	1.50	1.57
Standard Dev.	0.99	1.04	1.01	1.10
25^{th} Percentile	1.09	1.10	1.07	1.12
Median	1.30	1.30	1.30	1.34
75^{th} Percentile	1.64	1.65	1.64	1.69
N	174,581	183,381	177,386	181,173

Note: Markups measures. COMPUSTAT sample 1960-2017, firms observed from initial year of available accounting data.

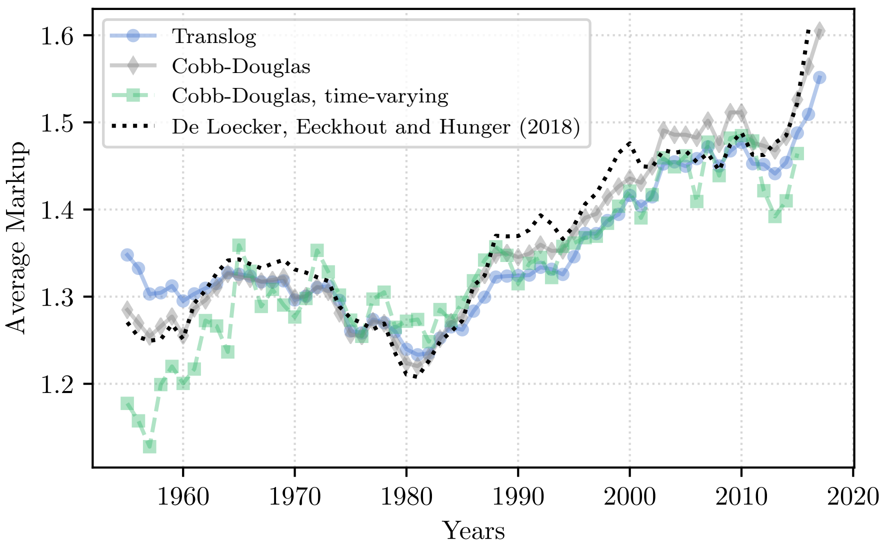
1.B Additional figures

Figure 1.B.1: Joint distribution of *First Available Year-IPO Year* and *First Available Year-Founding Year*



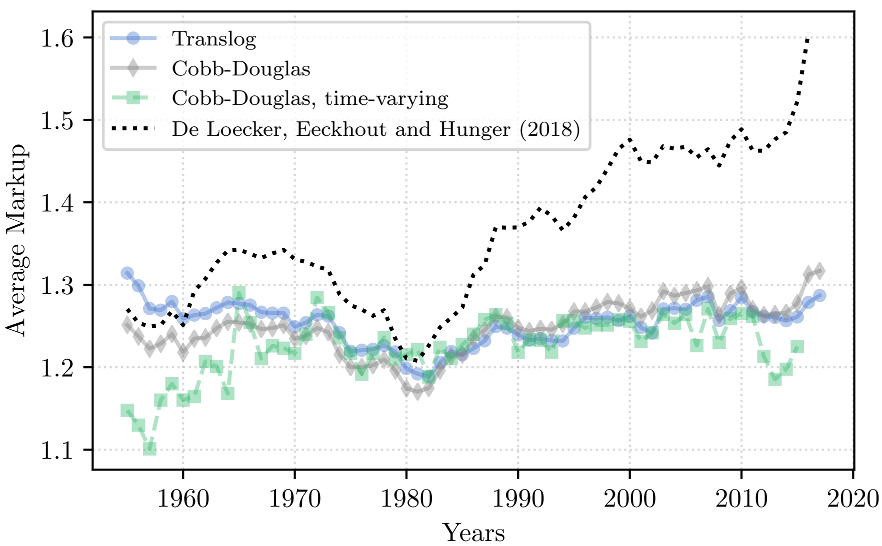
Note: The figure plots the joint distribution of *First Available Year-IPO Date First Available Year-Founding Year* for the subset of Compustat firms for which I have the date of the IPO. Each dot is an observation in the sample and darker colors indicate higher mass.

Figure 1.B.2: Sale weighted aggregate markups with different estimates for the elasticity of output to variable inputs



Note: The figure plots the sale weighted average markup in the sample. The dotted line is from De Loecker *et al.* [2018].

Figure 1.B.3: Cost weighted aggregate markups with different estimates for the elasticity of output to variable inputs



Note: The figure plots the cost-weighted average markup in the sample. The measure of cost is *COGS*. Adding capital costs to the weighting generates a trend more similar to Figure 1.B.2.

Chapter 2

Initial aggregate conditions and heterogeneity in firm-level markups

Abstract

In this paper, I explore the role of aggregate fluctuations as a persistent determinant of heterogeneity in firm-level markups. To analyze how business cycles generate dispersion in firm-level markups, I estimate the age profile of markups at the firm level and I use it to calibrate a general equilibrium model that features heterogeneous product markets, customer base accumulation and firm dynamics. In the model, firms' demands are constrained by the size of their customer bases. However, firms can accumulate customers using two complement channels: i) increasing sales by lowering prices and ii) making direct investments in customer acquisition. As the value of operating in each product market fluctuates endogenously with business cycles, aggregate conditions generate a selection on the product market composition of the cohorts of listing firms that results in heterogeneity in the optimal customer acquisition processes of active firms. This heterogeneity, is persistent and is able to significantly affect the response of the economy to aggregate shocks.

2.1 Introduction

Recent research has found significant and persistent effects of initial aggregate conditions on firm sizes. Firms that start their activities during bad economic times suffer from a size gap compared to similar types of businesses that start during booms. Notably, rather than firm-level productivity, whereby more productive firms are born during booms, aggregate conditions at the time of inception and the resulting easiness of expanding their customer bases appear to drive these size differences. [1] Interestingly, as aggregate conditions can influence other aspects of firms' operations, a similar mechanism could potentially affect also how firms manage variables that are commonly considered more volatile and less history-dependent, like prices and consequently markups.

The prominence of price-cost markups as one of the critical variables for the transmission of aggregate shocks is well established in macroeconomics. Until recently, however, estimates of their co-movements with business cycles have relied mostly on aggregate or industry-level measures.[2] The increased availability of firm-level data and the development of methods that allow estimating markups at finer levels of aggregation have spurred an increase in the number of papers studying the dynamics of markups at the firm level. As most of these studies focus on long-run dynamics,[3] shedding more light on the determinants of markups heterogeneity and their interplay with business cycles, is a promising avenue to uncover new channels for the transmission of aggregate shocks at frequencies higher than the ones currently dominating the literature.

In this chapter, I build on the empirical analysis developed in chapter 1 on the persistence of business cycle effects on firm-level markups and I complement it with an analysis of the contemporaneous correlation of markups with business cycles along firms' age profiles. In addition, I develop a theoretical framework that is able to match three main facts about firm-level markups and their business cycle behavior.[4]

[1] See Sedláček and Sterk [2017], Moreira [2015].

[2] See, among others, Rotemberg and Woodford [1999] and Bils [1987]. For a recent analysis of how aggregate macroeconomic shocks transmit to aggregate and industry-level markups see Nekarda and Ramey [2013].

[3] Relevant contributions in this field are, among other, De Loecker *et al.* [2020], Traina [2018], Autor *et al.* [2017], Hall [1988].

[4] I discuss the details of the empirical methodology used to estimate markups and the persistence of business cycles on markups in Chapter 1.

First, I show that the estimated age profile of markups is mildly increasing in firms' age. Second, I estimate the effect of aggregate conditions at the time of listing on the age profile. Third, I show that firm-level markups are countercyclical and that the countercyclicality is not constant throughout firms' lives. Importantly, older and larger firms are associated with less countercyclical, if not procyclical, markups. Contrary to other studies that highlight how markups' countercyclicality is decreasing as firms age, I show that the cross-sectional heterogeneity in markups' can be at the root of the observed difference between the cyclicality of aggregate measures of markups, that are usually found to be acyclical or mildly procyclical, and firm-level estimates of markup cyclicality that instead are countercyclical.

The reasons why these facts are particularly interesting from a macroeconomic perspective are twofold. First, in a world where the composition of firms in the economy is not fixed, the macroeconomy's response to aggregate shocks is bound to be shaped by the characteristics of active firms. Second, the cross-sectional composition of active firms can be persistently affected by lagged realizations of the business cycle, as aggregate conditions early in firms lives or close to firms' structural decisions, like listing, have persistent effects on how firms manage their markups. In light of these results, a model that is able to characterize both the firm-level and the aggregate behavior of markups and, at the same time, incorporate firms' life cycles can be useful to identify the relevant margins of heterogeneity in the cross-section of active firms and to quantify how relevant the composition of firms' characteristics is for the transmission of aggregate shocks.

With these empirical evidence at hand, I rationalize the empirical facts summarized above by building a firm dynamics model with product-market heterogeneity and customer base accumulation. The novelty of the mechanism proposed in this paper relies on the fact that firms use two joint channels to accumulate customers: i) increasing sales, and ii) directly investing in a range of activities that are specifically designed to increase the reach of their products.[5]

[5] The importance of pricing in models with customer markets dates back to Phelps and Winter [1970], Bils [1989] and more recently Nakamura and Steinsson [2011]. In the firm dynamics and pricing literature, many studies have looked at the effects of investing in goods' quality to acquire customers or in the role of sales for customers' acquisition. However, the combination of these two channels and its effects on markups has not been explicitly tackled. In this literature, the papers that are the closest in spirit to this one are Sedláček and Sterk [2017] and Hong [2017]. Among others, see Perla [2019], Gilbukh and Roldan [2018], Gourio

The mechanism I explore relies on an implicit selection on the cohorts of newly listed firms operated by business cycles. Specifically, in the model each prospective public firm's profitability depends on the product market in which it operates, which in turn dictates whether it is more important to list during a recession or an expansion. The fact that business cycles influence the listing choice of firms is based on the observation that the economy's aggregate state does not uniformly influence the profitability of firms operating in different product markets. As the option value of becoming a public firm fluctuates endogenously with the cycle and depends on firm product markets[6], the composition of listed firms in the economy can be shaped by the aggregate state at the time of their choice. As firms age, and gain even more relevance in the overall economy, they help transmit past business cycles realizations to the aggregate economy through the cross-sectional composition of active firms that gets shaped also by the history of business cycle realizations.[7]

In the model, business cycles have different effects on the profitability of firms because, depending on the product market in which they are active, firms face different trade-offs between using prices and directly investing in customer base. The main distinction between these two channels is in the nature of the investment they entail and how effective the two are in relaxing firms' demand constraints. For example, acquiring customers by selling more output can be easily achieved by lowering prices. This strategy could be an effective way of increasing demand for goods whose production can be easily scaled to satisfy large markets and where customers are more focused on prices and care less about brand values or other goods' characteristics. In these markets, therefore, the price lever could be particularly convenient as it enables firms to increase their output immediately and, at the same time, lock more customers in their specific varieties. Hence, for firms deciding to operate in these product markets, the acquisition of new customers relies on acting prevalently on prices. For firms that operate in less scalable markets or where customers are

and Rudanko [2014], Argente *et al.* [2018], Fitzgerald and Priolo [2018], Bilbiie *et al.* [2019], Edmond *et al.* [2018], Kueng *et al.* [2014], Foster *et al.* [2016].

[6] Among other, Stoughton *et al.* [2001], Chemmanur and He [2011], Chemmanur *et al.* [2018] provide empirical evidence that not only funding needs but also many factors, linked to firms' product market characteristics, influence their decisions of going public at specific times.

[7] Heterogeneous financial frictions have also been used in the literature to rationalize differences in firm-level markups, see Gilchrist *et al.* [2016] and Altomonte *et al.* [2018]. However, in the context of the mechanism proposed here, financial frictions would exacerbate the effects of business cycles on markups and therefore are not included. Nevertheless, their interactions with customer markets and markup cyclicality is an exciting topic that I leave for future research.

highly sensitive to other goods' characteristics besides prices, the expansion of their customer bases relies more heavily on targeted investments in product characteristics, like quality, marketing, or other forms of intangible capital. In these cases, investments in demand creation necessarily require a diversion of resources from production to direct customer base acquisition.

Given that firms face demands with different sensitivities to customer bases and the acquisition of new customers relies more heavily on firms' sales in some markets but not in others, aggregate conditions can affect the relative efficacy of these two investments and persistently impact the aggregate economy by skewing the composition of public firms' cohorts towards product markets that rely more heavily on either one of the two channels. Newly listed firms will operate in the product market that appears to be more profitable, depending on the aggregate state, at the time of their decision. As I assume a common cost of listing across markets, fluctuations in expected profitability create the premises for an endogenous selection of newly listed firms due to business cycles.

In addition, I show that it is necessary to allow for investments in customer base to depend on firms' sales to generate a realistic life-cycle path for firm-level markups. Without this dependence, the model would not be able to reproduce the estimated age profiles. Allowing firms to acquire customers with an alternative channel, instead, is essential to generate cohort-effects and cyclicality profiles that are close in magnitude to the empirically observed ones.

I calibrate the stationary solution of the model to match the estimated age profile of firm-level markups. Interestingly, the parametrization that delivers the best fit implies a negative relationship, across product markets, between the sensitivity of demands to customer bases and the relevance of firms' sales in attracting customers. On the one hand, product markets in which firms can benefit a lot from expansions in customer bases are also those where the price lever is less relevant. On the other hand, firms operating in product markets with relatively lower demand sensitivities exhibit higher dependence on sales for their customer accumulation process. Under this baseline parametrization, I show that the model is able to replicate the cyclicality profile of markups and generate at the same time strong countercyclical markups for small and young firms and pro-cyclical aggregate markups. This feature is consistent with the numerous empirical evidence both at the aggregate and at the firm level.

Relation to the literature. A large number of studies in macroeconomics have established the importance of firm dynamics for the behavior of output, investment and employment along the business cycle.[8] However, the connection between aggregate fluctuations and firm characteristics is a topic that has gained relevance only recently, particularly in relation to the deep causes of observed firm heterogeneity and how this heterogeneity can be connected to business cycles.[9] Are ex-ante firm characteristics the main driver of existing differences in the cross-section of firms or the history of idiosyncratic shocks for each firm is the main determinant of the observed firm outcomes? A compelling thesis of this literature is that firms are endowed with ex-ante heterogeneous characteristics that play a significant role in determining their economic outcomes (Pugsley *et al.* [2019], Hottman *et al.* [2016], Foster *et al.* [2016, 2008]).

Sedláček and Sterk [2017], Hoffmann [2017], Moreira [2015], Alp [2019] develop models where business cycles can select new firms on ex-ante characteristics and show that business cycles at inception is an important determinant of employment fluctuations and a persistent source of firm heterogeneity. These papers are very close in spirit to this one, albeit their primary focus is on firm size while I stress that the channels analyzed in these studies have the ability to affect also firms' markup choices persistently. Hence, I am contributing to this literature by adding a new margin along which firms can be permanently affected by business cycles, not only their size but also the way they manage their markups.[10]

Customers and their valuation of firms' products have a central role in the framework developed in this paper. Recently, the importance of firm and products' appeal in explaining the dispersion in firms' outcomes, such as sales and size, has been empirically investigated using various structural approaches. For example, Hottman *et al.* [2016] find that firms' appeal and the number of

[8] Among many others see Ottonello and Winberry [2019], Carvalho and Grassi [2019], Bloom *et al.* [2018], Clementi and Palazzo [2016], Fort *et al.* [2013] , Moscarini and Postel-Vinay [2012] and Khan and Thomas [2008].

[9] Notable examples in this literature are Luttmer [2007] and more recently Arkolakis [2016] and Bernard *et al.* [2019].

[10] Another paper explicitly models firm growth and heterogeneous markups along firms' life cycles is Peters [2018]. However, his focus is on long growth and the ensuing misallocation effects of markups rather than business-cycle consequences.

products they produce can account for a large share of the observed variation in firms' sales and are a key component in explaining firm heterogeneity.[11] Similarly, Argente *et al.* [2018] estimate the life cycle of revenues at the product level finding that a large share of the variation in firms' sales can be explained by product appeal. Using firm-level transaction data and a structural model of the production network, Bernard *et al.* [2019] find that downstream factors, especially the number of customers, are a key determinant of the observed variation in firms' sales.

In addition, I also touch upon the vast literature on the cyclical behavior of markups. This literature can be broadly divided into two strands. The first one that, mostly due to data limitations, focuses prevalently at the aggregate or industry-level measure of markups see, for example, Ramey [2016], Bils [1987], Nekarda and Ramey [2013]. The second one, more recent, that instead uses firm-level data to estimate markups at the firm-level and links markups with firm-level incentive mechanisms and time-varying market structures that determine the business cycle behavior of markups at the firm level. Notable examples in this literature are Jaimovich and Floetotto [2008], Bilbiie *et al.* [2012], Hong [2017], Gilbukh and Roldan [2018] and Burstein *et al.* [2019]. Anderson *et al.* [2018], using data on listed US firms find that firm-level gross margins, a measure highly correlated with markups, do not respond to high-frequency monetary policy and oil price shocks, suggesting that markups could be acyclical also at the firm-level. The main reason for the difference with the results discussed in this chapter is the use of a different subsample. In fact, to maintain consistency with other data sources used in their study, Anderson *et al.* [2018] focus only on firms belonging to the retail sector. For the analysis developed here, instead, I include firms in all sectors of the economy.[12]

The use of customer capital to deliver countercyclical markups is not novel, see for example Gourio and Rudanko [2014], Ravn *et al.* [2006], Bils [1989], the contribution of the analytical model developed in this paper is that it considers two complementary channels through which firms can acquire customers, and it highlights how business cycles can permanently influence firms' choices

[11] Specifically, using a structural model and product-level data they estimate that four-fifths of the variation in firms' sales can be attributed to firm appeal and product scope.

[12] Albeit large, the retail sector accounts for approximately 10% of aggregate US employment on average. Including all firms in the Compustat sample, instead, allows to cover approximately 30% of aggregate employment on average.

on how to accumulate customer capital most effectively .[13] More recently, the importance of customer base in firm dynamics model has been recently explored by Neiman and Vavra [2019], Bornstein [2018], Perla [2019]. Neiman and Vavra [2019] document an increase in niche consumption and they use an heterogeneous agents model to look at product level data. They use a framework close to the one used in this paper but they do not focus on the behavior of markups nor on business cycle implications of their documented consumer behavior. Similarly, Bornstein [2018] uses a model with heterogeneous taste shocks to highlight how population aging can impact business dynamism. Overall, this literature highlights how the connections between firms and customers can have profound repercussions on the macroeconomy. Paciello *et al.* [2019], instead, study how customer retention incentives influence the pass-through of productivity shocks to firm-level markups. I contribute to this strand of studies by analyzing the relationship between customer acquisition and markups when firms are allowed to use both prices and direct investments in customer acquisition to manage their customer base.

The importance of product selection to match many business cycle features of firms balance sheets, such as procyclical profits and countercyclical markups, has been investigated by Dhingra and Morrow [2019] and Bilbiie *et al.* [2012] among others. More recently, Bilbiie *et al.* [2019] focus on the importance of product selection as a determinant of market power. In this paper, I follow the spirit of the analysis carried out in these papers, but I consider firms' product market choices not as a continuous choice but something that can be done only infrequently and at particular junctures of firms' lives.[14]

Structure of the paper. The paper is structured as follows: Section 2.2 briefly summarizes the motivating evidences on the correlations between aggregate conditions early in firms lives and markups that are explored in more details in Chapter 1. Section 2.3 develops a theoretical model that incorporates incentives to accumulate customer base together with the documented effects of aggregate conditions on markups. Section 2.4 discusses the calibra-

[13] On a similar note, Crouzet and Eberly [2018], Cavenaile and Roldan [2019] and Belo *et al.* [2019], analyze the importance of intangible capital for market concentration and firms' valuations. In this paper's context, investment in customer base is akin to a form of intangible investments, but the focus is again shifted from the analysis of market power to the interaction of business cycles and markups.

[14] See Stoughton *et al.* [2001], Wies and Moorman [2015], Bernstein [2015] and Chemmanur *et al.* [2018], Chemmanur and He [2011] to see how the decision of going public is linked not only linked to funding requirements but also to product-market considerations. González [2020]...

tion of the model and the main functional form assumptions while Sections 2.5 and 2.6 present the main implications of the model. Finally, Section 2.7 concludes.

2.2 Motivating evidence

In this section, I summarize the empirical evidences on firm-level markups used to motivate and then calibrate the model developed in this paper. More details on the empirical analysis, its limitations and various robustness checks are discussed in Chapter 1.

2.2.1 Persistent effects of business cycles on markups

There is robust evidence of persistent effects of business cycles on firm sizes (Moreira [2015], Sedláček and Sterk [2017]). In this section, I extend the existing analysis on firm sizes to firm-level markups. In particular, I break the multicollinearity in a standard cohort-age-time model by proxying cohort fixed effects with aggregate conditions close to the time of listing. This strategy allows to study how business cycle realizations at the time of relevant structural decisions about firms operations and organization, such as listing, can be informative on how firms manage their markups and the persistence of business cycle realizations on the cross-sectional composition of active firms.

The main firm level specification that I bring to the data to estimate the persistence of the correlation between initial aggregate conditions and firm level markups is the following:

$$\log(\mu_{j,c,t}) = \alpha + \phi_a + \phi_t + \beta Y_c + \mathbf{X}_{j,t}\gamma + u_{j,c,t}, \tag{2.1}$$

where $\mu_{j,c,t}$ is the markup charged by a firm j belonging to cohort c at time t; ϕ_a and ϕ_t are respectively age and time fixed effects; Y_c is a measure of initial aggregate conditions for firms belonging to cohort c and $\mathbf{X}_{j,t}$ is a vector of controls that includes sector fixed effects and a second order polynomial in firm j's sale share in her three digits sector.

The main coefficient of interest is β, that, under the standard exogeneity restrictions, pins down the percent change in average markups resulting from a one percent variation in the initial business cycle conditions, after controlling for aggregate conditions faced by firms throughout their lives and the aging process. The age fixed effects, instead, are estimates of a counterfactual age profile of markups that reflects only the dynamics that can be attributed to the aging process, controlling for cohort and time effects.

The specification in equation (2.1) allows to measure the average effect on yearly markups but does not allow to verify if the effects of business cycles are persistent or tend to vanish as firms age. To assess these features of aggregate conditions around the time of listing, I rely on the following specification,

$$\log(\mu_{j,c,t}) = \alpha + \phi_a + \phi_t + \beta_0 Y_c + \beta_1 Y_c \times a_{j,t} + \mathbf{X}_{j,t}\boldsymbol{\gamma} + u_{j,c,t}. \tag{2.2}$$

The coefficient of interest in this case are β_0, that captures the effect on aggregate conditions in the first year, and β_1, that estimates the supplementary effect of business cycle realizations for each additional year firm j is observed.

The specification in equation (2.2) assumes a monotonic effect of age on the effects of aggregate conditions at birth. As a robustness check, I also consider a less demanding specification that allows for a non-monotonic effect of age by interacting the realizations of aggregate GDP with a set of dummy variables for each age.

2.2.2 Cyclicality of firm level markups along the age profile

It is reasonable to think that the elasticity of markups with respect to contemporaneous aggregate conditions is not constant over firm life cycles. For example, Gilchrist *et al.* [2017] show that for reasons linked to the availability of internal funds, smaller firms tend to react more strongly to aggregate shocks.[15]

To quantify how the elasticity to current business cycle conditions changes along firms' age profiles I am considering the following specification:

$$\log(\mu_{j,t}) = \alpha + \beta_0 Y_t + \beta_1 Y_t \times a_{j,t} + \beta_2 Y_t \times a_{j,t}^2 + \phi_a + \phi_j + \mathbf{X}_{j,t}\boldsymbol{\gamma} + u_{j,t}, \tag{2.3}$$

where ϕ_a are age fixed effects, ϕ_j are firm fixed effects and $a_{j,t}$ is firm j's age at time t measured in years since the first available year of balance sheet data. The object of interest is the set of β_i with $i = \{0, 1, 2\}$ as they estimate

[15] Hong [2017] estimates the cyclicality profile for a subsample of European firms providing additional evidence for the counterciclicality of markups at the firm level.

the strength of the contemporaneous elasticity between the business cycle and the firm-level markup at different stages of firms' life-cycle, depicting the evolution of the correlation between output and markups as firms age. I am not including time fixed effects as I would like the identifying variation for the cyclical response of markups to come from within-firm changes in markups. The inclusion of firm fixed effects in the specification ensures that any variation that could be attributed to cohort-level factors is already controlled for by this set of fixed effects, allowing me to not include aggregate conditions at the time of listing as an additional control in this specification. As discussed in the previous section instead, for the estimation of cohort effects and the age profile, I use time fixed effects rather than GDP realizations to control for contemporaneous aggregate conditions as they allow to simultaneously control for all time-dependent characteristics that could otherwise bias the estimation of markups' age profile.

2.2.3 Summary of empirical evidence

The empirical analysis unveils three main facts on the relationship of firm-level markups with business cycles:

1. Markups increase as firms age, see Figure 2.2a;
2. Aggregate conditions at the time of listing exhibit a negative correlation with the average markups firms charge. Moreover, firms that are first observed during recessions report higher initial markups but flatter age profiles, see Figure 2.13a;
3. The contemporaneous correlation between firm-level markups and business cycles is negative and decreasing with age and size, see Table 2.2, Column(1).

In the next sections of the paper, I explore how the combination of sales and direct demand investment in a model where firms are required to accumulated customers, can account for these empirical facts and help reconciling the fact that aggregate measures of markups are less countercyclical, if not acyclical or even procyclical, with respect to firm-level ones.

2.3 A model with product market selection and customer base accumulation

I consider an economy where firms operate in different product markets and where demands are constrained by the size of firms' customer bases. Firms can expand their customer bases in two ways: i) either by increasing sales, something easily achievable by cutting prices or ii) by devoting some resources specifically to the creation of customer base.[16]

2.3.1 Demographics and preferences

The economy comprises two sides: a production side and a consumption side. Time is discrete and denoted by t.

The production side of the economy is focused on listed firms. In particular, it is populated by heterogeneous firms, indexed by j, that produce a set of differentiated varieties and operate in distinct product markets, indexed by i. These firms have the incentive to accumulate customers using both sales and direct marketing investments. The degree to which firm demands are responsive to the size of customer bases and how relevant sales are in the accumulation of new customers depend on the product market in which firms operate. Importantly, firm types are fixed and are not allowed to change also at time of listing. Nonetheless, the composition of listing firms is allowed to fluctuates endogenously. This is because the profitability of listing responds endogenously to aggregate shocks and its response is heterogeneous across product markets. Exit is exogenous but determined by each firm's age after listing, denoted by a.

The consumption side of the economy, instead, is populated by a representative household that makes standard consumption, saving and labor supply choices.

2.3.2 Consumption side

The economy is populated by a unit measure of identical households. They derive utility from consuming an aggregate index of all varieties produced by

[16] In Sedláček and Sterk [2017] customer bases are managed only through firms' direct investments that, in that context, are assimilated to marketing expenditure. Other papers in the literature (see among others Moreira [2015], Hong [2017] and Gourio and Rudanko [2014] as recent examples), instead, assume that firms' customer capital is accumulated exclusively through firms' output. In this paper I focus on the combination of both channels.

listed firm. Notably, households form habits over each variety produced by listed firms.[17]

I assume that households aggregate varieties produced by listed firms using the following aggregator

$$C_t = \left(\sum_i \int_{j(i) \in \mathcal{J}_t(i)} k_i(b_{j(i),t})^{\frac{1}{\eta}} c_{j(i),t}^{\frac{\eta-1}{\eta}} \mathrm{d}j(i) \right)^{\frac{\eta}{\eta-1}} \tag{2.4}$$

where, with a slight abuse of notation, $\mathcal{J}_t(i)$ is the set of listed firms operating in product market i, $b_{j(i),t}$ is the size of the customer base for firm j in product market i and $c_{j(i),t}$ is consumption of firm $j(i)$'s variety.

Importantly, the utility weight for the household is an increasing function of the customer base served by each variety. This formulation captures the notion that households attach more weight to consumption of varieties that are perceived to be relatively more attractive. Throughout the paper I am assuming that households take the size of the customer base for each variety, $b_{j,t}$, as given when making their consumption choices.

The function $k_i(\cdot)$ captures the sensitivity of each incumbent's demand to the size of its customer base and varies across different product markets.[18] Hence, the elasticity of the function $k_i(\cdot)$ tells by how much firms' demands increase (in percentage terms) when their individual customer bases increase.

To capture the idea that higher demand sensitivity to customer bases should directly translate to higher demand I rely on the following reduced form specification for function $k_i(\cdot)$:[19]

$$k_i(b_{j(i),t}) = \kappa_i b_{j(i),t}^{\varepsilon_b^i}.$$

It is a useful formulation as it guarantees a constant elasticity of the utility weights to the size of firm's customer base, a feature that is going to be helpful

[17]See Ravn *et al.* [2006] for a discussion of the introduction of deep habits in macroeconomic models.

[18]For customer base I mean the set of potential customer of a company, i.e. those that are aware and willing to buy a certain good/service.

[19]I take this functional form for the utility weights from Sedláček and Sterk [2017]. As they show, it is possible to micro-found the utility function by considering the customer base of each firm as a measure of awareness of their products among household members. Goods in which an increase in awareness leads to large increases in demand can be thought of as *mass* goods. In contrast, demands of *niche* goods are not very sensitive to this awareness measure.

in the calibration of the model and to introduce some heterogeneity across product markets that differ primary by the value assigned to ε_b^i.

Household problem

I model the demand side of the economy using a representative household that every period maximizes its utility choosing the level of habit-adjusted aggregate consumption, C_t, savings B_{t+1}, and labor supply, H_t to solve the following maximization problem:

$$\max_{C_t,H_t,B_{t+1}} \mathbb{E}_0 \sum_{t=0}^{+\infty} \beta^t U(C_t, H_t) \tag{2.5}$$

subject to

$$P_t C_t + q_t B_{t+1} = w_t H_t + B_t + \Pi_t,$$

where B_t is the stock of risk-free bond in period t, w_t is the wage and Π_t denotes aggregate profits from the production side of the economy while P_t is the habit-adjusted, aggregate price index.[20]

2.3.3 Production side

The production side of the economy models explicitly only on the activities carried out by listed firms. Given their relevance in the economy,[21] listed firms are the main focus of this paper and they differ between each other along two margins: i) the type of products they sell and ii) their age since listing.

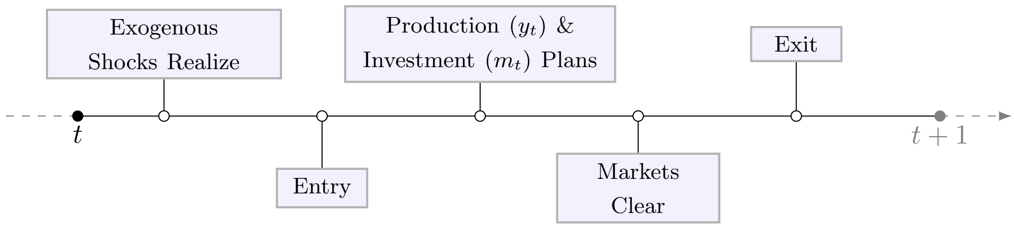
In the background, I assume that the economy is populated by a mass of smaller firms that, given their limited size and resources, do not compete for customers with listed firms.[22] Firm types are fixed and are not affected by the listing process. Every period, some firms go public and the attractiveness of becoming listed varies by the product type each firm sells. For some types, it is more attractive to go public during expansions while for others during recessions. For example, firms that sell products that rely heavily on marketing and similar investments to attract customers might find the option of going

[20] From household expenditure minimization $P_t = \left(\sum_i \int_{j(i)\in\mathcal{J}(i)} p_{j(i),t}^{1-\eta} k_i(b_{j(i),t}) dj(i)\right)^{\frac{1}{1-\eta}}$.

[21] On average, firms included in Compustat account for approximately 30% of non-farm business employment in the US.

[22] An implication of this assumption is that I can abstract from explicitly modeling the behavior of non listed firms as they are not going to affect listed firms' decisions on customer base accumulation.

Figure 2.1: Timeline



Note: Decisions timeline for listed firms in each product market.

public more appealing in expansions rather than in recessions. This formulation is useful in this context as it allows the point of the business cycle at the time of listing to be correlated with ex-ante firm characteristics.[23]

Hence, product types are going to affect two aspects of firms' production. First, given the presence of deep habits on household preferences, firms in each product market will face demands that are sensitive to the size of their customer bases and this sensitivity varies across products. Second, the relative importance that firms' output has in the acquisition of customer is also a characteristics that is specific to the product type in which each firm operates. This margin of heterogeneity, therefore, links the optimal strategy to accumulate customers to the product market faced by each firm.

Incumbent firms

In this section, I describe the behavior of incumbent firms. They are active in a finite number of product markets, indexed by i, and their demands are constrained by the size of their customer bases. Hence, at any point of their life-cycle they face a trade-off between accumulating new customers and harvesting the maximum possible profits from their current customer base.[24]

[23] However, it is also reasonable to consider that, as firms decide go public, they can use the extra funding and publicity to redirect their marketing operations towards markets that appear more profitable when making the listing choice. It is difficult to empirically identify these two motives, but, in the context of the theory developed in this paper, it is important that the correlations summarized in Section 2.2 are informative of either ex-ante firms' types or their choices upon listing. In both cases, there exists a link between business cycles at the time of listing and the cross-section of firms' characteristics that can be used to discipline the model and study the effects of this dimension of firm heterogeneity on the macroeconomy.

[24] The invest/harvest trade-off is common in models of customer markets and it is also documented empirically in Galenianos and Gavazza [2017]. From now on, when there is no risk of confusion, I drop the dependence on the product market in firm subscripts, so that a firm in product market i is denoted simply by j instead of $j(i)$.

Firms' demands. As customers are able to form habits on listed firms' varieties, firm demands are constrained by the size of the customers they are able to reach.

Proposition 1. *Given the consumption index for listed firms as in equation* (2.4)*, a firm operating in product market i faces the following demand function:*

$$y_{j,t} = \left(\frac{p_{j,t}}{P_t}\right)^{-\eta} k_i(b_{j,t})Y_t. \tag{2.6}$$

Proof. Details on the derivation of firms' demand functions are reported in Appendix 2.C. □

Accumulation process for customer base. To model the fact that firms can use multiple channels to relax their demand constraints I am assuming that firms' customer bases evolve according to the following rule,

$$b_{j,t} = (1-\delta)b_{j,t-1} + Q_t F_i(y_{j,t}, m_{j,t}), \tag{2.7}$$

where $m_{j,t}$ is the expenditure on direct demand investments (e.g. marketing, product quality improvements, sale strategies etc) expressed in units of labor and $y_{j,t}$ are firm j's sales, and Q_t is an aggregate stochastic shock that influences the ability of firms to accumulate customers and is meant to capture, in reduced form, the effects of aggregate demand shocks.[25]

Previous studies have mostly focused on formulations of the law of motion for customer base where the acquisition of new customers depended only on either m or y.[26] The novelty of this formulation, therefore, relies on the *joint* role of firms' sales and direct investments in the definition of firms' investment in customer base. In the next sections, after having described the rest of the model, I discuss more in detail how the dependence of customer accumulation

[25] Mechanically Q increases the ability of firms to recruit new customers along all channels. This effect is comparable to aggregate demand shocks if periods of higher than usual aggregate demand are also periods when it is easier to acquire customers. Albeit in very reduced form, this shock is meant to capture this effect.

[26] Sedláček and Sterk [2017], Perla [2019] construct models where firms accumulate customer bases by investing resources in the accumulation of new customers but their sales are irrelevant for the determination of the pool of potential customers. Moreira [2015], Hong [2017], Gilbukh and Roldan [2018], among others, assume that firms build customer bases by locking customers to their products but the only possibility for them to acquire new customers is only through sales. Bornstein [2018] uses taste shocks to lock-in customers in relationships with specific firms and prices are a tool that increases the probability of a good match.

to both channels of investment deeply affects the economy's behavior over the business cycle thanks to the selection effect embedded in this framework.

The function that governs the acquisition of customers by firms is indexed by the product market in which each firm operates as the dependence of investments in customer base to firms' output varies across product markets. Therefore, product markets differ in the sensitivities of demand functions to customer bases and in how strongly the customer accumulation process in each product market depends on sales. To keep the interaction between these two investment channels as tractable as possible I am modeling firms' investment in customer base with the following CES function:

$$F_i(y_{j,t}, m_{j,t}) = [\psi_i y_{j,t}^{\sigma} + (1-\psi_i) m_{j,t}^{\sigma}]^{\frac{1}{\sigma}}. \tag{2.8}$$

The advantage of this formulation is that it allows to capture both the relevance of each channel in the acquisition of customers, through the dependence parameters ψ_i, that are also allowed to vary across product markets, and the complementarity between the two channels, captured by the elasticity of substitution, $\frac{1}{1-\sigma}$.

Incumbents' problem

An incumbent in product market i chooses how much to produce and invest in customer base taking into account the household's demand for its specific variety and the law of motion of its customer base.

Each incumbent internalizes that the size of its customer base can be used to relax the demand constraint, hence, in any period, firms will optimally adjust their prices and direct demand investments to balance their contrasting incentives to increase their customers and extract revenues in every period.[27]

Formally, the optimal choice of the incumbent is the solution to the following problem:

$$V_i(b_{j,t-1}, a_{j,t}; S_t) = \max_{\substack{p_{j,t}, y_{j,t}, h_{j,t}, \\ m_{j,t}, b_{j,t}}} \left\{ \begin{array}{c} p_{j,t} y_{j,t} - w_t \left(h_{j,t} + \zeta(m_{j,t})\right) + \\ +(1-\rho(a_{j,t}))\mathbb{E}_t \left[q_{t,t+1} V_i(b_{j,t}, a_{j,t+1}, S_{t+1})\right] \end{array} \right\}, \tag{2.9}$$

[27]These harvest-investment motives are common in model of customer markets where prices can be used to relax the demand constraint.

subject to the relevant constraints

$$b_{j,t} = (1-\delta)b_{j,t-1} + Q_t F_i(y_{j,t}, m_{j,t}), \tag{2.10}$$

$$y_{j,t} = \left(\frac{p_{j,t}}{P_t}\right)^{-\eta} k_i(b_{j,t})Y_t, \tag{2.11}$$

$$y_{j,t} = A_t h_{j,t}^{\alpha}, \tag{2.12}$$

$$a_{j,t+1} = a_{j,t} + 1, \tag{2.13}$$

where equation (2.10) is the law of motion of customer base, equation (2.11) is the demand constraint, equation (2.12) is the technology available to listed firms and the conditions in (2.13) are the simple evolution of firm j's age.

Proposition 2. *The optimal markup management of listed firms can be summarized by the following condition:*

$$\mu_{j,t}^{-1} - \bar{\mu}^{-1} = Q_t F_y(\cdot|_t)\frac{\varepsilon_b^i}{\eta}\frac{y_{j,t}}{b_{j,t}}\left[1 + \mathbb{E}\sum_{\tau=1}^{\infty}\prod_{s=0}^{\tau-1}\widetilde{\rho}(a_{j,t+s})(1-\delta)^{\tau} q_{t,t+\tau}\frac{p_{j,t+\tau}}{p_{j,t}}\frac{y_{j,t+\tau}}{y_{j,t}}\frac{b_{j,t}}{b_{j,t+\tau}}\right] \tag{2.14}$$

where $\widetilde{\rho}(a_{j,t}) = (1-\rho(a_{j,t})$ *is the surviving probability up to* t, $\mu_{j,t} = \frac{w_t h_{j,t}}{\alpha y_{j,t}}\frac{P_t}{p_{j,t}}$ *is the firm-level markup and* $\bar{\mu}$ *is the markup that would prevail under standard monopolistic competition and without dynamic incentives in pricing.*

Proof. Details on the proof and derivations in Appendix 2.C. □

Looking at the right hand side of (2.14) is helpful to uncover the effects of the main incentive at work in the model. Noticing that it is always positive implies a positive difference between the inverse markup charged by firm j and the inverse of the standard monopolistic competition markup. The fact that the process for the acquisition of new customers induces some dependence on firms' output creates the incentive to set markups *below* the standard monopolistic competition value. This is because firms internalize the positive effects that higher levels of output have on their current and future demands. This incentive structure is also responsible for an increasing age profile of firm-level markups. Note that, as firms age and the size of their customer base increases, the trade-off between keeping prices low to acquire customers and harvesting profits from the customer base already in place moves in favor of the latter, putting an upward pressure on markups. Firm-level markups then converge towards the monopolistic competition level, but only in the limit.

In addition, the fact that firms discount profits using the household stochastic discount factor is responsible for the cyclical behavior of markups. With a procyclical stochastic discount factor, on the margin firms value customers more in booms than in recessions. This implies that, as firms can use sales to attract customers, they are more willing to charge lower markups to optimally increase their customer base in expansions rather than in recessions, generating countercyclicality in firm-level markups.[28] Even if all firms discount the future with the same discount factor, and hence would face identical countercyclical incentives, age fully determines the probability of survival and therefore is a fundamental variable for the determination of the benefit of having a larger customer base in the future. As young firms face a higher probability of exiting the marker, they will have an higher incentive to exploit their current customer base compared to older firms in the same product market. This feature, therefore, is helpful in generating an age profile in the countercyclicality of firm-level markups.

The presence of both sales and direct investments in the determination of firms' investments in customer base is a key element for markup dynamics along the age profile of firms. A model in which firms were able to attract new customers without changing markups, would not generate an increasing age profile of markups as the one observed in the data.

Listing process

I use the listing process and its strict link to product markets as a modeling shortcut to account for a broader range of business decisions that firms take in consideration when going public and that are linked to the way they manage their products. In the finance and marketing literature there are numerous studies that show how the decision of going public is a transforming event for a firm that is influenced not only by funding needs but also by product market considerations.[29]

[28] As noted in Beaudry and Guay [1996], Cooper and Willis [2014] for RBC and Clementi and Palazzo [2016], Winberry [2020] for state-of-the-art heterogeneous firms models, including habit-formation in the utility function is sufficient to generate a procyclical stochastic discount factor. This feature makes the model consistent with recent empirical evidence in finance that finds countercyclical risk-free rates at various time horizons.

[29] For example, Wies and Moorman [2015] founds that after going public firms tend to introduce more varieties of the same product in order to increase sales but holding back on the development of new products. On a related note, Bernstein [2015], finds that going public reduces firms' patent quality. Chemmanur and Yan [2009] established another link between product markets and the decision of going public studying how IPO outcomes can be affected by product market advertising. Chod and Lyandres [2011], instead, highlights how the different ownership structure between private and public firms allows the latter to take

The economy is divided into a fixed number of product markets, $i = 1, ..., I$, and, for each of them, I assume that there is a positive mass of firms in every period. Every period, in each of these product markets, firms compete for access to at most ω_i listing opportunities.[30]

The mass of successful new listings in every product market i, $\Gamma_{i,0,t}$, therefore, can be described by the following matching function linking available opportunities with the mass of firms competing to become listed in product market i at time t, $e_{i,t}$:

$$\Gamma_{i,0,t} = \omega_i^{\phi} e_{i,t}^{1-\phi}. \tag{2.15}$$

Once a match is formed, firms have to pay a common cost denominated in units of aggregate consumption, X_t, to finalize the listing process. Firms that successfully list in a product market, then, will face the same problem as other incumbents with the only difference that they will start their operations with the customer base that they have accumulated in their pre-listing years. To keep the model tractable I do not model explicitly the pre-listing years but in simulating the model I allow for different initial conditions by drawing values for the starting level of customer bases in each product market from distinct uniform distributions. I calibrate the distributions so that the evolution of the sales profiles in the model is consistent with the data.[31]

Free entry implies that, in equilibrium, the expected value of becoming a public firm in each product market with initial customer base $b_{i,0,t-1}$, has to be equal to the cost of listing, hence, for each product market i:

$$\frac{\Gamma_{i,0,t}}{e_{i,t}} V_i(b_{i,0,t-1}, 0, S_t) = X_t. \tag{2.16}$$

Selection mechanism. The free-entry condition ensures that the cost of listing is equal to the expected value of being matched with an available listing

on riskier product market strategies thanks to higher risk sharing among shareholders while Grullon *et al.* [2006] shows how firms capital structure affects their advertising strategies.

[30] Competition among firms in their listing process might ensue for a variety of reasons. For example, firms might compete in their listing process for investors and markets' attention. As the strategic choice of listing, albeit very interesting, is not the main focus of this paper, this margin not explicitly modelled in the paper and is left for future research.

[31] It is possible to easily extend the framework developed in this paper to allow for a mechanical non-listed sector without deeply affecting the results, as model simulations with exogenous processes for the initial customer base highlight that, in this framework, the average sizes of initial conditions are more relevant that their variances. On a similar note, Vardishvili [2018] and Carvalho and Grassi [2019] relax the assumption of a continuum of homogeneous potential entrants in models with firm entry and exit showing that this has a profound effect on the aggregate economy.

opportunity in each product market. It is important to note that, while the cost of listing is common across all product markets, the expected value of becoming a listed firm depends on the demand characteristics in each product market because, depending on the aggregate state of the economy, the value of being listed strongly depends on the incentive to acquire customers, which in turn is governed by the elasticity of the demand shifter to customer base, and whether firms rely more on sales or direct demand investments to acquire customers. For example, the value of becoming listed in product markets where prices are not as effective as direct investments to increase demand is higher in periods where these investments are very effective.

This detachment between a common cost of listing and heterogeneous firm values across product markets is the main driver of the selection effect operated by aggregate conditions at the moment of listing. Given that the value of operating in a specific product market, $V_i(b_{i,0,t-1}, 0, S_t)$, responds differently to the same aggregate shock across product markets but firms that compete to become listed face the same cost regardless of the product market in which they operate. Free entry, therefore, guarantees that the expected value of listing is equalized across product markets, as firm will compete to go public as long as the expected value of being public is equal to the listing cost. Thus, success probabilities in each product market, $\frac{\Gamma_{i,0,t}}{e_{i,t}}$, have to adjust to guarantee that the expected value of listing is equalized across product markets. As the number of opportunities is fixed, any change in success probabilities has to be due to a change in the mass of firms competing and ultimately succeeding to go public in each product market.

2.3.4 Aggregation and market clearing

From (2.14) it is possible to see that incumbents with the same age, a, operating in the same product market, i, face the same problem and therefore will make the same decisions. This observation makes the aggregation very easy as each cohort of listed firms operating in a specific product market behaves as a mass of homogeneous firms.

The resource constraint for listed firms therefore is

$$C_t + \sum_{i}^{I} e_{i,t} X_t = Y_t, \tag{2.17}$$

Labor market clearing requires that

$$\sum_{i}^{I}\sum_{a}^{\bar{a}}\Gamma_{i,a,t}\left(h_{i,a,t}+\frac{m_{i,a,t}^{\chi}}{\chi}\right)=H_t, \tag{2.18}$$

and the law of motion for the mass of listed firms in each cohort-product cell is

$$\Gamma_{i,a,t}=(1-\rho(a-1))\Gamma_{i,a-1,t-1}. \tag{2.19}$$

By definition, aggregate output produced by listed firms has to be such that

$$\sum_{i}^{I}\sum_{a}^{\bar{a}}\Gamma_{i,a,t}p_{i,a,t}y_{i,a,t}=P_tY_t. \tag{2.20}$$

In this economy households are the unique owner of firms, hence, firms' discount their future values using the stochastic discount factor of the representative household that in turn depends on the distribution of listed firms in each product market-cohort cell. As a consequence, the aggregate state of the economy, S_t, includes the distribution of active firms, $\Gamma_{i,a,t}$, in each product market-cohort cell (i,a), plus the set of exogenous processes for aggregate productivity, A_t, the listing cost, X_t and aggregate demand conditions, Q_t.

2.3.5 Equilibrium definition and solution method

Definition (Equilibrium). *A recursive equilibrium in this economy is a set of policy and value functions, for each product market and cohort pair, such that the household and the incumbents' problems, equation* (2.5) *and equation* (2.9) *are satisfied. The free entry condition, equation* (2.16)*, is respected and markets clear following equations* (2.17) *and* (2.18)*. In addition, the distribution of listed firms over age and product markets follows equation* (2.19) *while the exogenous states* (A_t, X_t, Q_t) *follow standard AR(1) processes.*

Solution method. The aggregate state is an infinite dimensional object due to the fact that firms are potentially infinitely lived. As an approximation, I assume that firms have finite lives, as in Sedláček and Sterk [2017]. Paired with the assumption of finite number of product markets, this approximation makes it possible to keep track of each point of the firms' distribution and to solve the model using standard perturbation techniques. The accuracy of the solution increases the longer the life-span of firms. I discuss the model details more in depth in Appendix 2.C.

2.4 Calibration

In this section I discuss the functional forms that I have not yet specified in the description of the model and I discuss the approach followed in the calibration of the model.

One period in the model corresponds to one year in the data. The calibration of the model in geared towards capturing the correct size and shape for the age profile of markups. In particular, the calibration of the subset of parameters controlling product market heterogeneity, i.e. the share of current demand in the accumulation of customer base and the sensitivities of firms' demands to customer base, it is based on matching moments from the stationary solution with their closest empirical counterparts. I discuss the detail more in depth in the following paragraphs.

2.4.1 Functional forms

The production of customer base is governed by the CES function in equation (2.8). Importantly, I am allowing the weight on current output in accumulating customer base to vary across product markets. Direct investments in customer base are subject to the following cost function: $\zeta(m_{i,a,t}) = \frac{m_{i,a,t}^{\chi}}{\chi}$. The exit probabilities at each age are set according to the following rule: $\rho(a) = \rho_0 + \rho_1 a$, with ρ_0 and ρ_1 measured using a simple regression on Compustat exit rates.

With these specified functional forms the model economy is governed by the following parameters: $[\alpha, \beta, \delta, \nu, \chi, X, \phi, \eta, \sigma, \{\varepsilon_b^i\}_{i=1}^I, \{\psi_i\}_{i=1}^I, \{\kappa_i\}_{i=1}^I, \{b_{0,i}\}_{i=1}^I, \{\omega_i\}_{i=1}^I, , \rho_0, \rho_1]$.

I choose the number of product types, I, to match the dispersion in firm sizes conditional on age. Specifically, to allow the model to have firms of the same age but with different sizes, I set $I = 10$ and target the average sizes in each size decile in the data. The reason why I choose to match the dispersion of firm sizes with product markets in the model relies on the observation that firm sizes can be informative of the type of products they produce and act as proxies for the product market in which they operate. Firms that operate in large, scalable markets, for example, where the ability to reach a large number of customer is an important driver of demand, will also be bigger in sizes while firms that operate in more niche markets will tend to remain relatively smaller as their ability to serve their customers does not depend heavily of their size.

The maximum age, $\bar{a}$, is set at 60, which is approximately the 99th percentile of the age distribution in the sample used in the empirical section of the paper.

The utility function of the household is chosen to keep the consumption side of the model as tractable as possible, hence

$$U(C_t, H_t) = \log(C_t) - \nu H_t.$$

Labor disutility, ν, is set so that the habit-adjusted real wage, $\frac{w}{P}$, is equal to the inverse of the monopolistic competition markup, as this particular value is very convenient when solving for the stationary solution of the model.

2.4.2 General parameters

As labor is the only input in production, I set α to 1 so that firms operate with constant returns to scale technologies. The discount factor for the representative household, β, is set to 0.96 for consistency with the macro literature that uses yearly data.

The elasticity of the cost function of direct investments in customer base, χ, is set to 2 and the elasticity of the matching function between potential entrants and available blueprints, ϕ, is 0.156 as in Sedláček and Sterk [2017]. The price elasticity of firms' demands, η, is set to 3.857 to ensure that the markup under monopolistic competition is equal to the average markup of older firms (25 years and more). This parametrization results in a monopolistic competition markup of 1.35 that is higher than the usual values used in the literature. However, the dynamic incentives of the model ensure that the resulting cost-weighted average markup in the economy is approximately 1.3, which is a value closer to the the range usually seen in the literature.

The stationary solution is very sensitive to the depreciation rate of the customer base, δ. I set δ equal to 0.2 which is on the high range of the documented customer capital depreciation rates but it is still much lower than depreciation rates estimated for advertising and marketing expenditures or customer turnover rates documented in the literature.[32]

The elasticity of substitution between sales and direct demand investments for the creation of customer base, $\frac{1}{1-\sigma}$, is set approximately to 0.67 ($\sigma = -0.5$) to capture the fact that current output and direct investments are not perfect substitutes in their ability to attract new customers.

As in the model firms belonging to the same cohort and product market are identical by construction, I use the cyclicality profile for cohort-level markups to calibrate the volatility of the exogenous process so that the model profile matches the estimate one.

2.4.3 Product markets and demand parameters

After having set the values for the general parameters, I am left with the following product market specific parameters that have to be calibrated: $\kappa_i, \psi_i, \omega_i$, ε_b^i and $b_{i,0}$. These are respectively: the scale of firms' demands in each product market, the relevance of sales in customer acquisition, the mass of available listing opportunities, the elasticity of firms' demands to their customer base and the average size of initial customer base upon listing. All these parameters are calibrated so that the model is able to reproduce the reduced form evidence on the age-profiles of firm-level markups discussed in section 2.2.

In particular, the model features two margins of heterogeneity: age and product markets. While age is directly observable, even if with some error, product market types are more difficult to identify directly from firms balance sheet data. However, through the lenses of the model, the product market in which a firm operates is the main determinant of its optimal size. This is because firms have to use labor inputs to attract customers, either through sales or through direct investments, therefore firms that operate in product markets that exhibit higher demand sensitivity will also have bigger optimal sizes. As a consequence, the size distribution of relatively older firms can be informative of their product types as these firms are more likely to have reached their optimal production scale.

The number of listing opportunities in each product market, ω_i, are calibrated to match the share of firms in each size decile and are scaled so that aggregate

[32]For a discussion on recent estimates of customer turnover rates and marketing investments see Gourio and Rudanko [2014] and Foster *et al.* [2016].

demand at the stationary solution of the model is equal to 1. The average sizes of initial customer bases are the last parameters to be calibrated so that the stationary solution of the model exhibits a convex age profile for sales growth in line with the data.[33]

I calibrate ψ_i and ε_b^i so that the average age profile of markups implied by the stationary solution of the model matches the profile estimated using equation (2.1). The intuition behind this choice is linked to the fact that the elasticity of demand with respect to the customer base is a key determinant of the growth rate of firms as it regulates how much firms benefit from an extra unit of customer base.

The values of κ_i, instead, are chosen to match the size distribution of old firms in the data. In practice, for each product market, I choose the value of κ_i such that the size of 25 years old firms in the model equals the average size of firms between 25 and 30 years in the data. As mentioned before, the intuition for this relies on the fact that it is more likely that older firms have already reached their optimal production scale hence while the elasticity of the demand shifter $k_i(b_{j,t})$ is informative of the growth path as firms age, its scale is tightly linked to the optimal size of firms.[34] The full baseline calibration of the model is reported in Table 2.1.

Moments matching procedure

In order to pin down the values of ψ_i and ε_b^i I am using the following procedure. Starting from guesses for each ψ_i and ε_b^i, $\tilde{\psi}^{(0)} = \{\psi_i^{(0)}\}$, and $\tilde{\varepsilon}^{(0)} = \{\varepsilon_b^{i,(0)}\}$, I solve for the stationary distribution of the model and I measure the age profiles for markups as

$$\mathcal{P}_a^{\mu}(\tilde{\psi}^{(0)}, \tilde{\varepsilon}^{(0)}) = \log\left(\frac{\mu_a(\tilde{\psi}^{(0)}, \tilde{\varepsilon}^{(0)})}{\mu_0(\tilde{\psi}^{(0)}, \tilde{\varepsilon}^{(0)})}\right),$$

where $\mu_a = \sum_i \Gamma_{i,a}\mu_{i,a}$ is the average markup at each age across product types. The empirical counterparts for these age profiles are given from the age fixed effects estimated using the model in equation (2.1). The condition I am using

[33] In the model simulations, initial customer bases in each product market are allowed to fluctuate between $[0, 2b_{0,i}]$ by drawing i.i.d. shocks from a set of uniform distributions between $(-1, 1)$. Varying these parameters does not significantly affect the overall behavior of the model. I provide more details in Appendix 2.C.

[34] For example, the size of 25 years old firms in product market one is matched to the average size of firms in the first decile of the size distribution of 25-30 years old firms.

to pin down the values for $\{\psi_i, \varepsilon_b^i\}$ is:

$$(\tilde{\psi}^*, \tilde{\varepsilon}^*) = \arg\min_{\psi_i, \varepsilon_b^i} \left\{ \left(\mathcal{P}_a^{\mu}(\tilde{\psi}, \tilde{\varepsilon}) - \hat{\phi}_a \right)' \left(\mathcal{P}_a^{\mu}(\tilde{\psi}, \tilde{\varepsilon}) - \hat{\phi}_a \right) \right\}.$$

In practice, I include the first forty age fixed effects from equation (2.1) and the growth rates relative to age-zero at the stationary solution to pin down twenty parameters. The model, however, does not perfectly replicate the age profile of markups due to the non-linearities in the mapping between the age profile of markups at the stationary solution and the estimated age fixed-effects.[35]

[35] A procedure similar in spirit to the one described here is adopted in Christiano *et al.* [2005] and Hong [2017].

Table 2.1: Calibration

Parameter	**Value**	**Target**
α Technology	1	CRS technolgy
β Discount factor	0.96	Annual discount rate in macro-literature
γ Utility weight of non-listed goods	2.17	Employment in non-listed firms equal to 2/3
δ Customer base depreciation	0.20	Customer turnover ratio [Gourio and Rudanko, 2014]
η Price elasticity of demand	3.86	Monopolistic markup equal to 1.35
ν Labor disutility	0.766	Habit-adjusted real wage equal to $\frac{\eta-1}{\eta}$
χ Curvature of cost function for customer base investments	2.0	Quadratic cost [Sedláček and Sterk, 2017]
ϕ Matching function elasticity	0.16	[Sedláček and Sterk, 2017]
$\bar{X}$ Average entry cost	33.60	Average success in product market (1) equal to 0.99
σ Elasticity of sales and markups for customer base accumulation	-0.5	Elasticity of markups and advertising $\approx$ 0.67 (Compustat)
ρ_0, ρ_1 Exit rates	0.08, -0.001	Exit rates by age (Compustat)
ρ_A, σ_A Productivity process	0.95, 0.15	Markup ciclicality profile, see Figure 2.5
ρ_Q, σ_Q Demand process	0.95, 0.4	“ ”
ρ_X, σ_X Entry cost process	0.95, 0.6	“ ”
Ω Measure of listing opportunities	$1.105e^{-5}$	Normalization so that aggregate output equals 1 at the stationary solution

Product Market Parameters	**Target**	**Values**									
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ψ_i Share parameter in $F_i(y_{j(i),t}, m_{j(i),t})$	Age profile of markups	0.647	0.568	0.529	0.539	0.524	0.490	0.427	0.336	0.216	0.058
ε_b^i Elasticity of $k_i(b_{j(i),t})$	Age profile of markups	0.424	0.437	0.537	0.740	0.886	1.005	1.101	1.195	1.335	1.635
κ_i Scale of $k_i(b_{j(i),t})$	Average size of mature firms	4.875	14.947	17.543	9.560	6.243	4.529	3.979	3.832	3.026	0.954
$b_{0,i}$ Avg. initial customer base	Sales' growth profile	2.779	6.009	9.194	15.064	21.301	27.313	32.367	36.831	43.284	69.350
$\frac{\Gamma_{i,0}}{e_i}$ Success Probability in i	Share of firms in i when old	0.999	0.231	0.116	0.071	0.045	0.029	0.018	0.01	0.006	0.002

Note: The table reports the main calibration of the model. Note that instead of the values for the masses of listing opportunities in each product market, ω_i, that are of small significance given the normalization, the table reports the implied success probabilities.

2.4.4 Stationary solution and model's fit

The model is calibrated so that the moments implied by the stationary solution and the empirical targets are matched.

Figure 2.2 reports the three main calibration targets: i) growth profile of markups; ii) the size distribution of firms and iii) the age distribution. The growth profile of markups shows the growth rates of markups relative to the initial year for both the model and the data. The non-linearities in the mapping between the model-based growth profiles and the age fixed-effects prevent the model of achieving a perfect fit. Looking at the ratio of average size of firms by product markets with the average size of firms by size decile, the model is able to generate a realistic size distribution.[36] Despite the very mechanic evolution of firm sizes, due to the mechanic law of motion for the distribution of active firms, the model replicates also the age distribution of firms quite well.

Figure 2.3, reports the values for the product market characteristics under the baseline calibration. In particular, the model implies a negative relationship between the share parameters of firms' sales for customer acquisition and the elasticity of demand to customer base. This stark negative relationship between the two implies that firms operating in product markets with high demand sensitivity to customer base are also characterized by a low incentive to use the price lever to attract customers. These characteristics of the demand properties in each product market turn out to be heavily responsible for the dynamic response of the model to aggregate shocks and for the transmission of aggregate shocks as they directly control the way firms use both prices and employment for the acquisition of new customers.[37]

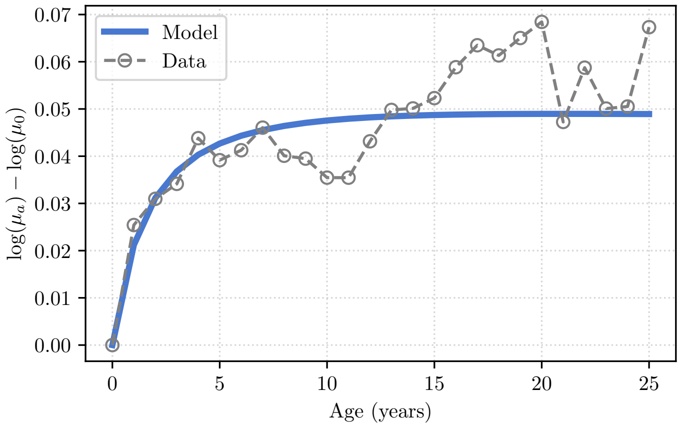
Even if the model is geared toward replicating the age profile of markups it is important to check how it performs in replicating some non-targeted moments in the data. To this end, Figure 2.4 compares the growth profile by age groups for markups and firms size, measured by employment.

[36] The model achieves a perfect fit for firms of age 25 in the model with the average of mature (25 to 30 years) firms in the data. The figure plots the ratio between the averages in the model and in the data across all ages

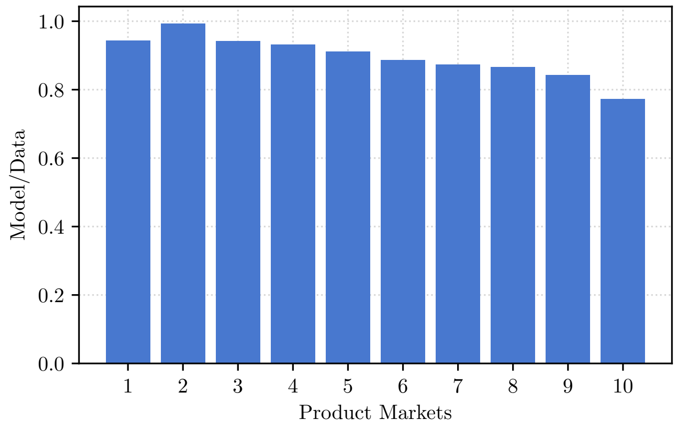
[37] A more careful estimation of these demand parameters, rather than a simple calibration is particularly interesting and of clear importance for the mechanisms outlined in this paper. However, a more detailed discussion and assessment of the identification of these parameters in the class of models used in this paper are left for future research.

Figure 2.2: Stationary solution targets

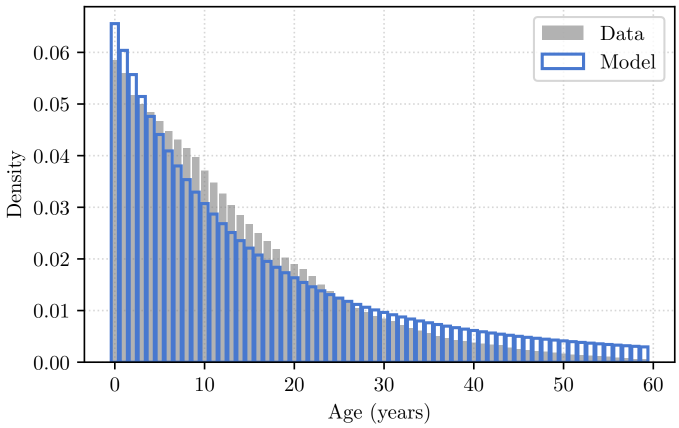
(a) Age profile of markups



(b) Firm sizes

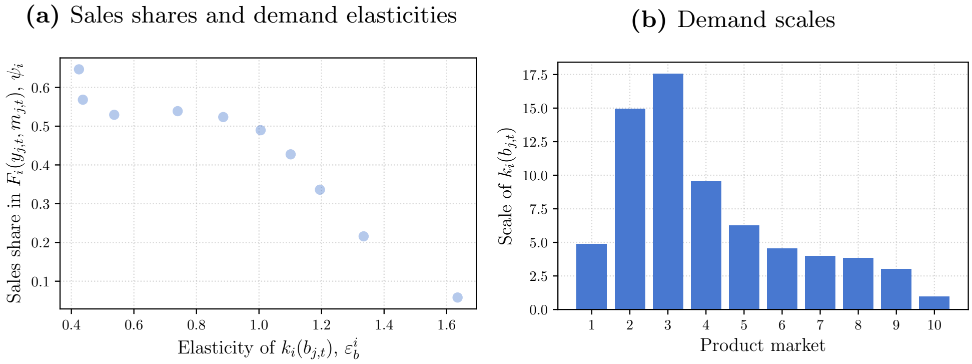


(c) Age distribution



Note: The figure reports the main calibration targets. Panel 2.2a reports the age profile of markups in the model and in the data up to 25 years, from estimating age fixed effects in equation (2.1) in the main sample and at the stationary solution. Panel 2.2b reports the average size of firms in each product market relative to the average size of firms in each size decile in the data. The calibration targets firms age 25 years old in the model, the panel reports the average for all ages. Panel 2.2c reports the age distribution of firms in the the data and at the stationary solution.

Figure 2.3: Product market characteristics in baseline calibration

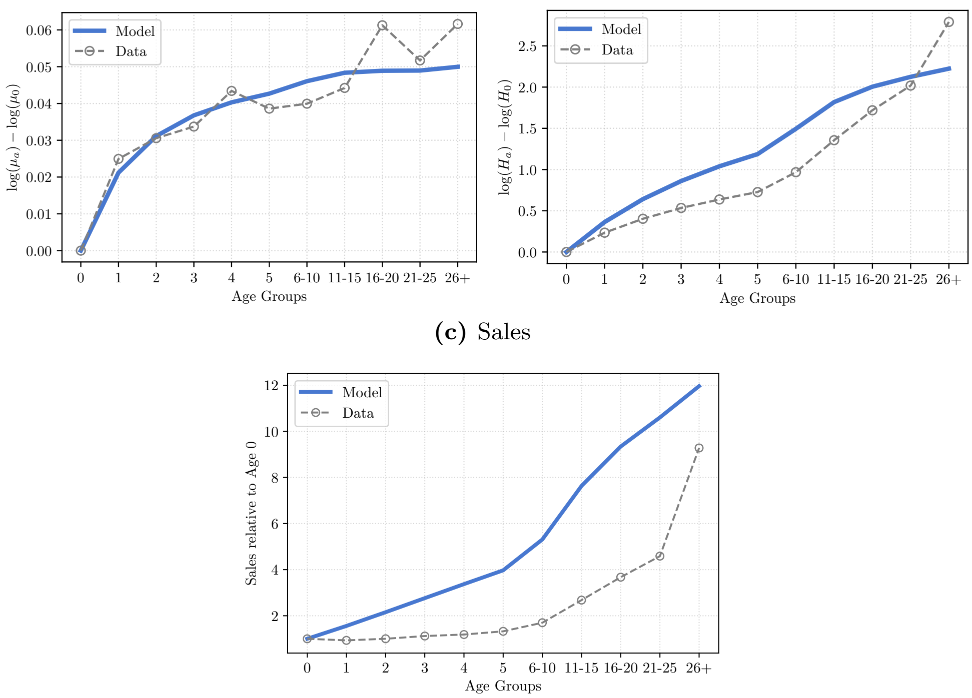


Note: The figure reports the relationship between the parameters determining the heterogeneity across product markets in the model. Panel 2.3a plots the relationship between the elasticity of demand to customer base versus the share of current sales in direct demand investments creation. Panel 2.3b plots the scale parameter in the demand shifter $k_i(b_{j(i),t})$.

The markup profile is tightly linked to the target moments, thus is not surprising that the model replicates this one as well. The employment profile, however, is not targeted and the model is able to fully generate a growth profile that follows the data only for the first years of firms lives. In particular, the model is not able to generate the strong kink observed in the employment growth rates measured in the data. As firms in the model tend to grow relatively fast, the model base profile slightly over-represents the growth rates in early years of firms' lives and then gradually slows down at later years. The main reason for this last effect is mainly mechanical and linked to the solution approach I use. In fact, as I solve for the policy functions and the stationary solution by truncating the lives of firms at 60 years, firms in the model start actively shrinking in size when they get closer to the end of their life. This dynamic puts a downward pressure on firms sizes, preventing the model to fully capture the higher growth of older age groups. A possible solution to this limitation comes at the cost of greater computational complexity. Pushing the end of firms' lives further in time, thereby extending the number of years firms are allowed to live is likely to improve the solution and address this issue better. However, as the main focus of this paper is the behavior of markups and given that the size profile implied by the model, albeit not perfect, follows the one estimated in the data reasonably well, especially for young firms I opt for the most tractable solution and keep the maximum allowed age close the one observed in the data. Linked to the issue discussed in the previous paragraph, comes another limitation of the model again related to the fact that firms grow too fast in the model relatively to the data. As

Figure 2.4: Age profiles - model vs data

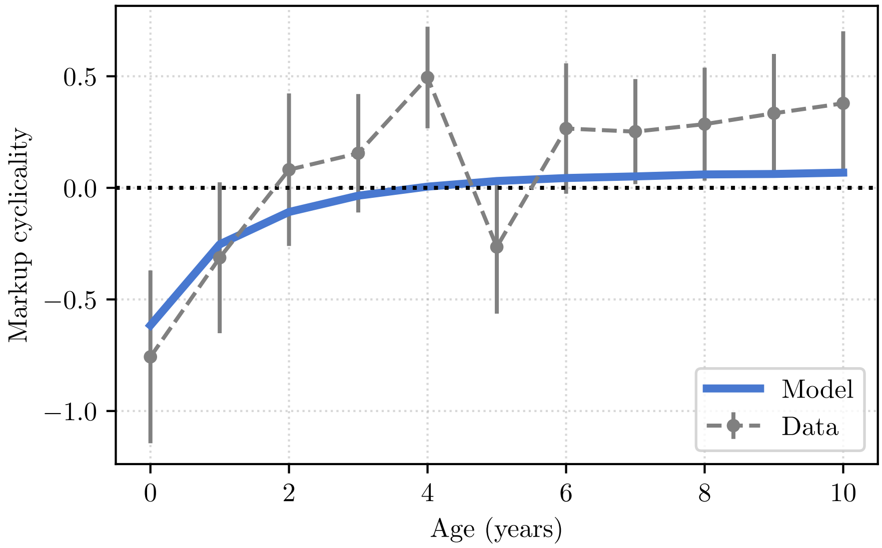
(a) Markups **(b)** Employment **(c)** Sales



Note: The figure plots the implied age profiles for markups, employment and sales in the data and in the stationary solution of the model. Note that while the age profile of markups and the shape of the sales profile are target in the calibration, the age profile of employment is not.

shown in Figure 2.4c, the average sales by age groups in the model exhibits a less convex profile compared to the data. This is due to firms in the model growing too fast in their initial years and slowing down as they reach maturity. This discrepancy can be deduced also by looking at the average growth rates of employment by age group in Figure 2.4. Firms tend to grow too fast early in their lives and then they stop as they approach the end of their lives. The intuition for this behavior hinges on the fact that, in the baseline calibration, sales and direct demand investments are not perfect substitutes. As a consequence, accumulating and maintaining a customer base requires a constant investment of resources, either through direct demand investment or indirectly through lower markups to boost sales. As firms age, the incentive to extract rents from the current customer bases dominates the accumulation incentive and firms start optimally increasing their markups, decreasing their investments in demand creation and optimally shrinking their sizes. Figure 2.5 plots the cyclicality profile for cohort-level markups in the data and the model. The model is able to replicate the ciclicality pattern observed in the data, where

Figure 2.5: Markup cyclicality profile



Note: The figure plots the cyclicality profile estimated at the cohort-level in a sample of cohorts followed up to 10 years and in a sample of model generated data obtained from simulating the model for 2500 periods. More specifically the figure plots the coefficients of the interaction between age dummies and contemporaneous realizations of the cycle component of GDP, i.e. $\hat{\beta}_a$ from $\log(\mu_{c,a,t}) = \alpha + \phi_a + \phi_c + \sum_a^{10} \beta_a D_a \times Z_t + u_{c,a,t}$ where ϕ_a, ϕ_c are respectively age and cohort fixed effects, Z_t is a measure of contemporaneous business cycles and D_a is a set of dummies for each age. *Data* reports the point estimate and a one standard deviation error band. The estimates are based from aggregating the main sample at the cohort level and following cohorts for up to 10 years. *Model* is based on following cohorts for 10 years from a dataset based on model simulations of 2500 periods. In the model the measure of business cycles are deviations of output from the stationary solution.

the correlation of the cohort-level markup is increasing as we consider cohorts of older firms.

2.5 The role of product market heterogeneity and aggregate shocks

In this section, I assess the role of differences across product markets by discussing the model's response to aggregate shocks under different calibrations of the model. Specifically, I progressively eliminate the main channels of heterogeneity across product markets: the sales share in customer base investments and the sensitivity of demand to customer base. I compare the baseline calibration to two alternatives: a first one, in which investment in customer base accumulation have the same sales sensitivity in each product market (constant ψ_i case), and a second one, in which firms' demands exhibit the same elasticity to customer base (constant ε_b^i case).[38]

[38]Figure 2.A.1 plots the two different calibrations together with the baseline. To preserve the size differences across product markets and maintain some consistency with the baseline

In addition, I discuss the role of the exogenous processes in shaping the correlations between aggregate markups and business cycles. Specifically, I simulate the model under different mixes of exogenous shocks and report the correlation coefficients between aggregate output and two measures of aggregate markups, the cost-weighted average, and the simple average.

2.5.1 Impulse response analysis

To assess the role of product market heterogeneity for the transmission of aggregate shocks, I compare the economy's responses under different calibrations of the parameters governing product market heterogeneity. First, I analyze the transmission of productivity shocks, then I look at aggregate demand shocks and finally to listing costs shocks.[39]

Aggregate productivity shocks

Figure 2.6 shows the impulse responses of the economy to a positive one-percent aggregate productivity shock for the baseline calibration, the constant sale share case, and the constant demand elasticity case. I discuss these two alternatives in turn.[40]

Common sales share, ψ_i. Under this calibration, the only difference across product markets is how sensitive firms' demands are to the size of their customer bases. Interestingly, as shown in the first panel of Figure 2.6a, the response of output is practically identical to the baseline case. The intuition for this result follows from noting that, compared to the baseline calibration, the composition of newly listed firms, shown in Figure 2.6b, is not highly affected by the change in TFP when ψ_i is constant. This result indicates that incumbents' profitability is not affected by the relevance of sales in the acquisition of customers. However, as shown from the last two panels of Figure 2.6a, the responses of the cost-weighted and the simple average markups are quite different from the baseline calibration. In particular, setting all product markets to the same average value for ψ induces a large procyclical response of aggregate markups. To grasp some intuition, consider that the aggregate productivity

economy, I recompute the scale parameters in firms' demands, κ_i, to match the average size of old firms in both alternative calibrations.

[39] Appendix 2.A reports the impulse responses to each shock for markups and the mass of firms for each product market and age cohort.

[40] Figure 2.A.2 reports the IRFs to aggregate productivity shocks for markups and masses of active firms by cohort and product market.

shock makes production more efficient, lowering marginal costs. Given that sales are still a method to acquire customers, firms can trade-off the boost in productivity from the aggregate shock with an increase in markups.

This mechanism is particularly strong for bigger and older firms, contributing to the high response of cost-weighted markups with constant ψ_i. In the baseline calibration, instead, firms with high demand sensitivities operate in product markets where sales are relatively ineffective in accumulating customers. This implies that to exploit the higher profitability generated by higher TFP, firms are incentivized to invest in relaxing their demand constraints via direct demand investments. However, as sales and direct investments are complements in the investment function for customer base, higher TFP strengthens these large firms' incentive to accumulate customers, putting downward pressure on aggregate markups.

In addition, given higher productivity, the profitability of firms in all product markets increases, boosting listing. Consequently, the share of young firms increases, lowering the simple average markups in the economy below its steady-state level for some periods. This is because the increase in the share of new firms outweighs the fact that active firms can charge higher markups. When the sale share in customer acquisition, ψ_i, is constant across product markets, the average markup's initial response is less negative than in the baseline calibration. In this case, fewer firms have to rely on direct investment in customer acquisition and can exploit the benefits of being more productive. As a larger fraction of newly listed firms chooses product markets where demand is not very sensitive to customer base accumulation, the average markup overshoots its steady-state value before converging back to it as more firms can exploit the productivity boost to charge higher markups.

Common demand elasticity, ε_b^i. Under this calibration, the only element that differentiates product markets is the role of sales in the acquisition of new customers.

The main effect caused by this alternative calibration is a significant change in the composition of newly listed firms, as shown in the third panel of Figure 2.6b.

Compared to the baseline case and the calibration with differences in ε_b^i, now a positive TFP shock induces a similar increase in listing across product markets. This implies that cohorts of newly listed firms are more homogeneous. This, in turn, generates a response of the average markup that is smaller than under the previous calibration. The intuition hinges on the fact that, in this case, the economy does not feature firms that are very eager to accumulate customers, and hence the initial value of markups that they choose to set is higher compared to the baseline for newly listed firms.

The initial response of the cost-weighted measure of markups, instead, is more muted under this calibration because firms that were previously operating in product markets with high-ψ and low-ε are now facing stronger incentives to accumulate customers. Hence, as productivity increases, the incentive to extract rents from the current customer base is counterbalanced as higher production can lead to larger future profits thanks to the role of sales in building customer base. These two contrasting forces cancel out in the aggregate contributing to the minimal response of cost-weighted markups to TFP shocks. Through the same reasoning, we can rationalize the larger response of output. In this calibration, more firms are eager to use prices to expand their customer bases, which implies that as productivity improves, incumbent firms do not increase markups as much and more output is produced.

It is worth mentioning that the transmission of TFP shocks on markups in this economy is very different from the standard procyclical results induced by price rigidities. In this model, markups' cyclical behavior, both in the aggregate and at the firm level, is dictated more by the dynamics of the incentives to accumulate customers than by the response of marginal costs.

Aggregate demand shocks

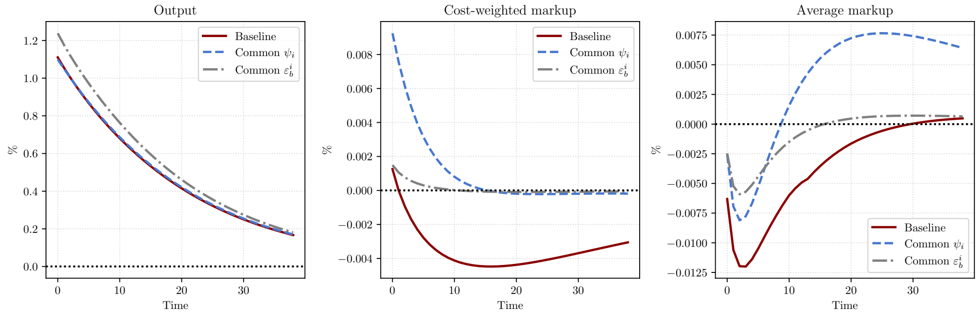
Figure 2.7 plots the impulse responses to a positive one-percent aggregate demand shock for the baseline economy, the economy with constant demand elasticities and the one with constant sales share in customer base investments.[41]

Common sales share, ψ_i. As discussed in the previous section, when the only difference across product markets is the elasticity of demand to customer base, the composition of newly listed firms does not change dramatically compared to the baseline case.

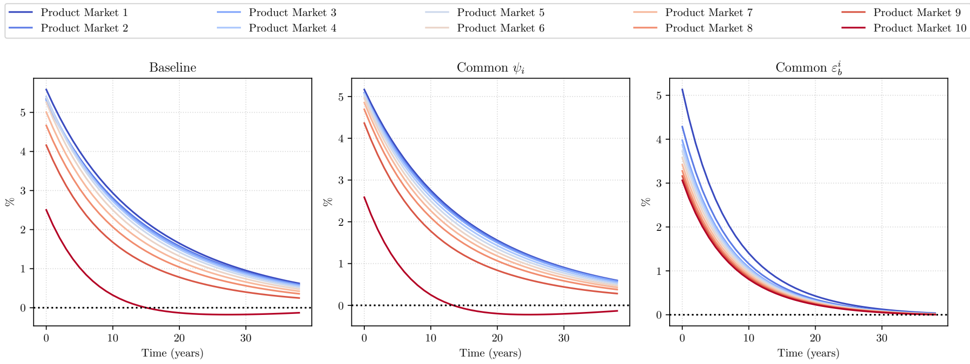
[41] Figure 2.A.3 reports the IRFs to aggregate demand shocks for markups and masses of active firms by cohort and product market.

Figure 2.6: Impulse responses to aggregate productivity shock

(a) Aggregates



(b) New firms in each product market

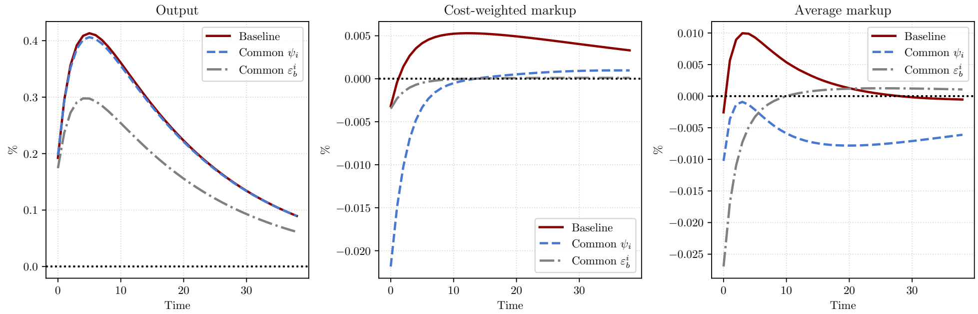


Note: The figure plots the impulse responses to a 1% aggregate productivity shock for selected aggregate variables under three different calibration of product market parameters ψ_i, ε_b^i. *Baseline* shows the IRFs to the baseline calibration of the mode where both ψ_i and ε_b^i vary across product markets. *Common* ψ_i shows the IRFs for a calibration of the model where product markets differ only across the demand elasticity ε_b^i and ψ_i is set at the average value for all i. *Common* ε_b^i, instead reports the IRFs for a calibration of the model where product markets differ only across ψ_i and ε_b^i is set at the average value for all i.

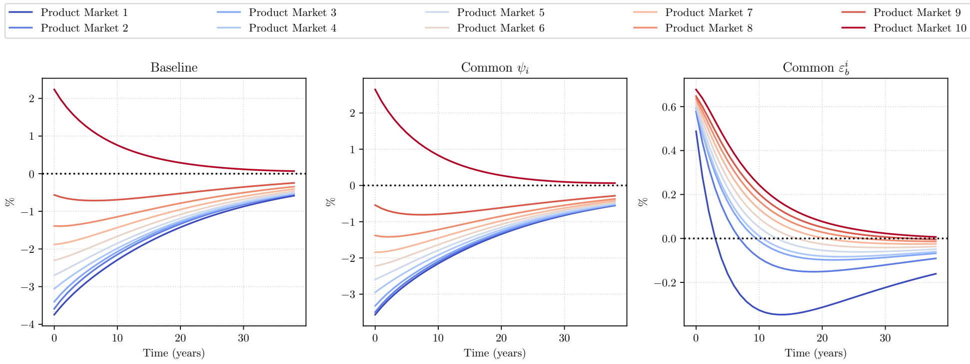
Nonetheless, the response of markups to aggregate demand shocks presents stark differences across the three calibrations. This is due to the deep changes in the composition of newly listed firms caused by fluctuations in Q. In fact, a positive shock to Q triggers a period in which the efficiency of customer base investment is high, which improves the profitability of firms operating in product markets that depend heavily on customer acquisition. In turn, this improvement induces a shift in the composition of new listings towards product markets characterized by highly sensitive demands and low dependence on sales to accumulate customers, as these product markets are relatively more profitable in periods of high demand. However, incumbent firms in these markets can enjoy the enhanced efficiency of their investment and hence, are incentivized to increase their markups. Consequently, both the cost-weighted and the average markups in the economy overshoot their steady state values, reflecting the fact that after an aggregate demand shock, the composition of active firms in the economy shifts towards product markets that rely a lot on

Figure 2.7: Impulse responses to aggregate demand shock

(a) Aggregates



(b) New firms in each product market



Note: The figure plots the impulse responses to a 1% aggregate demand shock for selected aggregate variables under three different calibration of product market parameters ψ_i, ε_b^i. *Baseline* shows the IRFs to the baseline calibration of the mode where both ψ_i and ε_b^i vary across product markets. *Common* ψ_i shows the IRFs for a calibration of the model where product markets differ only across the demand elasticity ε_b^i and ψ_i is set at the average value for all is. *Common* ε_b^i, instead reports the IRFs for a calibration of the model where product markets differ only across ψ_i and ε_b^i is set at the average value for all i.

customers but in which acquiring new customers does not depend a lot on sales and therefore prices.

When the sales share in the customer base investment function is fixed at its average value across product markets, instead, more firms are eager to reduce markups to expand their customer bases in periods of favorable aggregate conditions. This is because firms in highly sensitive product markets now rely relatively more on sales compared to the baseline calibration. As sales and direct demand investment are not substitute, these firms respond to their incentive to accumulate customers by aggressively cutting prices to benefit from the favorable economic climate.

The differences between the cost-weighted share of markups and the average markup can be again rationalized by looking at the change in the composition of newly listed firms. As more young firms are going to operate in product markets where they are eager to accumulate customers relative to the steady-state, on impact, the response of cost-weighted average markups is larger than the simple average. Also, old and large firms are more willing to lower their prices to increase their sales and benefit from easier customer acquisition due to the shock, contributing in keeping aggregate markups below their stationary solution levels.

Common demand sensitivity, ε_b^i. Under this calibration, the line of reasoning outlined in the previous paragraph is partially reversed. As all product markets feature the same demand sensitivity aggregate shocks induce a smaller change in composition, confirming the finding that this particular margin of heterogeneity across product markets is highly relevant for the selection effect of business cycles.

The third panel of Figure 2.7a shows that the average markup exhibit strong counterciclicality. This is a direct consequence of the combination of two facts. First, contrary to the baseline and the calibration with common ψ_i, with homogeneous demand sensitivities, the economy experiences an overall increase in the number of young firms. Second, not only these firms are relatively more eager to accumulate customers, but they also operate in product markets where this incentive is stronger.[42] This pushes down the average markups more than the cost-weighted one as more young firms enter the market and both incumbents and newly listed firms are eager to use the price lever to build their customer base.

Listing cost shocks

Figure 2.8 plots the impulse responses to a one-percent increase to listing cost for the baseline economy, the economy with constant demand elasticities, and the one with constant sales share in customer base investments.

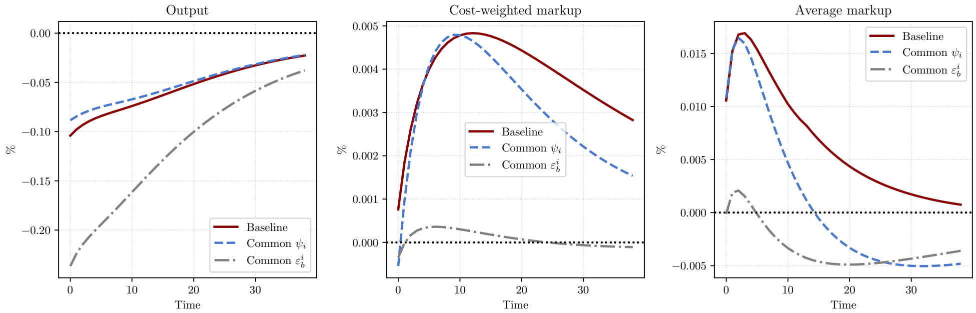
The overall behavior of the economy in response to this shock resembles closely the inverse of aggregate productivity shocks.[43] The intuition relies on noticing that while better productivity pushes firms to go public more in product markets that do not rely heavily on customer acquisition (low ε_b^i), higher listing costs do the opposite, making listing *less* profitable for these firms as they

[42]Notice that high ψ_i product markets, under this calibration, face stronger incentives to accumulate customer as the common ε_b is higher than the one associated in these product markets under the baseline calibration.

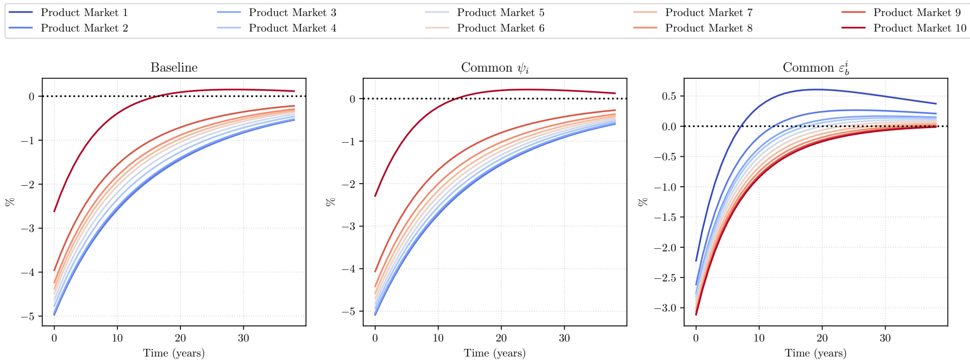
[43]Figure 2.A.4 reports the IRFs to listing cost shocks for markups and masses of active firms by cohort and product market.

Figure 2.8: Impulse responses to listing cost shock

(a) Aggregates



(b) New firms in each product market



Note: The figure plots the impulse responses to a 1% listing cost shock for selected aggregate variables under three different calibration of product market parameters ψ_i, ε_b^i. *Baseline* shows the IRFs to the baseline calibration of the mode where both ψ_i and ε_b^i vary across product markets. *Common* ψ_i shows the IRFs for a calibration of the model where product markets differ only across the demand elasticity ε_b^i and ψ_i is set at the average value for all is. *Common* ε_b^i, instead reports the IRFs for a calibration of the model where product markets differ only across ψ_i and ε_b^i is set at the average value for all i.

operate in product markets that do not have strong growth prospects. As a consequence, as output falls, both cost-weighted and average markup increase countercyclically.

Common sales share, ψ_i. When the model is calibrated to homogenize firms across their incentive to use sales as a tool to acquire customers, the overall response of the economy is very similar to the baseline one. Again, the reason is that having product markets with similar sales share parameters does not dramatically affect the composition of new listings. The transition back to the stationary solution however is different. The average and the cost-weighted markups revert back to their stationary solution levels more quickly, with the average markup overshooting its stationary value. The positive entry cost shock reduces the number of young firms in the economy, and as these firms age, they can increase their markups more quickly than the baseline.

Common demand sensitivity, ε_b^i. The calibration with common demand sensitivity generates very different responses to a one-percent increase in the listing cost. As in the previous section, muting the heterogeneity of demand sensitivities generates a big change in the composition of newly listed firms. In this case, the increase in listing costs, increases the fall in listings of firms that have weaker dependence on customer base accumulation on sales. These firms are also the ones that are bigger and have more growth potential, thus output falls by more than in the baseline cases.

As now all firms use prices to increase their customer base, the responses of both cost-weighted and average markups are muted. In this case, average markups overshoot their stationary value even sooner than under common sales shares. This is because the early cohorts of newly listed firm age are relatively more populated by firms that use prices to attract customers, while they try to reach their optimal size, pushing the average markups below its stationary level.

2.5.2 The role of aggregate shocks

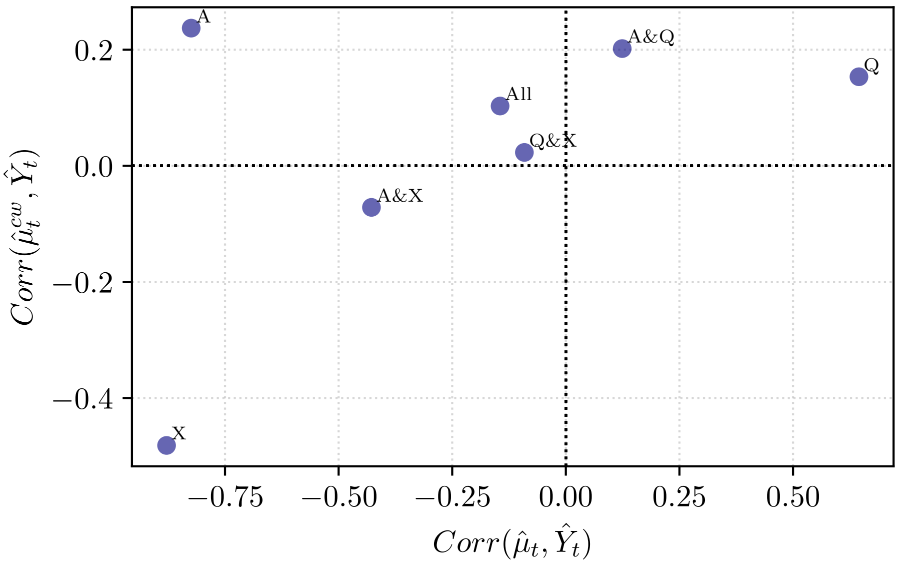
The model developed in this paper features three sources of uncertainty: aggregate productivity shocks, aggregate demand shocks, and listing cost shocks. In this section, I discuss the role of different shock mixes in shaping the correlations between output and aggregate measures of markups.

To this end, Figure 2.9 summarizes the relationships between correlation coefficients of the average and the cost-weighted average markup obtained from simulating the model with different shock mixes. Looking at these two measures of aggregate markups is helpful as they allow to build some intuition on the role of changes to the composition of active firms for markup cyclicality.[44] Two interesting patterns emerge.

First, when aggregate productivity shocks are the unique source of business cycles in the economy, cost-weighted and average markup exhibit a stark distinction in their cyclical behavior. While the former measure is strongly procyclical, the latter is mildly countercyclical. This pattern is consistent with the observation that, as productivity improves, incumbent firms increase their production as well as markups. However, higher volumes due to higher produc-

[44] For mode details on the IRFs to each shock for markups and firm masses by age and product market see Figures 2.A.2, 2.A.3 and 2.A.4.

Figure 2.9: Correlation coefficients with output of cost-weighted and average markup for different shock mixes



Note: The figure plots the correlation coefficients between different shock mixes of the model. $\hat{}$ variables denote deviations from a Hamilton filter trend (one lag, two leads); μ_t^{cw} and μ_t are respectively cost-weighted and average markup. A: TFP shocks, Q: Aggregate demand shocks and X: Listing cost shocks.

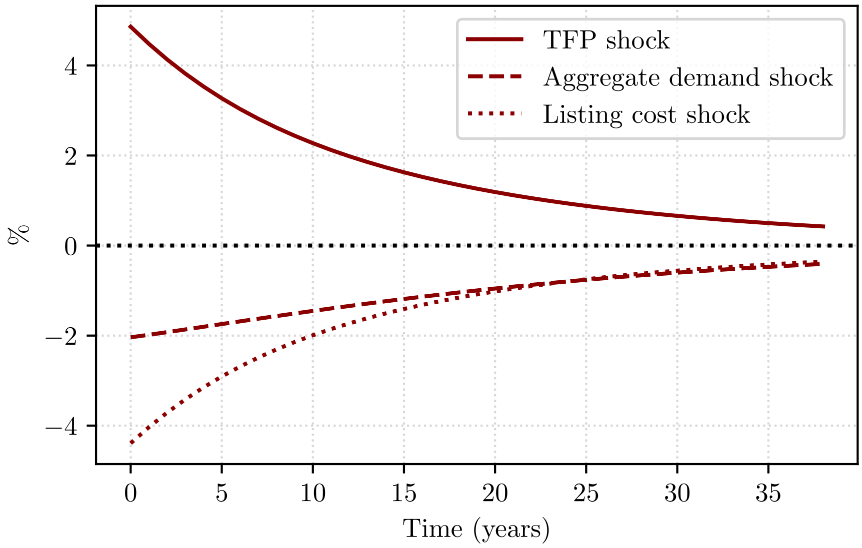
tivity more than compensate the customer loss caused by increase in markups. As noted in the previous section, the average markup in the economy, which reflects the extensive margin more strongly, is countercyclical as positive productivity realizations are associated with a spur in new listings that increase the share of young firms in the economy putting downward pressure on the average markup, see Figures 2.6b and 2.10.

Second, when listing cost shocks are the only sources of aggregate fluctuations, both cost-weighted and average markup exhibit countercyclicality. In contrast, aggregate demand shocks make both measures of markups procyclical, albeit the average exhibits a higher correlation with output than the cost weighted one.

The explanation of these patterns has to searched again in the changes to the composition effects at the extensive margin that these shocks generate, see Figure 2.10 for the IRFs for the total mass of newly listed firms for the three different shocks.

When output is driven exclusively by aggregate demand shocks, high-output periods are also times when investment in customer base is more efficient. This creates incentives for incumbent firms to lower their markups regardless of their product market. As these firms are also relatively larger, at the in-

Figure 2.10: Impulse responses for the total mass of newly listed firms



Note: The figure plots the IRFs for the total mass of newly listed firms to one-percent shocks to aggregate productivity, aggregate demand and listing costs.

tensive margin, cost-weighted markups tend to decrease as output increases following a positive demand shock. At the extensive margin, however, positive aggregate demand shocks reduce the total mass of firms in the economy, as more firms attempt listing in product markets that have smaller success probabilities, those that are highly sensitive to customer base acquisition.[45] This behavior, therefore, skews the cross-section of active firms toward older and larger firms that are both less responsive to the business cycle and that have higher average markups. Hence, as fewer young firms get listed, periods of high demand and high output are also associated with a pool of active firms overpopulated by old incumbents. This change in the composition at the extensive margin pushes for an increase in cost-weighted and average markups and thus a positive correlation with total output. The resulting correlation in Figure 2.9 is the combination of these two forces. The higher procyclicality of average markup can be explained by noting that, by construction, the simple average is a measure more sensitive to changes at the extensive margin, thus is more sensitive to the fact that periods of high-demand will reduce the total number of listings.

The first-order consequence of an increase in listing costs, instead, is that fewer firms are listed in all product markets. This extensive margin response then

[45] See Figure 2.7b. Notice that as customers are easier to acquire, operating in product markets that are highly sensitive to customer base becomes more attractive as these are markets that are ex-ante more profitable.

increases the relative weight of older and larger firms in the economy, putting an upward pressure on average markup. As the reduction in the number of firms is also associated with a decline in total output, this dynamics rationalize the countercyclical behavior of both average and cost-weighted markups, as shown in Figure 2.9.In addition, this effect is stronger for firms operating in product markets with low demand sensitivity to customer base that are also less willing to use prices to manage their customer base. The intuition for this stronger selection effect is that, in these markets, firms reach their optimal size earlier in their life-cycle, hence, their present discounted value is more affected by aggregate conditions at the time of listing, given that business cycles at the beginning of their listed lives have a greater scope to persistently affect their values. All these effects compound to generate a negative correlation between output and the average markup in the economy.

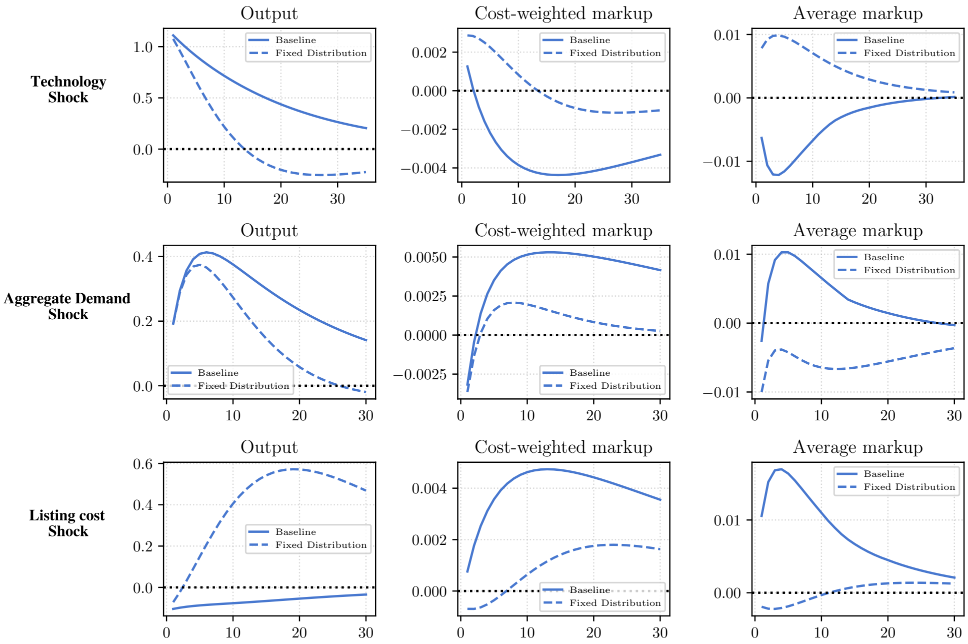
The recessions caused by an increase in listing cost shocks, instead, force especially young firms to lower markups, attempting to counteract the effects of lower sales with lower prices. Older firms, instead, being less sensible to the loss of customers caused by the initial reduction in output are able to recoup their losses in customers faster than young firms and therefore can benefit from relatively higher demand and increase their markups, leading to a lower counterciclicality of cost-weighted markups compared to that of average markup when listing cost shocks are the only source of business cycles in the economy.

As shown by Figure 2.9, combining the three shocks allows the economy to generate both countercyclical average markup, and slightly pro-cyclical aggregate (cost-weighted) markups. Under the baseline calibration, increasing the strength of productivity shocks would push the economy towards generating more pro-cyclical aggregate markups while increasing the role of aggregate demand shocks would induce stronger compositional effects of business cycles.

2.5.3 Extensive versus intensive margin

The response of aggregate markups to aggregate shocks is driven by the combination of the change in incumbents' responses and the change in the product market composition of cohorts of newly listed firms. To disentangle the contribution of the two channels to the correlation of aggregate markups with output

Figure 2.11: Role of firm composition for the transmission of aggregate shocks



Note: The figure plots impulse response to a one-percent change in technology, aggregate demand and listing cost processes. In the simulations, the fixed-distribution variables are constructed using log-linearized versions of (2.21) and (2.22).

I adopt the following strategy. I simulate the model fixing the distribution of active firm in each product market-age cell at its stationary solution value and I recompute the correlations of aggregate markups with output under different shock mixes.

Specifically, I construct average ($\tilde{\mu}_t$) and cost-weighted ($\tilde{\mu}_t^{cw}$) markups with a constant mass of firms as follows:

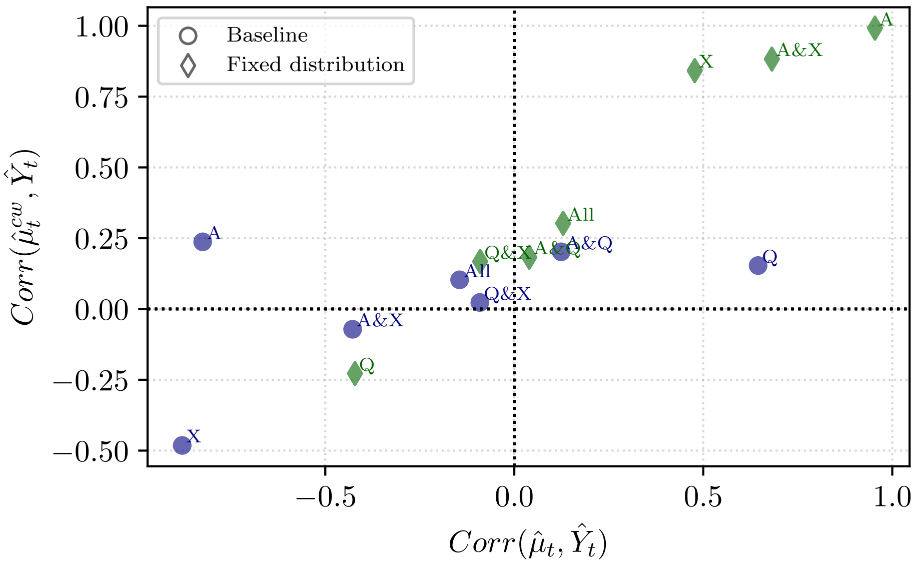
$$\tilde{\mu}_t = \sum_i \sum_a \bar{\Gamma}_{i,a,t} \mu_{i,a,t} / \sum_i \sum_a \bar{\Gamma}_{i,a,t} \tag{2.21}$$

$$\tilde{\mu}_t^{cw} = \sum_i \sum_a \bar{\Gamma}_{i,a,t} h_{i,a,t} \mu_{i,a,t} / \sum_i \sum_a \bar{\Gamma}_{i,a,t} h_{i,a,t}, \tag{2.22}$$

where $\bar{\Gamma}_{i,a}$ indicates the mass of firms with age a that populate product market i absent any aggregate shocks.

Figure 2.11 plots the impulse responses to a one-percent change in the exogenous variables for the average and cost-weighted measures of markups in two versions of the economy: i) the baseline one and ii) one in which the distribution of firms in the economy is kept constant at its stationary solution throughout the simulations.

Figure 2.12: Correlation coefficients with output of cost-weighted and average markup for different shock mixes



Note: The figure plots the correlation cofficients between different shock mixes of the baseline model and a model where average and cost-weighted markups are constructed keeping a the distribution of firms at the stationary-solution values at each simulation period. ˆ variables denote deviations from a Hamilton filter trend (one lag, two leads); μ_t^{cw} and μ_t are respectively the cost-weighted and average markup. A: TFP shocks, Q: Aggregate demand shocks and X: Listing cost shocks.

The figure helps understanding how compositional changes in the cross-section of active firms influences the response of the economy to aggregate shocks. As noted before, the average measure of markups is more responsive to changes to the composition of firm population in the economy. Intuitively, this is due to the fact that differences in product markets and the implicit selection operated by business cycles have more bite on the average that is a measure more sensitive to changes in the overall mass of firms. From the impulse responses in Figure 2.11, we can see that this is particularly true for productivity and listing cost shocks.

The differences in the propagation of aggregate shocks caused by the missing compositional channel has deep repercussions also on how different shocks impact the correlations between output and markups, as show in Figure 2.12. Notably, keeping the distribution of firms at the stationary solution levels, radically affects the correlation between output and markups, especially for productivity and listing cost shocks. These correlations are consistent with the shock propagations highlighted by the impulse response functions presented in Figure 2.11. For simulations in which productivity is the only source of business cycles the model with a constant distribution generates highly procyclical markups, both average and cost-weighted. To see why, consider that a positive productivity shock pushes all firms to increase markups, as the potential loss from an increase in prices is compensated from the gain in customers derived from increased production. However, by keeping the mass of firms in the economy fixed, both the average markup and the cost-weighted one move in the same direction resulting in a highly positive correlation with output.

2.6 Main results

In this section, I describe the main model results. In particular I document the ability of the model to generate age profiles and cohort effects that are in line with the empirical evidence presented in Section 2.2. In addition, I show that the model replicates the observed correlations between markups and output both at the firm and aggregate level. Moreover, I test the ability of the model to capture relevant incentives at the firm-level by estimating cohort effects and age profiles for advertising expenditure, a close empirical counterpart for direct demand investments.

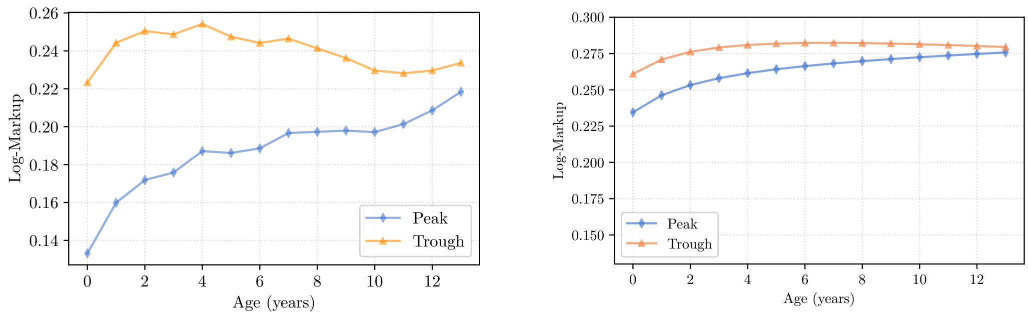
2.6.1 Persistent effects of business cycles

Cohort effects on firm-level markups. The selection mechanism embedded in the listing phase of the model allows business cycles to have persistent effects on the life-cycle behavior of firm-level markups.

Figure 2.13b shows the age profiles estimated using equation (2.2) on a dataset of model-generated data. The figure shows that the model can qualitatively reproduce the age profile and the cohort effects estimated in real data and reported in Figure 2.13a. Firms that start their operations during a recession end up charging higher markups and operate on a flatter age profile. Through the lenses of the model, these effects are highly tied to differences in the demand characteristics of the different product markets. As recessions are periods in which it is difficult to attract customers, new listings are skewed towards product markets that guarantee higher up-front profits. These are specifically those that do not require big investment in demand acquisition and where the optimal size is reached earlier in firms lives. These product markets are characterized by a relatively low dependence of demand on customer base, and this implies that firms operating in these markets have a low need to expand their customer base. This lack of incentives to attract new customers

Figure 2.13: Cohort effects and age profile of markups

(a) Data **(b)** Model



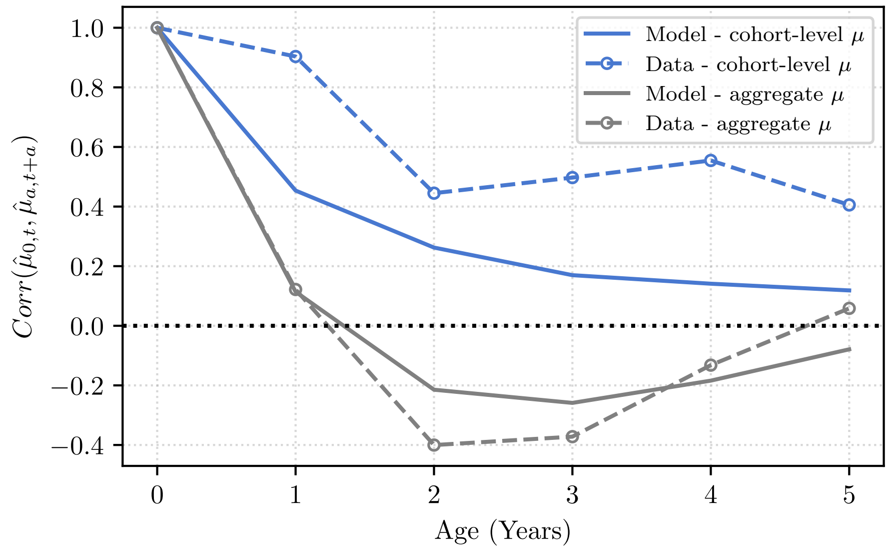
Note: The figure plots the age profile for markups estimated from equation (2.2) for the main sample and for a panel dataset constructed from simulating the model for 2500 periods. Specifically, at each age a, I am plotting $\hat{m}_0 + \hat{\phi}_a + (\hat{\beta}_0 + \hat{\beta}_1 a)Z$, where $\hat{m}_0$ is the average markup in the first available year, $\hat{\phi}_a$ are the estimated age fixed effects, $(\hat{\beta}_0, \hat{\beta}_1)$ capture the estimated persistence of aggregate conditions in the first available year. Z is a measure of business cycle outcomes that takes value 2σ for *Peak* and -2σ for *Trough*, with σ being the standard deviation of listed firm output. In the model the measure of business cycles are deviations of output from the stationary solution.

makes them particularly appealing during bad economic conditions when the option value of expanding the scale of operations is smaller. Moreover, as firms that operate in these markets have a smaller optimal size than firms operating in other product markets, they can afford to charge higher markups from the start of their activities as they have weaker incentives to accumulate customers.

Quantitatively, running the same regression as in (2.2) on model-generated data, reveals that a two-standard deviations positive increase from the stationary solution value of aggregate output induces an initial negative change in firm-level markups of approximately -1.4%, compared to an effect of approximately -4% in the baseline empirical specification. Therefore, the initial effect of aggregate conditions in the model is approximately 35% of the one estimated in the data for similarly sized changes in the cycle components of GDP. One of the main driver of this effect is the role of demand shocks in business cycles. As also noted in the previous section, see Figure 2.7b, increasing the dependence of business cycles to demand shocks increases the ability of the model to generate dispersion in the characteristics of entrants, increasing the magnitude of business cycle effects on the average age profile of markups.

Alternatively, the persistence of business cycles can be checked by comparing the correlation of cohort-level markups at different ages with their initial level.

Figure 2.14: Autocorrelation of aggregate markups in the data and in the model



Note: The figure plots the cohort-level and aggregate autocorrelation of markups in the model and in the data. *Cohort-level* refers to correlations of average markup by cohorts of firms with the average markups of the same cohort a years into the future, i.e. $Corr(\hat{\mu}_{0,t}, \hat{\mu}_{a,t+a})$, where ˆ indicates deviations from an Hodrick-Prescott trend taken *across* cohorts of the same age. *Aggregate* refers to autocorrelations of cost-weighted markup in the economy between t and $t+a$, i.e. $Corr(\hat{\mu}_t, \hat{\mu}_{t+a})$ where ˆ indicates deviations from an Hodrick-Prescott trend.

Figure 2.14 shows a comparison of these correlations for the data and the model. As business cycles persistently affect the cohort of newly listed firms along their product market, the autocorrelation of markups across cohorts - i.e. the correlation of the average markup by cohort with the average markup of the same cohort some years in the future - exhibits an higher persistence than the aggregate measure of markup, remaining higher than zero up to age five. This result indicates that cyclical variations in markups across cohorts persist into later years without mean-reverting. This particular feature is not shared by the measure of aggregate markups that does not show any persistent autocorrelation beyond one year, both in the model and in the data.[46]

2.6.2 Markups' co-movements with aggregate conditions

A well-known fact in models with customer markets where firms' sales affect the number of customers they can acquire is that markups are strongly countercyclical. The intuition hinges on the fact that, due to deep-habits and procyclical stochastic discount factors, firms value extra customers more in

[46]The cohort-level autocorrelation in the model can be made even stronger allowing for initial conditions in customer base accumulation to be correlated with the aggregate state of the economy.

Table 2.2: Markups' co-movements with the cycle

Dep.Variable: Log-Markup	Data	Model
Cycle Measure	-0.359***	-0.352
	(0.124)	
Cycle Measure × Age	0.051**	0.102
	(0.020)	
Cycle Measure × Age^2	-0.002**	-0.003
	(0.001)	
Controls	Yes	Yes
R^2	0.62	0.67
N	123,997	445,500

Note: Robust standard errors in parentheses, p :$^{***} < 0.01$,$^{**} < 0.05$,$^{*} < 0.1$. The table reports estimates of the co-movements between firm-level markups and a measure of business cycle. *Data* reports the coefficients of interest from estimating equation 2.3 on the main sample of firms followed up to 25 years; controls include firm fixed effects, sector shares, HHI index in three digit NAICS, cash holdings and log-employment; the measure of business cycle is quadratically detrended log-real GDP. *Model* shows the coefficients of running a regression as in equation 2.3 on a model simulated dataset; controls include age fixed effects, firm sizes and cohort fixed effects; the measure of business cycle is the log-deviation of output from the stationary solution value. The data are constructed by simulating the model for 2500 periods and then keeping firms up to 25 years to maintain consistency with the empirical counterpart.

booms than in recessions. Hence, when a recession hits, firms will find it profitable to exploit their current customer base rather than keep markups slightly lower to benefit from an higher customer base in the future. These incentives are at play also in the model presented in this paper with the addition that age is also an important force in determining the co-movements of markups with aggregate conditions. Table 2.2 reports the coefficients for a version of equation (2.3) estimated on model simulated data, showing that the model can replicate the age profile for the cyclicality of firm level markups.

A direct consequence of the fact that markups become less countercyclical as firms grow is that weighted and unweighted measures of aggregate markups in the model exhibit different cyclicality.[47] In particular, under the baseline calibration, cost-weighted measures of markups in the model features a 0.1 correlation with aggregate output, while the unweighted average shows a -0.14 correlation with aggregate output.[48] The root of this difference is once again linked to the nature of firm heterogeneity in the model. As firms age and grow bigger, their incentives to adjust markups in reaction to business cycles

[47] Note that unweighted measures of markups overrepresent the behavior of small firms relatively to their contribution to total output.

[48] As noted by Edmond *et al.* [2018] in a large class of models the correct model based measure of aggregate markups is a cost-weighted average rather than a sale-weighted average. This is true also in the model developed here as I show in Appendix 2.C

diminish as they get closer to the monopolistic competition limit and reach their optimal scale of operations. This intuition is also shown more formally in Proposition 2.

Moreover, the firms that are destined to grow in size and command a larger share of the overall economy are also those that operate in product markets with a large demand elasticity to customer base. Under the baseline calibration, these product markets are associated with relatively low sales relevance in the acquisition of new customers. This implies that the largest firms in the model do not respond strongly to business cycles not only because they are larger, hence closer to the monopolistic competition limit, but also because they operate in product markets where the price lever is not very effective in managing the size of their customer bases.

The age and product market heterogeneity present in the model, therefore, is important to generate dynamics of aggregate markups that speak both to the literature on firm-level estimates, that, as documented also in this paper, founds strong negative co-movements between markups and aggregate conditions and to the broader and older literature on markup cyclicality that instead estimates acyclical or even procyclical markups.[49]

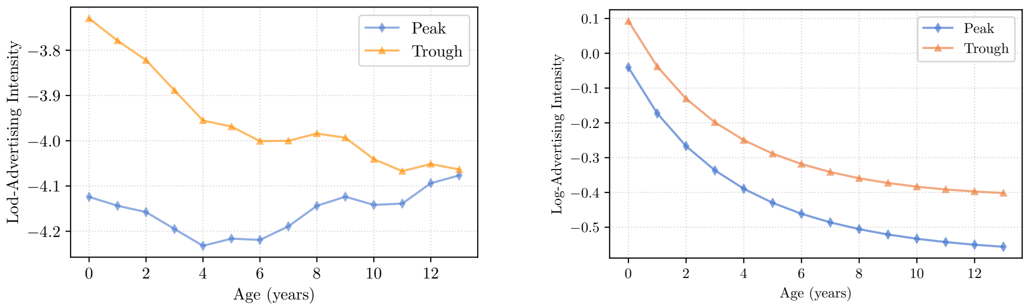
2.6.3 Testing the model's predictions using Advertising Expenditure.

The model presented in the paper is geared towards explaining the behavior of markups using the heterogeneity in customer accumulation motives that firms operating in different product market have. However, the fact that firms in the model can explicitly devote resources to the acquisition of customers allows them to compare the behavior of direct investment in demand with a similar empirical counterpart. To check if the model delivers sensible results on this margin, which is completely untargeted, I estimate the cohort effects on a measure of advertising intensity in a small sub-sample of firms that report their expenditure on advertising. I define advertising intensity as the ratio on advertising expenditure to total operating expenditure. As direct demand investments in the model are denoted in units of labor, the model counterpart for this measure is the ratio of firms' wage bill that directly depends on these direct investments, i.e. $\zeta(m_{j(i),t})$ over $[\zeta(m_{j(i),t}) + h_{j(i),t}]$. I then estimate the

[49] See among others, Nekarda and Ramey [2013].

Figure 2.15: Cohort effects and age profiles for intensity of advertising expenditure

(a) Data **(b)** Model



Note: The figure plots the age profile for the intensity of advertising expenditure (*Advertising expenditure*/*Operating Expenditure*) from equation (2.2) for the main sample and for a panel dataset constructed from simulating the model for 2500 periods. Specifically, at each age a, I am plotting $\hat{m}_0 + \hat{\phi}_a + (\hat{\beta}_0 + \hat{\beta}_1 a)Z$, where $\hat{m}_0$ is the average advertising expenditure in the first available year, $\hat{\phi}_a$ are the estimated age fixed effects, $(\hat{\beta}_0, \hat{\beta}_1)$ capture the estimated persistence of aggregate conditions in the first available year. Z is a measure of business cycle outcomes that takes value 2σ for *Peak* and -2σ for *Trough*, with σ being the standard deviation of listed firm output. In the model the measure of business cycles are deviations of output from the stationary solution.

effects of initial aggregate conditions on this measure using equation (2.2) in the data and in a model simulated dataset.

The comparisons of the age profiles is reported in Figure 2.15. Figure 2.15a shows the estimated age profiles for firms that are first observed in periods of above trend GDP in the data while Figure 2.15b replicates the exercise in the model. While the size of the initial effects of aggregate conditions is very similar, a recession of two-standard deviations in the model generates an increase in the intensity of advertising intensity that is approximately 90% of the one estimated in the data. However, the overall impact on the age profiles is very different, especially for firms that are first observed in booms.

The main reason for this detachment between the model prediction and the data is likely due to the very stark assumption in how the model deals with direct demand investments. As firms' age, they expand both their customer base and their production capabilities. As sales is a channel through which firms can accumulate customers, in the model, the need to divert resources from production to customer acquisition decreases mechanically as firms age. The additional mechanism that makes firms that start their operations more reliant on direct demand investments is the fact the these firms are self-selected in product markets where customers are less responsive to sales and more sensitive to the direct investments (low ψ product markets).

A potential explanation for the very different age profiles in the data instead can be linked to the fact that firms that start their operations in booms are facing also higher competition and therefore need to keep up spending on advertising as they age. These competition effects are not present in the model hence the product market choice and the aging structure of the economy fully determine the age profiles of direct demand investments.[50]

2.7 Conclusions

In this paper, I study how aggregate conditions at relevant junctures of firms lives can have persistent effects on firm choices and how these choices affects the behavior of both firm-level and aggregate markups.

I describe new empirical evidences on the effect of business cycles at the time of listing finding that firms that experience economics booms close to their listing date tend to start their activities as listed firms with lower initial markups and steeper age profiles.

In addition, I characterize the correlation of firm-level markups along the age profile, confirming recent existing evidence on the cyclicality of firm-level markups that finds a declining countercyclicality as firms grow. I then show that these empirical evidences are consistent with a model of firms dynamics that features heterogeneous product markets and realistic firms life cycles.

A novel feature of the model is that firms rely on both sales and direct demand investments to acquire customers and relax their demand constraints. Notably, product markets differ among each other along two main margins: i) how sensitive firms demands are to the size of customer base and ii) how relevant firms' sales are for the acquisition of new customers.

I discuss how varying the margins of heterogeneity across product markets influences the transmission of aggregate shocks and how different shock mixes can results in different correlations between a cost-weighted measure of aggregate markups and the simple average markup in the economy. As the cost-weighted measure is more affected by the behavior of incumbents, looking at the differences between the two measures is helpful to build some intuition

[50] Explicitly modeling the interaction of higher competition and product market characteristics with business cycles is left for future research.

of the role of the extensive and intensive margin for the cyclical behavior of aggregate markups in the economy.

I then show that business cycles have persistent effects on markups as they operate a selection along product markets on the cohorts of firms that start they operations in booms or in recessions. In particular, firms that get listed in periods of high aggregate demand, when it is relatively easier to attract new customers operate in product markets that require large investment in customer base and guarantee large long-term values at the cost of charging lower initial markups to build the necessary customer base.

The fact that sales are a necessary component for the acquisition of new customers makes the pricing choice of firm akin to an investment: lowering prices allows to increase sales and therefore enjoy a larger customer base in the future. As young firms are more likely to exit the market, they will also respond more strongly to business cycle matching the decline in the counter-ciclicality of markups along firms' age profiles and allowing to reconcile the differences in the measurement of cyclicality between aggregate and firm-level markups.

Appendices

2.A Additional tables and figures

2.A.1 A reduced form approach: autocorrelation of cohort-level markups

To estimate the autocorrelation of cohort level markups I am exploiting the following specification at the cohort level

$$\log(\mu_{a,t}) = \alpha + \beta_0 \log(\mu_{0,t-a}) + \beta_1 \log(\mu_{0,t-a}) \times a + \beta_2 \log(\mu_{0,t}) + \beta_3 a + \beta_4 a^2 + u_{a,t}, \tag{2.A.1}$$

where $\mu_{a,t}$ is the average markup of cohort a in year t; $\mu_{0,t-a}$ is the average markup of cohort a in the year of birth; $\mu_{0,t}$ is the average markup of entering firms in year t; and a is age. The elasticity of each cohort markups to their initial conditions is therefore given by β_0 and the elasticity at each subsequent age is $\beta_0 + \beta_1 \times a$. The comparison of the coefficients of interest for equation (2.A.1) estimated in the main sample and in a dataset of model generated data is reported in Table 2.A.1.

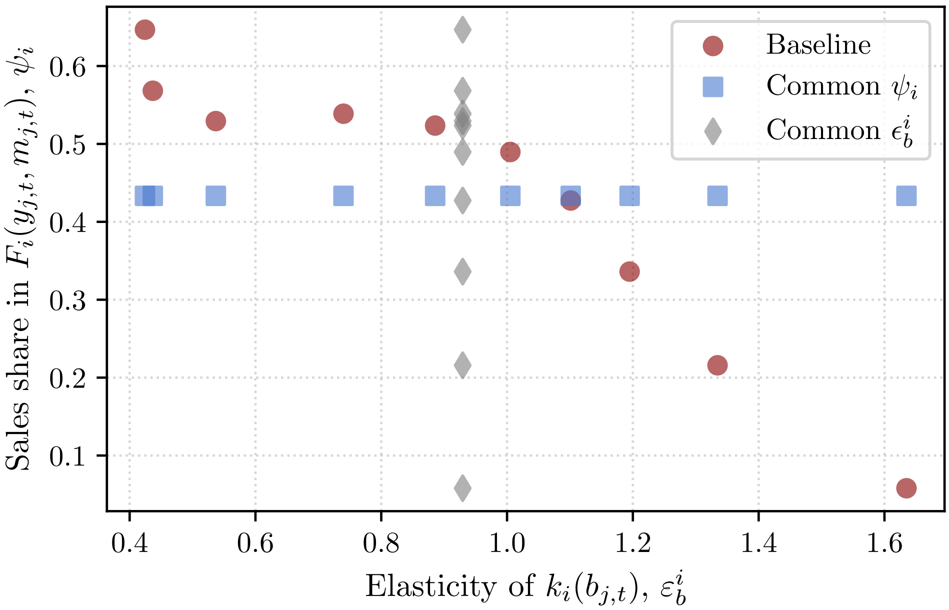
The model slightly overestimates the correlation with the markup charged in the initial year and the decline in the correlation induced by age. However, both coefficients are similar in magnitude and indicate that each cohort of firms in the data and in the model shares some common feature that only slowly fades away with time. In the model, this feature is the product market composition of different cohorts of firms.

Table 2.A.1: Auto correlation of firm-level markup

Dep.Variable: Log-Markup	Model	Data
Log-Markup at A_0	0.583	0.503***
		(0.072)
Log-Markup at A_0 × Age	-0.082	-0.022*
		(0.012)
Controls	Yes	Yes
R^2	0.63	0.34
N	370	370

Note: The table reports the elasticity of cohort-level markups with the markup charged in the first year a cohort is observed. In the data each cohort of firms starting from 1970 to 2006 is followed up to 10 years of age (37 cohorts followed for 10 years), robustness checks for 5 and 20 years give qualitatively similar results. For the model simulated data I run a simulation of 2500 period for each cohort and then I keep 37 cohorts of firms for 10 periods to have a comparable number of cohorts in the model simulation and in the data, increasing the number of simulated cohorts does not significantly change the results. Robustness checks for cohorts followed for 5 and 20 years deliver qualitatively similar results.

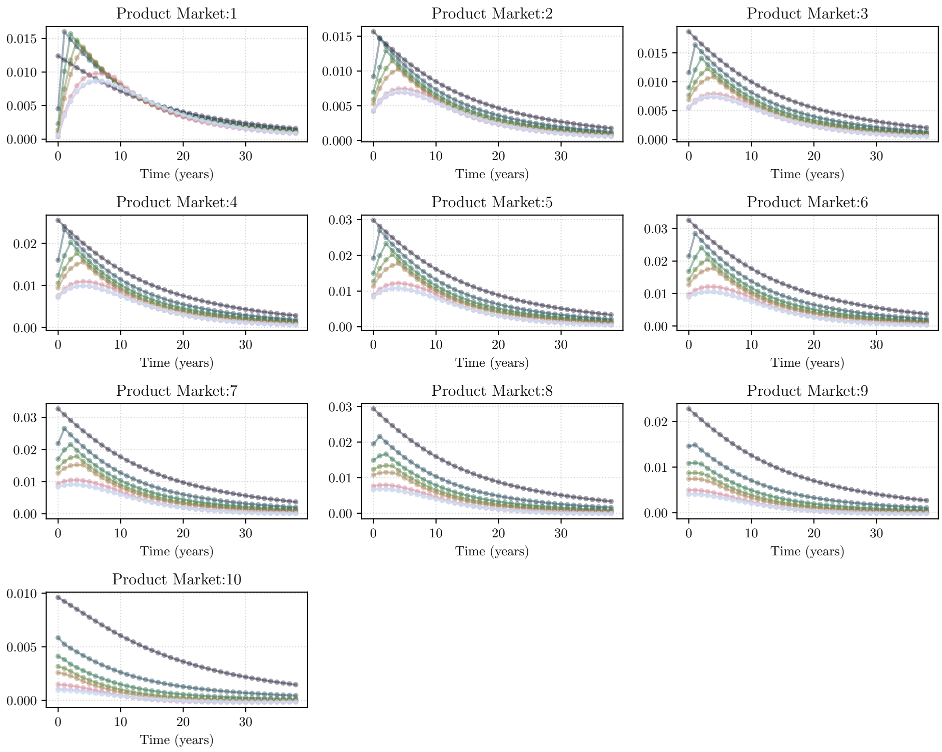
Figure 2.A.1: Product markets characteristics under different calibrations



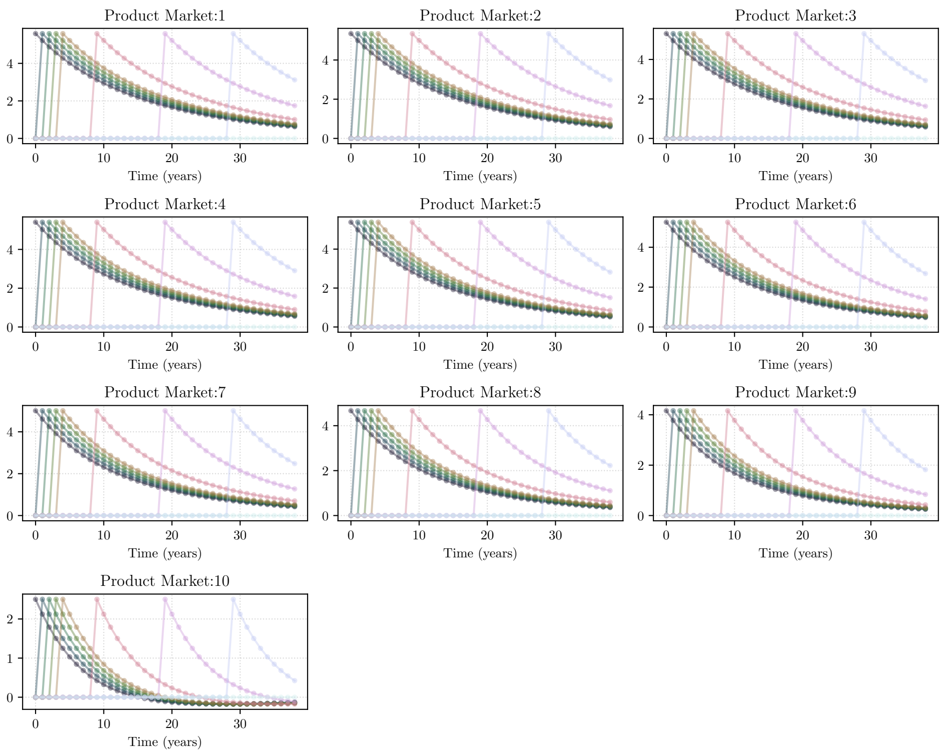
Note: The figure plots the parameters defining product market characteristics under the different calibrations discussed in Section 2.5.

Figure 2.A.2: Impulse response functions to aggregate productivity shocks for markups and mass of active firms by cohort and product market

(a) Markups



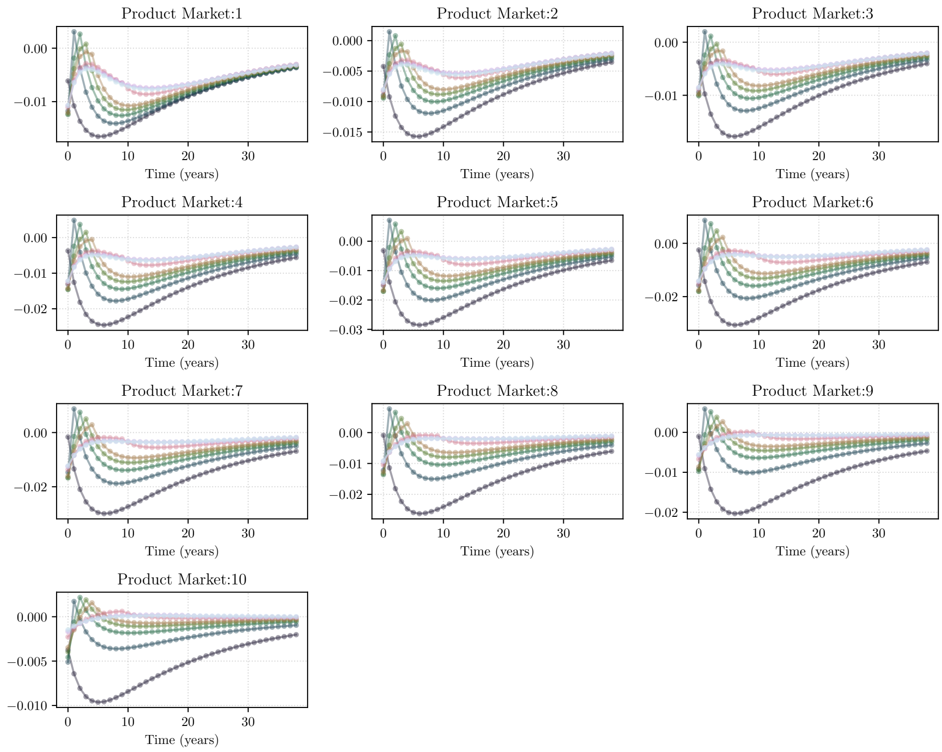
(b) Firm masses



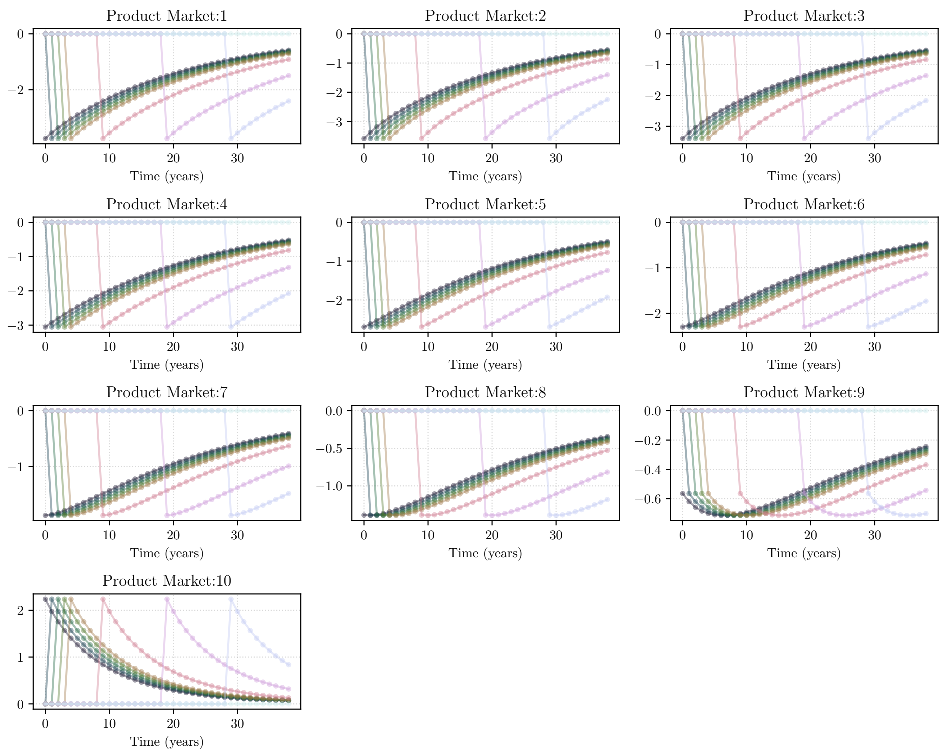
Note: The figure plots impulse response functions and percent deviations from the stationary solution to a one-percent productivity shocks. Each panel reports the IRFs for selected cohorts in a given product market. Darker colors are younger cohorts and selected cohorts are firms aged $a \in \{0, 1, 2, 3, 4, 5, 10, 20, 30, 40\}$. The time periods are years since the shock realization. Therefore, each dot shows the response of firms aged a years old t periods after the shock.

Figure 2.A.3: Impulse response functions to aggregate demand shocks of markups and masses of active firms by cohort and product market

(a) Markups



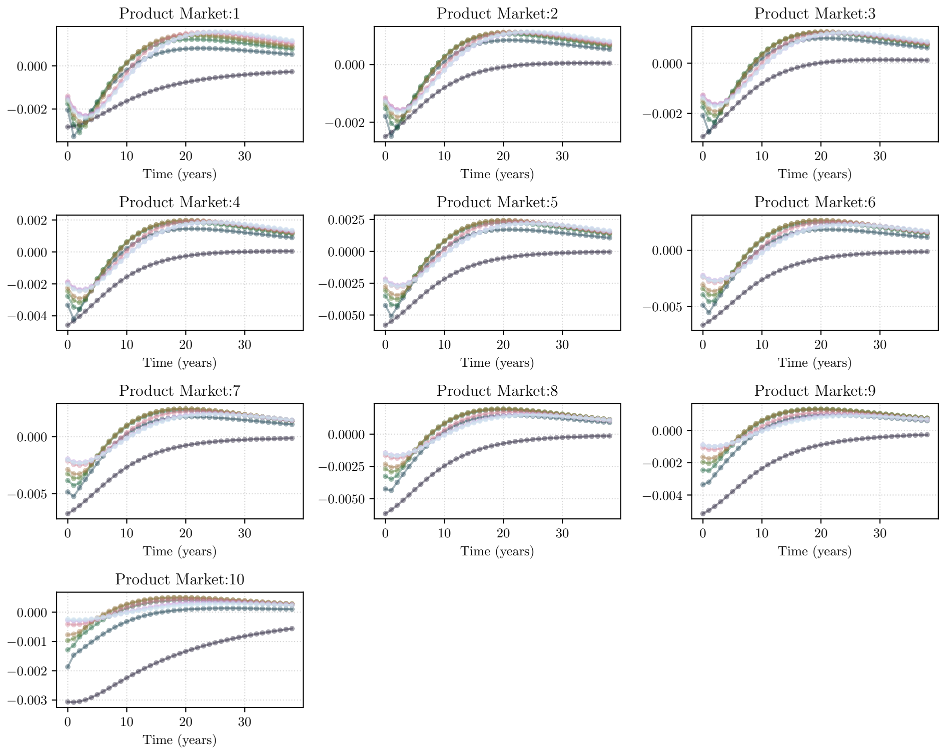
(b) Firm masses



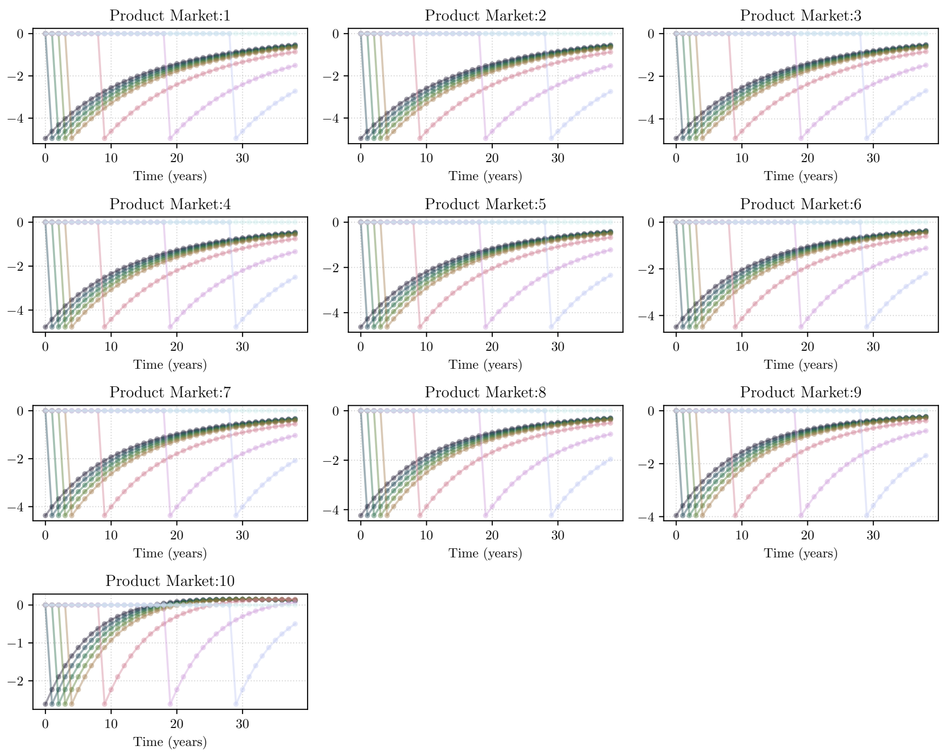
Note: The figure plots impulse response functions and percent deviations from the stationary solution to a one-percent aggregate demand shocks.Each panel reports the IRFs for selected cohorts in a given product market. Darker colors are younger cohorts and selected cohorts are firms aged $a \in \{0, 1, 2, 3, 4, 5, 10, 20, 30, 40\}$. The time periods are years since the shock realization. Therefore, each dot shows the response of firms aged a years old t periods after the shock.

Figure 2.A.4: Impulse response functions to listing cost shocks for markups and masses of active firms by cohort and product market

(a) Markups



(b) Firm masses



Note: The figure plots impulse response functions and percent deviations from the stationary solution to a one-percent listing cost shock. Each panel reports the IRFs for selected cohorts in a given product market. Darker colors are younger cohorts and selected cohorts are firms aged $a \in \{0, 1, 2, 3, 4, 5, 10, 20, 30, 40\}$. The time periods are years since the shock realization. Therefore, each dot shows the response of firms aged a years old t periods after the shock.

2.B Measuring markups in a model with customer base accumulation and dynamic pricing

The estimation strategy for firm-level markups outlined in the previous chapter relies on the assumption that pricing decisions are static. In other words, firms reset their prices every period without being subject to price rigidities (like menu costs or other adjustment costs).

However when firms' customer base accumulation is influenced by the level of their demand, even without price rigidities, their pricing choices are going to be dynamic as their current production contributes to the determination of their customer base and hence to the level of future demand.

To see how, consider a general setting where a firm j produces output using only labor and faces a demand function $D(p_{j,t}, b_{j,t})$ that depends on prices and the level of customer base b. Firms accumulate customer base through sales, so that customer base investments are a function of firm j's sales $F(y_{j,t})$.

The Lagrangian of a firm that has incentives to use prices to expand production and accumulate customer takes the following form:

$$\mathcal{L} = \max_{p_{j,t}, y_{j,t}, b_{j,t+1}} \mathbb{E}_0 \sum_{t=0}^{+\infty} \Lambda_{t,t+1} \left\{ \begin{array}{c} p_{j,t} y_{j,t} - P_t w_t h_{j,t} \\ -\lambda_{j,t}[y_{j,t} - f(h_{j,t})] \\ -\gamma_{j,t}[y_{j,t} - D(p_{j,t}, b_{j,t})] \\ -\phi_{j,t}[b_{j,t+1} - (1-\delta) b_{j,t} - F(y_{j,t})] \end{array} \right\} \tag{2.B.2}$$

where the λ constraint is the technology the firms uses, the γ constraint is the demand faced by firm j and the ϕ constraint is the law of motion of firm j's customer base. The optimal choice of the firm is described by the following first order conditions:

$$[p_{j,t}]: \quad y_{j,t} + \gamma_{j,t} D_p(\cdot) = 0 \tag{2.B.3}$$

$$[y_{j,t}]: \quad p_{j,t} - \lambda_{j,t} - \gamma_{j,t} + \phi_{j,t} F_y(\cdot) = 0 \tag{2.B.4}$$

$$[h_{j,t}]: \quad -P_t w_t + \lambda_{j,t} f_h(h_{j,t}) = 0 \tag{2.B.5}$$

$$[b_{j,t+1}]: \quad \phi_{j,t} = \mathbb{E}_t \left[\Lambda_{t,t+1}((1-\delta)\phi_{j,t+1} + \gamma_{j,t+1} D_b(\cdot|_{t+1})) \right] \tag{2.B.6}$$

Using equations (2.B.3), (2.B.4) and (2.B.5), keeping in mind that $\frac{D_p(\cdot) p_{j,y}}{y_{j,t}} \equiv \eta_{j,t}$ and $\frac{f_h(\cdot) h_{j,t}}{y_{j,t}} \equiv \theta_t^h$, I obtain the following condition that links firm-level markups and the expenditure share on variable inputs

$$\theta_t^h \left(\frac{P_t w_t h_{j,t}}{p_{j,t} y_{j,t}} \right)^{-1} = \frac{\eta_{j,t}}{\eta_{j,t} - 1} \left(1 - F_y(\cdot) \frac{\phi_{j,t}}{\lambda_{j,t}} \right). \tag{2.B.7}$$

The left hand side of equation (2.B.7) is the De Loecker and Warzynski [2012] measurement equation. Note that if current output does not play any role in creating customer base (i.e. $F_y(\cdot) = 0$) then this condition collapses back to the standard condition $\theta_t^h \left(\frac{P_t w_t h_{j,t}}{p_{j,t} y_{j,t}} \right)^{-1} = \frac{\eta_{j,t}}{\eta_{j,t} - 1}$. The willingness of firms to use prices to expand their customer bases however, introduces a negative bias to the standard measure of markups used in the literature. The size of the bias depends on the relative magnitude of the multiplier on the customer base accumulation process to the one on the technology used in production. In the model the size of the Lagrange multipliers depend on the state variable, hence we can think of the bias as being a function of firms' customer bases. I exploit this fact and I include a second order polynomial in firms' sector shares in three digits sectors to try to mitigate this problem as much as possible.

2.C Model and solution details

In this section I provide additional details about the model results and derivation discussed in the main body of the paper.

2.C.1 Rescaling of incumbent's value function

I solve the model with the habit-adjusted consumption good as the numeirare. The main consequence of this choice is that it is convenient to rescale the incumbent value function $V_i(b_{j,t-1}, a_{j,t}; S_t)$ to take the change into account. In particular, consider the original problem

$$V_i(b_{j,t-1}, a_{j,t}; S_t) = \max_{\substack{p_{j,t}, y_{j,t}, h_{j,t}, \\ m_{j,t}, b_{j,t}}} \left\{ \begin{array}{c} p_{j,t} y_{j,t} - w_t \left(h_{j,t} + \zeta(m_{j,t}) \right) + \\ +(1 - \rho(a_{j,t})) \mathbb{E}_t \left[q_t V_i(b_{j,t}, a_{j,t+1}, S_{t+1}) \right] \end{array} \right\} \tag{2.C.1}$$

subject to the relevant constraints

$$
\begin{aligned}
b_{j,t} &= (1-\delta)b_{j,t-1} + Q_t F_i(y_{j,t}, m_{j,t}) \\
y_{j,t} &= \left(\frac{p_{j,t}}{P_t}\right)^{-\eta} k_i(b_{j,t}) Y_t \\
y_{j,t} &= A_t h_{j,t}^{\alpha} \\
a_{j,t+1} &= a_{j,t} + 1.
\end{aligned}
$$

Now divide both left hand side and the right hand side of the Bellman equation in (2.C.1) by the aggregate habit-adjusted price index P_t and define $\tilde{V}(\cdot) \equiv V(\cdot)/P_t$.

Then, (2.C.1) becomes

$$
\tilde{V}_i(b_{j,t-1}, a_{j,t}; S_t) = \max_{\substack{p_{j,t}, y_{j,t}, h_{j,t}, \\ m_{j,t}, b_{j,t}}} \left\{ \begin{array}{c} \frac{p_{j,t}}{P_t} y_{j,t} - \frac{w_t}{P_t}\left(h_{j,t} + \zeta(m_{j,t})\right) + \\ +(1-\rho(a_{j,t}))\mathbb{E}_t\left[\frac{q_t}{P_t} V_i(b_{j,t}, a_{j,t+1}, S_{t+1})\right] \end{array} \right\}. \tag{2.C.2}
$$

From the household optimality conditions, however, we get that $q_t = \mathbb{E}_t\left[\beta \frac{u'(C_{t+1})}{u(C_t)} \frac{P_t}{P_{t+1}}\right]$, hence

$$
\tilde{V}_i(b_{j,t-1}, a_{j,t}; S_t) = \max_{\substack{p_{j,t}, y_{j,t}, h_{j,t}, \\ m_{j,t}, b_{j,t}}} \left\{ \begin{array}{c} \frac{p_{j,t}}{P_t} y_{j,t} - \frac{w_t}{P_t}\left(h_{j,t} + \zeta(m_{j,t})\right) + \\ +(1-\rho(a_{j,t}))\mathbb{E}_t\left[\beta \frac{u'(C_{t+1})}{u(C_t)} \tilde{V}_i(b_{j,t}, a_{j,t+1}, S_{t+1})\right] \end{array} \right\},
$$

so that, defining $\Lambda_{t,t+1} \equiv \beta \frac{u'(C_{t+1})}{u(C_t)}$ and substituting the price ratio and production technology from the constraints the incumbents' problem simplifies to

$$
\tilde{V}_i(b_{j,t-1}, a_{j,t}; S_t) = \max_{h_{j,t}, m_{j,t}, b_{j,t}} \left\{ \begin{array}{c} (k_i(b_{j,t})Y_t)^{\frac{1}{\eta}}(A_t h_{j,t}^{\alpha}) - \frac{w_t}{P_t}\left(h_{j,t} + \zeta(m_{j,t})\right) + \\ +(1-\rho(a_{j,t}))\mathbb{E}_t\left[\Lambda_{t,t+1} \tilde{V}_i(b_{j,t}, a_{j,t}+1, S_{t+1})\right] \end{array} \right\} \tag{2.C.3}
$$

subject to

$$
b_{j,t} = (1-\delta)b_{j,t-1} + Q_t F_i(A_t h_{j,t}^{\alpha}, m_{j,t}).
$$

2.C.2 Derivation of firm demands

For varieties produced by listed firms, household expenditure minimization requires that the agent solves the following problem:

$$
\max_{\{c_{j,t}\}_i} \left(\sum_i \int_{j \in \mathcal{J}_i} k_i(b_{j,t})^{\frac{1}{\eta}} c_{j,t}^{\frac{\eta-1}{\eta}} \, dj \right)^{\frac{\eta}{\eta-1}} - \lambda \left[\sum_i \int_{j \in \mathcal{J}_i} p_{j,t} c_{j,t} - E_t \right].
$$

Combining the optimal choices for each variety delivers the following household demand

$$c_{j,t} = \left(\frac{p_{j,t}}{P_t}\right)^{-\eta} k_i(b_{j,t})C_t,$$

where $P_t := \left(\sum_i \int_{j\in\mathcal{J}_i} k_i(b_{j,t})p_{j,t}^{1-\eta}dj\right)^{\frac{1}{1-\eta}}$. Now, note that entry requires X_t units of the habit-adjusted consumption good from listed firms , hence each entrant will demand a fraction

$$X_t\left(\frac{p_{j,t}}{P_t}\right)^{-\eta} k_i(b_{j,t})$$

of firm j's output. Therefore the total amount of firm j's output used for entry is:

$$x_{j,t} = \sum_{i=1}^{I} e_{i,t}X_t\left(\frac{p_{j,t}}{P_t}\right)^{-\eta} k_i(b_{j,t}),$$

then market clearing for each variety implies that

$$y_{j,t} = c_{j,t} + x_{j,t} = \left(\frac{p_{j,t}}{P_t}\right)^{-\eta} k_i(b_{j,t})\left[C_t + \sum_{i=1}^{I} e_{i,t}X_t\right].$$

As $C_t + \sum_{i=1}^{I} e_{i,t}X_t$ is the total expenditure in listed firms for this economy, then market clearing for the listed goods sector requires that

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

and therefore

$$y_{j,t} = \left(\frac{p_{j,t}}{P_t}\right)^{-\eta} k_i(b_{j,t})Y_t. \tag{2.C.4}$$

2.C.3 Equilibrium conditions

The incumbent problem for a firm in market i is

$$V_i(b_{j,t-1}, a; \boldsymbol{S}_t) = \max_{\left\{\substack{p_{j,t},y_{j,t},h_{j,t},\\ m_{j,t},b_{j,t}}\right\}} \left\{ \begin{array}{c} \frac{p_{j,t}}{P_t}y_{j,t} - w_t(h_{j,t} + \zeta(m_{j,t})) + \\ (1-\rho(a))\mathbb{E}[\Lambda_{t+1}V_i(b_{j,t}, a+1; \boldsymbol{S}_{t+1})] \end{array} \right\} \tag{2.C.5}$$

subject to

$$\begin{aligned} y_{j,t} &= \left(\tfrac{p_{j,t}}{P_t}\right)^{-\eta} k(b_{j,t})Y_t \\ b_{j,t} &= (1-\delta)b_{j,t-1} + Q_tF(y_{j,t}, m_{j,t})) \\ y_{j,t} &= A_t h_{j,t}^{\alpha} \end{aligned}$$

Using the individual firm demands to express the price ratio as a function of y and b, the technology to express y as a function of h then the incumbents problem can be simplified as follows[51]

$$V_i(b_{j,t-1}, a; \boldsymbol{S}_t) = \max_{\{h_{j,t}, m_{j,t}, b_{j,t}\}} \left\{ \begin{array}{l} \left(A_t h_{j,t}^{\alpha}\right)^{1-\frac{1}{\eta}} (k_i(b_{j,t})Y_t)^{\frac{1}{\eta}} - w_t(h_{j,t} + \zeta(m_{j,t})) \\ +(1-\rho(a))\mathbb{E}[\Lambda_{t+1} V_i(b_{j,t}, a+1; \boldsymbol{S}_{t+1})] \end{array} \right\} \tag{2.C.6}$$

subject to

$$b_{j,t} = (1-\delta)b_{j,t-1} + Q_t F(A_t h_{j,t}^{\alpha}, m_{j,t}))$$

The first order conditions for incumbents are:

$$\gamma_t = \tfrac{w_t \zeta'(m_{j,t})}{Q_t F_m(\cdot|_t)}, \tag{2.C.7}$$

$$\tfrac{w_t}{\alpha A_t h_{j,t}^{\alpha-1}} = \left(1 - \tfrac{1}{\eta}\right) \left(\tfrac{k_i(b_{j,t})Y_t)}{A_t h_{j,t}^{\alpha}}\right)^{\frac{1}{\eta}} + \tfrac{F_h(\cdot|_t)}{F_m(\cdot|_t)} \zeta'(m_{j,t}) w_t, \tag{2.C.8}$$

$$\tfrac{w_t \zeta'(m_{j,t})}{Q_t F_{m,t}(\cdot)} = \tfrac{1}{\eta} k_i'(b_{j,t}) Y_t \left(\tfrac{k_i(b_{j,t}Y_t)}{A_t h_{j,t}^{\alpha}}\right)^{\frac{1}{\eta}-1} + (1-\rho(a))(1-\delta)\mathbb{E}\left[\Lambda_{t,t+1} \tfrac{w_{t+1}\zeta'(m_{j,t+1})}{Q_{t+1} F_m(\cdot|_{t+1})}\right] \tag{2.C.9}$$

The same set of optimality conditions holds for all firms producing the same products and facing the same exit probability (i.e. firms in the same cohort), therefore the full equilibrium of the model can be described by the following conditions:

- For each cohort $a \in \{0, ..., \bar{a}\}$ and product type $i \in \{1, ..., I\}$:

$$\begin{aligned}
&\tfrac{w_t}{\alpha A_t h_{a,i,t}^{\alpha-1}} = \left(1 - \tfrac{1}{\eta}\right) \left(\tfrac{k_i(b_{a,i,t})Y_t)}{A_t h_{a,i,t}^{\alpha}}\right)^{\frac{1}{\eta}} + \tfrac{F_h(\cdot|_t)}{F_m(\cdot|_t)} \zeta'(m_{a,i,t}) w_t \\
&\tfrac{w_t \zeta'(m_{a,i,t})}{Q_t F_m(\cdot|_t)} = \tfrac{1}{\eta} k_i'(b_{a,i,t}) Y_t \left(\tfrac{k_i(b_{a,i,t})Y_t}{A_t h_{a,i,t}^{\alpha}}\right)^{\frac{1}{\eta}-1} + (1-\rho(a))(1-\delta)\mathbb{E}\left[\Lambda_{t,t+1} \tfrac{w_{t+1}\zeta'(m_{a+1,i,t+1})}{Q_{t+1} F_m(\cdot|_{t+1})}\right] \\
&b_{a,i,t} = (1-\delta) b_{a-1,i,t-1} + F(A_t h_{a,i,t}^{\alpha}, m_{a,i,t}) \\
&V_{a,i,t} = \left(A_t h_{a,i,t}^{\alpha}\right)^{1-\frac{1}{\eta}} (k_i(b_{a,i,t-1})Y_t)^{\frac{1}{\eta}} - w_t(h_{a,i,t} + \zeta(m_{a,i,t})) + (1-\rho(a))\mathbb{E}[\Lambda_{t+1} V_{a+1,i,t+1})] \\
&\Gamma_{a,i,t} = (1-\rho(a-1))\Gamma_{a-1,i,t-1} \\
&\tfrac{p_{i,a,t}}{P_t} = \left(\tfrac{k_i(b_{a,i,t})Y_t)}{A_t h_{a,i,t}^{\alpha}}\right)^{\frac{1}{\eta}}
\end{aligned}$$

  with initial conditions for customer bases in each product market $b_{-1,i,t-1} > 0$, calibrated to match sales profiles in the data. In practice, to simulate the model for each product market i, I draw idiosyncratic shocks $u_{i,t-1} \sim U(-1,1)$ and set $b_{0,i,t-1} = b_{0,i} \cdot (1 + u_{i,t-1})$.

[51] Recall that as I am using the rescaled version of the incumbent problem and I am solving everything in terms of the habit-adjusted price level, P_t, w_t is indicating the habit-adjusted real wage.

- Free entry in each sector $i \in \{1, ..., I\}$:

$$\frac{\Gamma_{0,i,t}}{e_{i,t}} = \frac{X_t}{V_{0,i,t}}$$
$$\Gamma_{0,i,t} = (\omega_i)^{\phi}\, e_{i,t}^{1-\phi}$$

  where ω_i are scaled by a common factor calibrated to ensure that aggregate output is equal to 1 in the stationary solution with no aggregate uncertainty.

- Household optimality condition, market clearing and exogenous processes:

$$\frac{w_t}{C_t} = \nu$$
$$\Lambda_{t+1,t} = \beta \frac{u'(C_{t+1})}{u'(C_t)}$$
$$H_t = \sum_{i=1}^{I} \sum_{a=0}^{\bar{a}} (h_{a,i,t} + \zeta(m_{a,i,t}))\Gamma_{a,i,t}$$
$$\sum_{i=1}^{I} \sum_{a=0}^{\bar{a}} \left(\frac{k_i(b_{a,i,t})Y_t}{A_t h_{a,i,t}^{\alpha}} \right)^{\frac{1}{\eta}} A_t h_{a,i,t}^{\alpha} \Gamma_{a,i,t} = Y_t$$
$$C_t + X_t \sum_{i=1}^{I} e_{i,t} = Y_t$$
$$A_t = \rho_a A_{t-1} + \epsilon_t^A$$
$$X_t = \rho_X X_{t-1} + \epsilon_t^X$$
$$Q_t = \rho_Q X_{t-1} + \epsilon_t^Q$$
$$\epsilon_t^j \sim N(0, \sigma_j), j = \{A, X, Q\}$$

2.C.4 Aggregation

As noted by Edmond *et al.* [2018], aggregate markups in a model economy like the one presented in this paper are equal to the inverse of the labor share.

Therefore, it is possible to define

$$\frac{1}{\mathcal{M}_t} = \frac{w_t H_t^g}{Y_t},$$

as the aggregate markup, where H_t^g denotes the amount of labor inputs dedicated to the production of the output good and not to the accumulation of customer base.

Firms' optimal behavior implies a firm-specific markup over marginal costs so that

$$\frac{1}{\mu_{j,t}} = \frac{P_t w_t h_{j,t}}{p_{j,t} y_{j,t}},$$

therefore

$$\frac{\frac{p_{j,t}}{P_t} y_{j,t}}{Y_t} = \mu_{j,t} \frac{h_{j,t}}{H_t^g} \frac{w_t H_t^g}{Y_t}. \tag{2.C.10}$$

As $Y_t = \sum_i \int_j \frac{p_{j,t}}{P_t} y_{j,t} dj$ then I can integrate both sides of the previous equation and, using the definition of aggregate markups for this economy, get

$$\mathcal{M}_t = \sum_i \int_j \mu_{j,t} \frac{h_{j,t}}{H_t^g} dj, \tag{2.C.11}$$

so that the aggregate markup is the cost-weighted average of firm-level markups.

Alternatively, it is possible to rewrite equation (2.C.10) as

$$\frac{\frac{p_{j,t}}{P_t} y_{j,t}}{Y_t} \frac{1}{\mu_{j,t}} = \frac{h_{j,t}}{H_t^g} \frac{1}{\mathcal{M}_t},$$

and, after integrating both sides over the distribution of active firms, get

$$\mathcal{M}_t = \left(\sum_i \int_j \frac{\frac{p_{j,t}}{P_t} y_{j,t}}{Y_t} \frac{1}{\mu_{j,t}} dj \right)^{-1}. \tag{2.C.12}$$

Therefore, consistently with Edmond *et al.* [2018], also in the model presented in this paper aggregate markups can be derived as the cost-weighted average of firm level markups or as the harmonic weighted average of firm markups with firms' sales shares over GDP as weights.

2.C.5 Proof of Proposition 2.

Proposition 2. The optimal markup management of listed firms can be summarized by the following condition:

$$\mu_{j,t}^{-1} - \bar{\mu}^{-1} = Q_t F_y(\cdot|_t) \frac{\varepsilon_b^i}{\eta} \frac{y_{j,t}}{b_{j,t}} \left[1 + \mathbb{E} \sum_{\tau=1}^{\infty} \prod_{s=0}^{\tau-1} \widetilde{\rho}(a_{j,t+s})(1-\delta)^\tau q_{t,t+\tau} \frac{p_{j,t+\tau}}{p_{j,t}} \frac{y_{j,t+\tau}}{y_{j,t}} \frac{b_{j,t}}{b_{j,t+\tau}} \right]$$

where $\widetilde{\rho}(a_{j,t}) = (1 - \rho(a_{j,t})$ is the surviving probability up to t, $\mu_{j,t} = \frac{w_t h_{j,t}}{\alpha y_{j,t}} \frac{P_t}{p_{j,t}}$ is the firm-level markup and $\bar{\mu}$ is the markup under standard monopolistic competition that would prevail without dynamic incentives in pricing.

Proof. From the first order conditions of an incumbent, equations (2.C.7),(2.C.8) and (2.C.9) we have that the Euler equation for the marginal value of customer base is

$$\gamma_t = \frac{1}{\eta} k_i'(b_{j,t}) Y_t \left(\frac{k_i(b_{j,t}) Y_t}{y_{j,t}} \right)^{\frac{1}{\eta}-1} + (1 - \rho(a_{j,t}))(1-\delta) \mathbb{E}\left[\Lambda_{t,t+1} \gamma_{t+1} \right].$$

given that $k_i(b_{j,t}) = \kappa_i b_{j,t}^{\varepsilon_b^i}$, we can multiply and divide by $k_i(b_{j,t})$ and $b_{j,t}$ to get

$$\gamma_t = \frac{\varepsilon_b^i}{\eta}\left(\frac{k_i(b_{j,t})Y_t}{y_{j,t}}\right)^{\frac{1}{\eta}}\frac{y_{j,t}}{b_{j,t}} + (1-\rho(a_{j,t}))(1-\delta)\mathbb{E}\left[\Lambda_{t,t+1}\gamma_{t+1}\right].$$

Denote the survival probability to the current period for a firm of age a as $\widetilde{\rho}(a_{j,t}) = (1-\rho(a_{j,t})$ and consider that $\left(\frac{k_i(b_{j,t})Y_t}{y_{j,t}}\right)^{\frac{1}{\eta}} = \frac{p_{j,t}}{P_t}$ from the demand constraint. Therefore iterating forward this equation for T period we get:

$$\gamma_t = \frac{\varepsilon_b^i}{\eta}\frac{y_{j,t}}{b_{j,t}}\frac{p_{j,t}}{P_t} + \mathbb{E}\sum_{\tau=1}^{T}\prod_{s=0}^{\tau-1}\widetilde{\rho}(a_{j,t+s})(1-\delta)^{\tau}\Lambda_{t,t+\tau}\frac{p_{j,t+\tau}}{P_{t+\tau}}\frac{\varepsilon_b^i}{\eta}\frac{y_{j,t+\tau}}{b_{j,t+\tau}} + \prod_{s=0}^{T}\widetilde{\rho}(a_{j,t+s})(1-\delta)^{T+1}\gamma_{t+T+1}\Lambda_{t,t+T+1},$$

taking the limit for $T \to \infty$:

$$\gamma_t = \frac{\varepsilon_b^i}{\eta}\frac{y_{j,t}}{b_{j,t}}\frac{p_{j,t}}{P_t} + \mathbb{E}\sum_{\tau=1}^{+\infty}\prod_{s=0}^{\tau-1}\widetilde{\rho}(a_{j,t+s})(1-\delta)^{\tau}\Lambda_{t,t+\tau}\frac{p_{j,t+\tau}}{P_{t+\tau}}\frac{\varepsilon_b^i}{\eta}\frac{y_{j,t+\tau}}{b_{j,t+\tau}} + \lim_{T\to\infty}\mathbb{E}\prod_{s=0}^{T}\widetilde{\rho}(a_{j,t+s})(1-\delta)^{T+1}\gamma_{t+T+1}\Lambda_{t,t+T+1}.$$

The term $\mathbb{E}\prod_{s=0}^{T}\widetilde{\rho}(a_{j,t+s})(1-\delta)^{T+1}\gamma_{t+T+1}\Lambda_{t,t+T+1}$ represents the discounted value, in utility terms, of an extra unit of customer base at infinity. Using an argument similar to the transversality condition in consumer problems, as profits are increasing in the size of customer base, at the optimal plan, an incumbent firm has to choose a path for its customer base such that this value is zero in the limit, thus

$$\gamma_t = \frac{\varepsilon_b^i}{\eta}\frac{y_{j,t}}{b_{j,t}}\frac{p_{j,t}}{P_t} + \mathbb{E}\sum_{\tau=1}^{+\infty}\prod_{s=0}^{\tau-1}\widetilde{\rho}(a_{j,t+s})(1-\delta)^{\tau}\Lambda_{t,t+\tau}\frac{p_{j,t+\tau}}{P_{t+\tau}}\frac{\varepsilon_b^i}{\eta}\frac{y_{j,t+\tau}}{b_{j,t+\tau}}.$$

Now, using equation (2.C.8), we can express the value of γ as deviations of firm-level markups from the monopolistic competition level. Using the demand constraint, we can rewrite equation (2.C.8) as follows:

$$\gamma_t Q_t F_y(\cdot) = \frac{p_{j,t}}{P_t}\left[\frac{P_t}{p_{j,t}}\frac{w_t}{\alpha A_t h_{j,t}^{\alpha-1}} - \left(1-\frac{1}{\eta}\right)\right],$$

and given that by definition $1-\frac{1}{\eta} = \bar{\mu}^{-1}$ and $\frac{P_t}{p_{j,t}}\frac{w_t}{\alpha A_t h_{j,t}^{\alpha-1}}$ is the inverse of the ratio of firm level prices to marginal costs, hence a measure of firm-level markups $\mu_{j,t}^{-1}$, we can rewrite the equation as

$$\frac{(\mu_{j,t}^{-1} - \bar{\mu}^{-1})}{Q_t F_y(\cdot)}\frac{p_{j,t}}{P_t} = \gamma_t.$$

Combining the definition of γ_t from the forward iteration with the expression above then implies that

$$\frac{(\mu_{j,t}^{-1} - \bar{\mu}^{-1})}{Q_t F_y(\cdot)} \frac{p_{j,t}}{P_t} = \frac{\varepsilon_b^i}{\eta} \frac{y_{j,t}}{b_{j,t}} \frac{p_{j,t}}{P_t} + \mathbb{E} \sum_{\tau=1}^{+\infty} \prod_{s=0}^{\tau-1} \widetilde{\rho}(a_{j,t+s})(1-\delta)^{\tau} \Lambda_{t,t+\tau} \frac{p_{j,t+\tau}}{P_{t+\tau}} \frac{\varepsilon_b^i}{\eta} \frac{y_{j,t+\tau}}{b_{j,t+\tau}},$$

which in turn can be rearranged as the condition in the proposition

$$\mu_{j,t}^{-1} - \bar{\mu}^{-1} = Q_t F_y(\cdot|_t) \frac{\varepsilon_b^i}{\eta} \frac{y_{j,t}}{b_{j,t}} \left[1 + \mathbb{E} \sum_{\tau=1}^{\infty} \prod_{s=0}^{\tau-1} \widetilde{\rho}(a_{j,t+s})(1-\delta)^{\tau} q_{t,t+\tau} \frac{p_{j,t+\tau}}{p_{j,t}} \frac{y_{j,t+\tau}}{y_{j,t}} \frac{b_{j,t}}{b_{j,t+\tau}} \right]$$

using the fact that $q_{t,\tau} = \mathbb{E} \Lambda_{t,t+\tau} \frac{P_t}{P_{t+\tau}}$ from household's optimality conditions.

□

Chapter 3

Leveraging on human capital: labor rigidities and sorting over the business cycle

Joint with Edoardo Maria Acabbi and Luca Mazzone

Abstract

This paper introduces a structural model of the labor market that features both worker and firm heterogeneity and where workers accumulate human capital and can search on the job. We analyze the optimal provision of insurance within the firm through an optimal dynamic contract, that, paired with the assumption of limited liability on the firm side, implies downward wage-rigidity. In our framework, insurance incentives and contractual rigidities are crucial in determining the pattern of job matches and separations along the business cycle. In particular, we show that aggregate fluctuations can alter the sorting between workers and firms by affecting workers' search strategies and, as a consequence, distort their human capital accumulation. In addition, we show that we can represent the optimal contracting problem using both the promised utilities framework and the more computationally feasible recursive Lagrangean approach.

3.1 Introduction

Business cycle fluctuations and the accumulation of human capital are strictly intertwined. Recessions have an impact on the matching process between workers and firms, altering the job ladder and thus accumulation of human capital on the job. At the same time, limits to investment in human capital can produce relevant feedback effects and delay recoveries from recessions.

This article proposes a theoretical framework to evaluate the effect of temporary aggregate fluctuations on workers' careers when human capital accumulation is linked to the quality of employer-employee matches. Notably, in our model, workers' human capital accumulation is linked to the quality of employers, giving an important role to workers' past careers for labor market outcomes. In addition, we highlight the role of frictions in shaping the worker search strategies and the worker-firm sorting along the business cycle. Specifically, we show that aggregate fluctuations have the ability to influence the search strategy of workers and consequently affect their human capital accumulation path. In addition, we show that limited liability on the firm side determines downward rigidity in wages and induces endogenous separations during recessions.

In order to generate these rich interaction between workers and firms, we develop a structural model of the labor market in which we nest a dynamic contract setting between risk-averse workers and risk-neutral firms. Our framework allows to characterize the interactions between workers human capital accumulation over the life cycle and the matching with heterogeneous firms. The model also characterizes the optimal amount of worker income insurance within the firm and its interaction with workers' human capital accumulation.

In the model, heterogeneous workers accumulate on the job experience which augments their skills and helps them climbing the job ladder. Search frictions and the presence of aggregate uncertainty distort sorting of workers with firms and generate inefficiencies in the economy. By studying the characteristics of the optimal contract between heterogeneous workers and firms we are able to draw some theoretical conclusions on how aggregate uncertainty influences both workers and firms' search strategies. In particular, we show that positive aggregate shocks are associated with an upgrade in the quality of the firms targeted by workers. As we assume that human capital accumulation depends on firms quality, our framework formalizes a potentially relevant mechanism through which aggregate conditions have persistent effects on workers careers by affecting the path of their human capital accumulation. In addition, the characterization of optimal contracts between heterogeneous workers and firms creates the scope for endogenous separations and therefore can be used to evaluate the impact of rigidities on the pattern of job destruction at the onset of recessions.

We populate the economy with overlapping generations of workers. Each cohort is exposed to different aggregate conditions at the start of the working career, a feature which determines a time-varying cross-sectional distribution of workers. The OLG structure allows us to identify the different sources of long-run changes in job sorting, wage growth and human capital accumulation affecting each age group.

Akin to physical capital, human capital can play a persistent role in affecting economic performance. An important difference between the two, however, is that the intensive and extensive margins of investment in, and acquisition of, human capital are likely much more limited given workers limited lifetime. This feature amplifies output fluctuations and keeps workers' productivity below potential for a period longer than the duration of a temporary negative TFP shock. Not only workers who enter the labor market in bad times, but also those who are displaced during recessions, face a worsened job ladder and trade worse employment prospects for a higher likelihood of exiting from unemployment. In this sense even a transitory shock, if intense enough or protracted enough, will generate a permanent loss in the human capital of the labor force that, on aggregate, is not going to be completely offset as long as the treated cohorts of workers are part of the workforce.[1]

We capture these rich business cycle effects as in Menzio and Shi [2010] and Menzio *et al.* [2016], by adopting a directed search framework in which we nest human capital accumulation. We also generate endogenous separations through endogenous wage rigidity, resulting in inefficient separations when a high wage induces negative continuation values for the firm. Rigidity results from an optimal insurance contract between firms and workers where

[1]The first study to advance this hypothesis is Ljungqvist and Sargent [1998], in which the *eurosclerosis* of the 1990s' is associated with an analysis of a rigid labor market with slow human capital accumulation by workers.

firms have limited liability, so they exit the contract when their continuation value is negative. Our framework has the advantage of explicitly modeling the dynamics of aggregate shocks and their effects on the entire earnings-skills distribution and clearly presenting the trade-off between insurance incentives, contractual rigidities and the efficient allocation of workers in the labor market. The model, thus, allows to assess how these trade-offs vary along business cycles and provides a framework to quantify what is the aggregate effect for an economy of having different cohorts of workers who experienced recessionary periods.

The analysis of how worker-firm relationships are shaped by business cycles has gained prominence in recent years, especially during the latest recessionary periods. Throughout the Global Financial Crisis (2007-2009) and the Sovereign Debt Crisis (2010-2012), it became evident that labor rigidities and missing investments in human capital could be potentially very harmful for workers who underwent periods of heightened instability and insecurity on the job, and in many cases were forced to accept under-qualified, precarious employment positions [OECD, 2014]. An assessment of aggregate dynamics of the labor market around these kinds of recessionary events is thus instrumental to inform the policy debate regarding the optimal level of flexibility of the market, and the possibility of enacting counter-cyclical policies, such as putting into place targeted unemployment benefits, targeted hiring subsidies, training programs or fiscal devaluations of labor costs to support unemployed cohorts and employment in recessions.

Relation to the literature. Our paper relates to strands of research in labor and macroeconomics analyzing the effects and costs that business cycles can have on workers' careers in the longer term.

From a theoretical standpoint, we contribute to the extensive literature analyzing long-term contracting with limited commitment (e.g. Harris and Holmstrom [1982], Thomas and Worral [1988], Krueger and Uhlig [2006], Xiaolan [2014], Lamadon [2016]). Contrary to these papers, however, our main focus is to quantify the aggregate effects of the interaction of insurance incentives, provided by the contract and crucial in determining the patterns of job separations along the business cycles, and the long-term term effects on the macroeconomy of having cohorts of workers exposed to worse employment relationship early in their careers.

Models of the labor market focusing on related topics have been recently proposed, among others, by Jarosch [2015] to analyze how workers value job security with respect to the salary, or by Burdett *et al.* [2016] to structurally estimate the cost of job loss. We complement these studies by explicitly modeling the contracting problem between the worker and the firm, taking into account the relevance of human capital accumulation for both the workers' careers and the aggregate economy.

In addition, we show that it is possible to describe the recursive structure of our contractual framework using both the formulation based on the inclusion of promised utilities as additional state variables, as in Spear and Srivastava [1987], that currently dominates the literature on optimal contracts between firms and workers, with the more computationally feasible formulation of recursive contracts developed by Marcet and Marimon [2019] [2].

At this time, and to the best of our knowledge, few studies have managed to incorporate the influence of the economic cycle together with the dynamics of human capital accumulation and firms-workers matches in a structural model of the labor market. Also, little is known about how the contractual framework influences firms decisions' on workforce mix, training programs and overall hiring strategies, especially in relationship to business cycle fluctuations. The search-and-matching literature has recently been trying to address how the sorting of workers and firms varies cyclically (see for instance [Lise and Robin, 2017]).

Structure of the paper. The paper is divided as follows: in Section 3.2 we briefly discuss some existing preliminary evidence regarding the dynamics of sorting, matching and the relevance of human capital accumulation along the business cycle; in Section 3.3 we present the model; in Section 3.4 we discuss features of the solution of the model and of the equilibrium conditions (proofs are in Appendix); finally, in Section 3.5 we conclude.

[2] The approach is known as the recursive Lagrangean method. Notable contributions, besides the seminal paper, include Cole and Kubler [2012], Messner *et al.* [2012], and Mele [2014]

3.2 Discussion of existing evidence

The model and the contractual environment developed in this paper imply a strong dependence of workers' careers on the history of aggregate shocks they are exposed to throughout their working lives. In addition, the fact that workers can accumulate human capital while working, paired with the assumption that working lives are finite, gives a disproportionate importance of aggregate shocks to matches early in workers lives. The intuition of why this is the case relies on noticing that early in workers lives human capital accumulation is more important both because the level of human capital workers are endowed with is lower and because the net present value of higher human capital levels in the future are greater the longer is the period a worker is able to reap the benefits of these higher levels. Both these reasons are stronger for younger workers searching for their first job than for older workers that are possibly trying to reallocate. As a consequence, any shock that induces missing investments or impairs the accumulation of human capital, especially early in workers careers, will generate a persistent loss in workers' earnings. This line of reasoning has numerous counterparts in the labor literature.

The existing empirical literature, in fact, has developed two main methodologies that are usually employed to measure the effects of recession on workers' career outcomes: (i) event studies and other kinds of reduced form empirical approaches, leveraging on an identification (necessarily in partial equilibrium) based on the observation of quasi-exogenous separation shocks or the possibility of matching workers in the data with very similar characteristics but different employment dynamics; and (ii) structural theoretical models, characterized by search frictions in job markets, which attempt to describe the effects of different kinds of shocks, both aggregate and idiosyncratic, on wage dynamics, distributions, and matching of workers to firms.

A number of recent studies have analyzed the long-run effects of unemployment for workers' earnings, and the importance of the economic cycle on lifetime outcomes. Many empirical studies, following the seminal paper by Jacobson *et al.* [1993], have analyzed the impact on earnings and work careers of losing a job during a recession finding large income penalties for workers that lose their job in bad economic times. A related literature, following in particular Kahn [2010], Oreopoulos *et al.* [2012], Schwandt and Von Wachter [2019], focused instead on the long term effects on graduates' and young workers' careers of entering the labor market in a recession, switching the analysis from matches that are already formed to the analysis of how aggregate conditions influence labor-market matches in the first place.[3] These studies use geographical variation in aggregate labor market conditions across cohorts of new graduates and they estimate a significant and persistent effect of entering the labor force under worse aggregate economic conditions.

In this line of research, and closely related in spirit to our work, Arellano-Bover [2020] estimates on Spanish data that the size of first employer can explain up to 15% of the scarring effects of recessions. In addition, he shows that the quality of initial matches between firms and workers have persistent effects on workers careers: when a worker is hired in a firm one standard deviation higher than the average, his lifetime income (20 years in the future) is estimated to be approximately one-third higher. In this case, firm size is taken as a proxy for firm specific attributes that can directly impact workers' careers, such as the availability of training programs or simply being exposed to better management practices, that are notably very difficult to measure accurately.[4] Similarly, Fernández-Kranz and Rodríguez-Planas [2017], Garcia-Cabo [2018] estimate the loss of an average recession in a range between 6-12% of annual earnings over 10 years, stronger and more persistent for less educated workers (7 years persistence for high-school graduates versus 5 years for college graduates). Other recent studies provide some evidence on the dynamic role of employment at heterogeneous firms also for job mobility and other labor market outcomes. Abowd *et al.* [2018], for example, show getting the first job in a top-paying firm can lead to up-ward movements in the earnings distribution later in workers careers. Bonhomme *et al.* [2019], instead, show how past firm types can have an impact on future earnings after changing jobs.

There is a large body of empirical evidence that estimates how the economic cycle can have a substantial impact on cohorts currently in the job market and especially on the ones entering the labor market for the first time in a recession. Even if the impact of this condition has already been analyzed as regards the impact on lifetime earnings by means of a reduced form approach, no analysis

[3]See also, among others, Schmieder *et al.* [2018], Lachowska *et al.* [2017], Altonji *et al.* [2016], Huckfeldt [2018].

[4]The model presented in this paper does not feature an explicit firm size distribution as we adopt the approach of one-job-one-firm, which is standard in the search and matching literature. As a consequence, we are silent on which are the specific channels through which firm quality influences human capital. The mechanism we have in mind, however, is akin to the effect of being exposed to better management practices and more efficient organizations.

has been carried out about the potentially persistent effects for the entire economy and for overall labor and firm productivity of having cohorts with a slower accumulation of human capital in the context of general equilibrium model of the labor market.[5]

Thanks to the computational tractability of the model developed in this paper, an interesting application that we leave to future research is the use of detailed matched employer-employee datasets to structurally estimate the effects of having cohort of workers exposed to negative aggregate conditions at the beginning of their careers and quantitatively asses the persistence of business cycles on the aggregate economy through labor market channels.

3.3 Model

In this section we present our model of the labor market. We start by discussing the environment, the timing and the preference structure of the economy. Then we discuss the features of a frictional labor market with directed search, and finally we characterize the workers problem and the optimal recursive contract.

3.3.1 Environment

Time is discrete, runs forever and is indexed by $t \in \mathbb{Z}$. The economy is populated by two kinds of agents: a unit measure of finitely-lived risk-averse households (workers) and a continuum of infinitely-lived, risk-neutral entrepreneurs who have the ability to invest in enterprises and thus run an endogenously chosen number of operating firms. All agents in the economy share the same discount factor $\beta \in (0, 1)$.

Following Menzio *et al.* [2016], we populate the economy with $T \geq 2$ overlapping generations of households, that face both aggregate and idiosyncratic risk. Each household lives for T periods deterministically, with age $\tau \in \mathcal{T} \equiv \{1, 2, 3, \ldots, T\}$. Every period workers participate to the labor market and, as in Shi [2009], Menzio and Shi [2010], direct their search towards different submarkets. Workers can only search for work and consume, as we do not

[5] Theoretical analyses of the impact of recessions on workers careers are presented in Audoly [2020], Guo [2018], Wee [2016]. Their models share some of the features of our model, but do not feature optimal dynamic contract and endogenous separations, or heterogeneous firms.

model saving decisions.[6] The objective of the household is to maximize its own life-time flow-utility from non durable consumption:

$$\mathbb{E}_{t_0}\left(\sum_{\tau=1}^{T}\beta^{\tau}u(c_{\tau,t_0+\tau})\right)$$

where t_0 characterizes the time of entry into the labor market and τ characterizes the age of the agent. Workers can either be employed or unemployed, and we denote by e and u their employment status. Workers are characterized by heterogeneous human capital levels h, with $h \in \mathcal{H} \equiv [\underline{h}, \overline{h}]$. Workers start off their life with a baseline level of human capital drawn from an initial exogenous continuous distribution with density $l(h)$ and can get training on-the-job over the course of their working career. The way in which they get training is another source of heterogeneity in the model across workers. Workers are matched with firms characterized by different levels of (permanent) firm quality $y \in \mathcal{Y} \equiv [\underline{y}, \overline{y}]$, which in our model are isomorphic to capital levels. This maps into the dynamics of human capital as explained below.

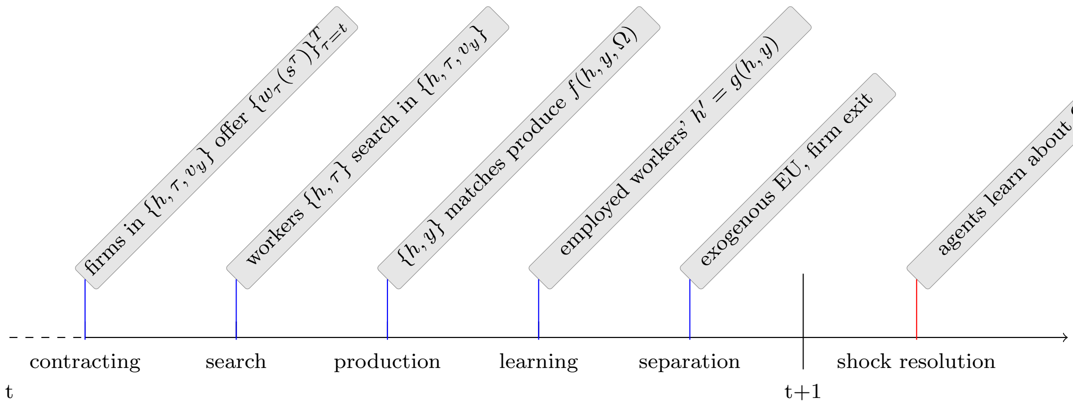
The only form of human capital accumulation in the model is on-the-job. Following Lise and Postel-Vinay [2019], we model human capital accumulation in this way: depending on the level of quality of the firm and their own level of ability, workers accumulate human capital according to some law of motion $g(h, y) : \mathcal{H} \times \mathcal{Y} \rightarrow \mathcal{H}$. In this setting, training is similar to "catching-up" of the firm quality with respect to the "training" ability of the firm, up to a point (depending on y) when the worker will not be able to learn anymore from the match and would possibly like to transit to a higher y match. At the same time, coherently with the concept of "mismatch", workers who lost their job and only manage to re-match with a low quality firm see their ability progressively deteriorating with the same $g(h, y)$ function. We assume that the function $g(h, y)$ is concave in both arguments.[7] Firms are, as common in labor-search studies, just one worker-one job matches, and we are thus abstracting from firm size.[8]

[6] Modeling wealth accumulation in a model with two-sided heterogeneity, life-cycle, human capital accumulation and directed search (also on-the-job) is undoubtedly interesting and important and left to future research.

[7] This is needed to ensure that the firm's optimal profit function is smooth and does not exhibit kinks as workers start to accumulate human capital.

[8] Modeling multi-worker firms in our context is an immensely interesting advancement that we leave for future research.

Figure 3.1: Timeline of worker-firm match



We denote future values in recursive expressions by adding a $'$ to them, or index elements by t in non-recursive ones. The aggregate state of the economy Ω is characterized by the level of aggregate productivity $a \in \mathcal{A} \subset \mathbf{R}^+$ and by the distribution of agents across states $\mu \in \mathcal{M} : \{e, u\} \times \mathcal{H} \times \mathcal{Y} \times \mathcal{T} \to [0, 1]$. Let $\Omega = (a, \mu) \in \mathcal{A} \times \mathcal{M}$ represent the aggregate state of the economy and let $\mathcal{M}$ represent the set of distributions μ over those states. Let $\mu' = \Phi(\Omega, a')$ be the law of motion of the distribution. Aggregate productivity evolves as a stationary Markov process, namely $a' \sim F(a'|a) : \mathcal{A} \to \mathcal{A}$.

The timing of each period is as follows: a productivity shock for the period is drawn; entrepreneurs open vacancies across the submarkets and post their offers; workers search from unemployment or on-the-job, and workers transition to a new job if on-the-job search is successful; production of both surviving and newly created firms takes place; employed workers accumulate human capital; an exogenous share of matches breaks down; at the same time, and before knowing what the next period productivity draw will be, incumbent firms decide whether to shut down, endogenously destroying their matches, or continue producing. State contingent policies prescribe an action for each realization of the story of worker-firm matches. For ease of notation, we denote the sequence of stories as $\{s^\tau\}_{\tau=1}^{T}$. The sequence of actions just described is summarized in Figure 3.1.

3.3.2 Labor markets

Search is directed. The labor market is organized as a continuum of submarkets indexed by the expected offers of lifetime utility $v_y \in \mathcal{V} \equiv [\underline{v}, \overline{v}]$. Each worker, characterized by an (h, τ) tuple of human capital and age, directs search while entrepreneurs decide which kind of firms y to open and, correspondingly, an offered lifetime value v_y.[9] There is free entry for entrepreneurs in submarkets. The process of opening a firm, which amounts to posting a vacancy at a quality-specific cost $\kappa(y)$, will be described in Section 3.3.6. We will also prove that, given a choice of worker (h, y) to whom an offer is made, there is going to be *only* one kind of firm y offering a defined value v_y. In other words, given (h, τ), v_y is an injective function $f_v : \mathcal{Y} \to \mathcal{V}$, and any vacancy in submarket (h, τ, v_y) is actually offered by the *same* y kind of firm.

The search process is characterized by a constant return to scale, twice continuously differentiable matching function $M(u, \nu)$ for each submarket, where the tightness of each submarket is as usual defined as $\theta = \frac{v}{\nu}$. Households job finding rates are defined as $p(\theta(h, \tau, v_y; \Omega)) = \frac{M(u(h,\tau,v_y;\Omega),\nu(h,\tau,v_y;\Omega))}{u(h,\tau,v_y;\Omega)}$, where $p() : \mathbf{R}^+ \to [0, 1]$ is twice continuously differentiable, strictly increasing and strictly concave function with $p(0) = 0$, $\lim\limits_{\theta \to +\infty} p(\theta) = 1$ and $p'(0) < \infty$, whereas the vacancy-filling is $q(\theta(h, \tau, v_y; \Omega)) = \frac{M(u(h,\tau,v_y;\Omega),\nu(h,\tau,v_y;\Omega))}{\nu(h,\tau,v_y;\Omega)}$, where $q()$: $\mathbf{R}^+ \to [0, +\infty]$ is twice continuously differentiable, strictly decreasing and strictly convex, with $q(0) = 1$, $\underset{\theta \to +\infty}{lim}\, q(\theta) = 0$ and $q'(0) < 0$. We have that $q(\theta) = p(\theta)/\theta$, and $p(q^{-1}(\cdot))$ is concave.

Upon match, workers produce next period according to a production function $f(a, h, y) : \mathcal{A} \times \mathcal{H} \times \mathcal{Y} \to \mathbf{R}^+$, increasing and concave in both arguments. The compensation of the worker depends on workers' and firms' kinds, and is defined by means of dynamic contracts through which firms deliver a promised utility. Contracts are going to be described in Section 3.3.5.

Matches are destroyed at an exogenous rate λ each period. Moreover, firms are subject to limited liability and matches separate endogenously either if the worker is poached by another firm (quit) or if the value of the match for the firm becomes negative (fires).[10] Workers are always allowed to search while unemployed and search while employed with probability λ_e. Notice that the timing of each period implies that newly separated workers can immediately search in the same period.

[9] As in Menzio and Shi [2010], given a menu of offers from any firm, workers are able to search in any submarket will separate by type in equilibrium, and any given type (h, τ) visits a particular market. For this reason submarkets can then be represented directly by (h, τ, v).

[10] Notice that separations happen in two waves during the same period. Fires and exogenous separations happen before the worker's search (at the end of previous period), whereas quits happen afterwards.

3.3.3 Informational and contractual structure

A contract defines a transfer of utility from the risk neutral firm to the risk averse worker with the match for all future possible histories of shocks. Given a match formed at a generic hiring time t_0, the state of the match is defined by $s_{t_0} = (h_{t_0}, \tau_{t_0}, a^{t_0}, \mu^{t_0}) \in \mathcal{S} = \mathcal{H} \times \mathcal{T} \times \Omega^{t_0}$, that is the worker skill, age and the history of aggregate productivity shocks and workers' distributions across employment states and submarkets (specifically, the specific worker's history of employment). We define $s^{t_0+(T-\tau_{t_0})}$ as the history of realizations between t_0, the time of hiring of the worker, and $t_0 + (T - \tau_{t_0})$, the time of maximum duration of the match with the worker before retirement (τ_{t_0} is the age at which the worker is hired and T is the retirement age).

The workers' history and the history of productivity are common knowledge, and histories are fully contractible. In this sense, the contract is fully state-contingent. Nevertheless, markets are incomplete as we assume that workers' actions are private knowledge in the search stage and firms are thus unable to counter outside offers. The contract offered by the firm can thus be defined as:

$$\mathcal{C} := (\mathbf{w}, \zeta) \text{ with } \mathbf{w} := \{w_t(s^{\tau_t - \tau_{t_0} + t_0})\}_{t=t_0}^{t_0+(T-\tau_{t_0})}, \text{ and } \zeta := \{v_t(s^{\tau_t - \tau_{t_0} + t_0})\}_{t=t_0}^{t_0+(T-\tau_{t_0})} \tag{3.1}$$

According to the contract the firm promises a series of state-contingent wages, to which the worker replies by enacting its own state-contingent search strategy, defined by the series of v_t sought at each node of the history.[11] As in Lamadon [2016], ζ is the action suggested by the contract, which in our analysis is bound to be incentive compatible for the worker. The contract is otherwise fully flexible in the degree to which the firm can determine wage levels and adjustment paths over the match histories.

3.3.4 Worker problems

Given the fact that the relationship of workers and firms is going to be characterized by a recursive contract with forward looking constraints, the state space of the problem needs to include the current utility promised to employed workers (or the current utility of unemployed agents), as in Spear and Srivas-

[11] Similarly to Menzio and Shi [2010], Tsuyuhara [2016], Lamadon [2016], and in order to guarantee, at least for the general proofs, that the problem is well behaved and the firm profit function is concave, the contract will require a randomization, a two-point lottery, which specifies probabilities over the actions prescribed.

tava [1987]. Job seekers, comprised of both unemployed agents and a share λ_e of incumbents, face similar search problems. Given a generic current lifetime utility V, any job seeker characterized by human capital h and age τ has to decide in which submarket to direct the search. Submarkets are indexed by the posted offered utility v_y. As it will be proved in Section 3.3.6, the choice over v will also indirectly determine which kind of firm y the worker matches with, and thus the human capital accumulation path. For now, assume that this mapping exists, and thus that, given (h, τ), the function $v_y(y)$ is an injective function $f_v : \mathcal{Y} \to \mathcal{V}$ such that any value v ends up being offered to the *same* kind of worker (h, τ) by one specific y firm only. This means that, even if workers only care about offered life-time utilities v, their choices determine which firm quality y they can match with and the human capital accumulation that concurs to deliver the promised utility v itself.

A worker characterized by (h, τ) who got the opportunity to search enters the search stage with lifetime utility $V + \max\{0, R(h, \tau, V; \Omega\}$, where the second component of the expression embeds the option value of the search, and R is the search value function. R is defined as:

$$R(h, \tau, V; \Omega) = \sup_{\{v_{y,\Omega}\}} \Big[p(\theta(h, \tau, v_{y,\Omega}; \Omega)) \big[v_{y,\Omega} - V \big] \Big] \tag{3.2}$$

We denote the solution of the search problem as $v_y^* = v^*(h, \tau, V; \Omega)$, and $p^*(h, \tau, v_{y,\Omega}^*; \Omega) = p(\theta(h, \tau, v_{y,\Omega}^*; \Omega))$ as the associated optimal job-finding probability. Notice that, given the timing of the choices outlined in Figure 3.1, a job seeker can devise search strategies that are *contingent* on the state in which the search actually takes place.

The lifetime utility of an unemployed worker at the beginning of the production stage can be define as

$$\begin{aligned} U(h, \tau; \Omega) = \ & u(b(h, \tau)) + \beta \mathbb{E}_\Omega \Big(U(h, \tau + 1; \Omega') \\ & + \max\{0, R(h, \tau + 1, U(h, \tau + 1; \Omega'); \Omega')\} \Big) \end{aligned} \tag{3.3}$$

where $b(h, \tau)$ is a (possibly) skill and age dependent unemployment benefit. Given finite workers' lives, $U(h, \tau; \Omega) = 0 \ \forall (h, \tau; \Omega) \in \mathcal{H} \times \mathcal{T} \times \mathcal{A} \times \mathcal{M}$ where $\tau > T$.

The corresponding lifetime utility of an employed worker with current promised utility $V_{y,\Omega}$ at the beginning of the production stage can be expressed as:

$$V_{y,\Omega} = u(w_y) + \beta \mathbb{E}_\Omega \bigg(\lambda U(g(h,y), \tau+1; \Omega') + (1-\lambda) \Big[V_{y,\Omega'} + \lambda_e \max\{0, R(g(h,y), \tau+1, V_{y,\Omega'}; \Omega')\} \Big] \bigg) \tag{3.4}$$

where w_y is the currently promised wage and $V_{y,\Omega'}$ is next period's state-contingent promised lifetime utility of remaining in the current firm, which becomes the "outside option" in the search problem. Notice that in the worker's problem there is nothing specific to firm quality y *per se*. The worker targets its search towards a desired level of promised utility, which we are going to show is going to be offered by one kind of y firm only in equilibrium. For this reason, we can index wages and utilities by y (and the aggregate state Ω, given the state-contingency of promises). The promised utilities V are an equilibrium object themselves, as they are the outcome of the firm optimization of the dynamic contract.

By means of their search strategy workers indirectly have an impact on their current contract too, as firms internalize workers' strategies in their optimization, and post wages and utility offers to maximize profits and thus retention. In fact, a worker quit drives profits to zero, independently of their previous level. Workers future promised utility incorporates both higher wages and higher option values of search, also through the human capital accumulation dynamics defined by $g(h,y)$.

The policy functions are uniquely defined, and allow to identify y uniquely as long as there exists a injective mapping between the offered utility v and y given $\{h, \tau, \Omega\}$, which we assume for now and prove given the structure of contracts in Section 3.3.6. Proofs for the uniqueness of the policy functions and the optimal policy are provided in Appendix 3.A.

The solution of employed workers' on-the-job search problem implicitly defines two "policy" functions, which incorporate workers' incentive compatibility which firms internalize in their optimization.

Definition 3.3.1 (Optimal retention probability and utility return)**.** *The solution of the worker's problem defines a retention function* $\widetilde{p} : \mathcal{H} \times \mathcal{T} \times \mathcal{V} \times \Omega \rightarrow$

$[(1-\lambda)(1-\lambda_e), 1-\lambda]$ *and a utility return* $\widetilde{r} : \mathcal{H} \times \mathcal{T} \times \mathcal{V} \times \Omega \to \mathcal{V}$*:*

$$\widetilde{p}(h, \tau, V_{y,\Omega}; \Omega) \equiv (1-\lambda)(1-\lambda_e p^*(h, \tau, v^*_{y,\Omega}; \Omega)) \tag{3.5}$$

$$\widetilde{r}(h, \tau, V_{y,\Omega}; \Omega) \equiv \lambda U(h, \tau; \Omega) + (1-\lambda)\Big[V_{y,\Omega} + \lambda_e \max\{0, R(h, \tau, V_{y,\Omega}; \Omega)\}\Big] \tag{3.6}$$

The two functions $\widetilde{r}$ and $\widetilde{p}$ incorporate the optimal behavior of the worker, and thus its incentive-compatible best replies when evaluating offers by firms. These functions are what the firms internalize while setting the optimal contract.

3.3.5 Contract

The contract is structured in such a way that the firm is subject to limited liability but commits to the delivery of a utility value to the worker, who on the other does not have to commit. Specifically, this means the worker is able to search at any period, and the firm is not able to counteract with another offer when its employee matches with another firm. The sequence of stories s^t is common knowledge and, while the firm cannot observe any of the actions of its workers, it has enough information to incorporate the worker's optimal policy decision.

Our choice of timing of exit decision is such that exiting firms know from the start of the period whether the productivity level is below the critical one a^* for the match (h, τ, y, W_y), and thus whether they will exit or not. In particular, given the current state, we can define the following indicator function:

Definition 3.3.2 (Exit policy)**.** *Incumbent firms make their exit decisions before the realization of aggregate productivity. The following indicator takes value one if the firm does not decide to exit in the following period:*

$$\eta_{t+1} = \begin{cases} 1 & \textit{if } a \geq \max\{0, a^*\} \\ 0 & \textit{otherwise} \end{cases}$$

with the productivity threshold defined as

$$a^*(h, \tau, y, W_{y,\Omega}) : \mathbb{E}_\Omega[J_{t+1}(h', \tau+1, y, W_{y,\Omega'}; a', \mu')|h, \tau, y, W_{y,\Omega}, a, \mu] = 0. \tag{3.7}$$

Given $\eta_{t+1} = 1$, the value function of a continuing incumbent in state $(h, \tau, W_{y,\Omega}; \Omega)$ can be rewritten recursively using the promised utilities to the workers as additional state variables as:

$$
\begin{aligned}
&J_t(h,\tau, y, W_{y,\Omega}; \Omega) = \\
&\qquad \sup_{\pi_i, w_i, \{W_{iy,\Omega'}\}} \sum_{i=1,2} \pi_i \Bigg(f(y, h; \Omega) - w_i \\
&\quad + \mathbb{E}_\Omega \Big[\widetilde{p}(g(h, y), \tau + 1, W_{iy,\Omega'}; \Omega')(J_{t+1}(g(h, y), \tau + 1, y, W_{iy,\Omega'}; \Omega') \Big] \Bigg) \qquad (3.8)
\end{aligned}
$$

$$
s.t.\ W_{y,\Omega} = \mathbb{E}_\Omega \left(u(w_i) + \mathbb{E}_\Omega \widetilde{r}(g(h, y), \tau + 1, W_{iy,\Omega'}; \Omega') \right), \qquad (3.9)
$$

$$
\sum_{i=1,2} \pi_i = 1 \qquad (3.10)
$$

where equation (3.9) is the promise keeping constraint ensuring that the current value of the contract is indeed based on the current wage and future utility promises with $\widetilde{r}_t()$ implicitly including the incentive constraint of the worker.

In this kind of contracts, the firm (principal) optimizes over its possible offers taking into account the utility of the worker (agent) and its incentive-compatible best replies. The resulting equilibrium is a subgame perfect Nash equilibrium of the kind identified in leader-follower sequential games, as in Von Stackelberg [1934]. The problem also resembles a Ramsey optimal policy problem, in that the principal in our case is akin to a policy-maker who maximizes aggregate utility according to some Pareto weights, taking into account optimization on the part of the agents in the economy (worker-firm match).

3.3.6 Vacancy opening and free entry

The economy is populated by a continuum of risk-neutral entrepreneurs. Each entrepreneur can invest to reach the desired level of firm quality y. The start-up costs of the firm are priced in terms of the consumption good and they consist of the posting a vacancy in the frictional labor market.

The cost of each vacancy is proportional to the quality of the firm being created, hence in order to post a vacancy for the creation of a firm with quality y the entrepreneur is forced to pay $c(y)$ in terms of the consumption good.

Thus, each entrepreneur at a generic time t, chooses in which submarket to post the vacancy selecting a lottery over the offered utility W_y, which maps into the set of firms' qualities $y \in \mathcal{Y}$, and worker characteristics $(h, \tau) \in \mathcal{H} \times \mathcal{T}$.

As the entrepreneur chooses the submarket in which to open a vacancy, he faces the following problem internalizing the optimal contract dynamics:

$$\Pi_t(h, \tau, y, W_{y,\Omega}; \Omega) = \sup_{y,h,\tau,W_{y,\Omega}} \quad - c(y) + q(\theta(h, \tau, W_y; \Omega))\beta[J_t(h, \tau, y, W_{y,\Omega}; \Omega)] \tag{3.11}$$

and, given perfect competition, free entry and the possibility for all entrepreneur to choose *any* possible firm kind y the profits from opening a vacancy should be driven down to 0 in submarkets which actually open:

$$\Pi_t(h, \tau, y, W_{y,\Omega}; \Omega) \leq 0 \;\text{ for } \forall\{h, \tau, y, W_{y,\Omega}; \Omega\} \in \{\mathcal{Y} \times \mathcal{V} \times \mathcal{S}\} \tag{3.12}$$

Assuming that $q(\cdot)$ is invertible, it delivers the equilibrium definition of marker tightness in each submarket:

$$\theta_t(h, \tau, W_{y,\Omega}; \Omega) = q^{-1}\left(\frac{c(y)}{\beta J(h, \tau, y, W_{y,\Omega}; \Omega)}\right). \tag{3.13}$$

3.3.7 Equilibrium definition

Recursive Equilibrium. Let $\Theta = \mathcal{A} \times \mathcal{M} \times \mathcal{H} \times \mathcal{T}$. A recursive equilibrium in this economy consists of a market tightness $\theta : \Theta \times \mathcal{V} \to \mathbb{R}_+$, a search value function $R : \Theta \times \mathcal{V} \to \mathbb{R}$, a search policy function $v^* : \Theta \times \mathcal{V} \to \mathcal{V}$, an unemployment value functions $U : \Theta \to \mathbb{R}$, a series of firm value functions, $\{J_t\}_{t=1}^T : \mathcal{S} \times \mathcal{V} \times \mathcal{Y} \to \mathbb{R}$, a series of contract policy functions $\{c_t\}_{t=1}^T : \mathcal{S} \times \mathcal{V} \times \mathcal{Y} \to \mathcal{C}$, a mapping between firm qualities and promised utilities $f_v : \mathcal{Y} \to \mathcal{V}$, an exit threshold for aggregate productivity $a^* : \mathcal{S} \times \mathcal{V} \times \mathcal{Y} \to \mathcal{A}$ and a law of motion for the aggregate state of the economy $\Phi_{\Omega,a'} : \mathcal{A} \times \mathcal{M} \to \mathcal{A} \times \mathcal{M}$ such that:

1. given the mapping f_v, market tightness satisfies equation (3.13)
2. the unemployment value functions solves equation (3.3)

3. the search value function solves the search problem in equation (3.2) and v^* is the associated policy function

4. the series of firm value functions and the associated contract policy functions are a solution to equation (3.8) for each $t \leq T$

5. the exit threshold satisfies equation (3.7)

6. the law of motion for the aggregate state of the economy respects the search and contract policy functions and the exogenous process of aggregate productivity

Definition 3.3.3 (Block Recursive Equilibrium). *A block recursive equilibrium is a recursive equilibrium such that the value and policy functions depend on the aggregate state only through aggregate productivity, $a \in \mathcal{A}$ and not through the distribution of agents across states $\mu \in \mathcal{M}$.*

We provide a proof for the existence of a BRE equilibrium in Appendix 3.D.

3.4 Discussion

In this section we briefly discuss the properties of the equilibrium of the model economy developed in the previous sections. All propositions and corresponding proofs are reported in Appendix 3.A and 3.B.

3.4.1 Workers optimal behavior

In the following proposition we summarize the main results regarding the behavior of the workers and their objective functions.

Proposition 3.4.1. *Given the worker search problem, the following properties hold:*

(i) The returns to search, $p(\theta(h, \tau, v_{y,\Omega}; \Omega))\big[v_{y,\Omega} - V\big]$, are strictly concave with respect to promised utility, $v_{y,\Omega}$.

(ii) The optimal search strategy

$$v^*(h, \tau, V; \Omega) \in \arg\max_{v_y} \big\{p(\theta(h, \tau, v_{y,\Omega}; \Omega))\big[v_{y,\Omega} - V\big]\big\}$$

is unique and weakly increasing in V.

(iii) For all promised utilities, the search gain $R(h, \tau, V; \Omega)$ is positive, weakly decreasing in V.

(iv) The survival probability of the match, given the optimal choice of the worker, is increasing in the value of promised utilities, so $\widetilde{p}_t(h, \tau, W_{y,\Omega}; \Omega)$ is increasing in $W_{y,\Omega}$.

Proof. See Proposition 3.A.1 in Appendix 3.A. □

The first statement implies that the marginal returns of searching towards better firms are decreasing. The intuition is that as workers search for work at better firms, their job-finding probability decreases as better employment prospects are also subject to higher competition.

As a consequence of the strict concavity established in the first statement, we can say that the optimal search strategy of each worker is increasing in the value of life-time utilities granted by the current contract. The intuition is that, as the search stage happens after the realization of the aggregate state and workers only care about the posted utility offers, given their type (h, τ), workers have a unique preferred option among utility offers.

The third statement follows from the fact that returns to search are decreasing and the set of utility promises is compact. The intuition is that employees at firms that promise an higher value of future utilities don't have a lot of incentives to search. To clarify this, imagine an individual working under a contract that guarantees her the best possible utility: there is no point in searching elsewhere as no other firms could match her current option. This leads directly to the finding that when firms offer better prospects to their workers, workers also have fewer incentives to leave. This guarantees a longer expected duration of the match, and generates retention probabilities that are increasing in promised utilities.

As human capital accumulation is tightly linked to the quality of the employer, workers that are able to start their working careers in good times have a higher chance of finding themselves on an higher path of human capital growth. As worker careers are limited and human capital accumulation follows a slow-moving process, business cycle effects are hard to fade and the quality of initial matches bears a long-standing effect on workers careers.

3.4.2 Characteristics of the optimal contract

The contracting problem between the worker and the firms allows us to analyze directly the trade-offs between the insurance provision and the optimal search behavior of the workers. The following proposition allows us to characterize the various incentives along the business cycle.

Proposition 3.4.2. *The Pareto frontier $J(h, \tau, y, W_{y,\Omega}; a, \mu)$ is increasing in the aggregate productivity shock a, while retention probabilities, $\widetilde{p}(h, \tau, W_{y,\Omega}; a, \mu)$ decrease in aggregate productivity.*

Proof. See Proposition 3.B.1 in Appendix 3.B. □

The intuition behind this proposition relies on the observation that higher productivity realization are associated not only with better outcomes on impact but also to better future prospects due to the monotonically increasing Markov process of productivity. In addition, the model allows us to characterize the optimal behavior of the workers along the business cycle. The following proposition summarizes how the search strategy changes depending on the realization of aggregate productivity.

Corollary 3.4.1. *The optimal search strategy of the workers is increasing in aggregate productivity.*

Proof. The claim follows directly from the fact that retention probabilities at the Pareto frontier, $\widetilde{p}$, are decreasing in a as discussed in Proposition 3.4.2. □

As discussed more in detail in Proposition 3.4.2, positive shocks to aggregate productivity will induce workers to search in submarkets that offer higher life-time utilities as firms optimally react to the positive shock increasing the number of vacancies posted. As firm promises and firm-qualities are linked by a one-to-one mapping, as discussed in Proposition 3.4.4, in booms workers manage to get matched with better firms.[12]

The combination of these two effects implies that, as firms became more productive, they are willing to post more vacancies in each submarket to increase the prospect of hiring in good times incorporating the higher benefits due to to

[12]In our model a better firm is an higher quality firm. We do not specifically model the determinant of quality heterogeneity but we take the existence of profound differences in firm quality as a reduced form meant to capture differences in management styles, organizational capital, technology, product qualities, availability of training programs and other firm level characteristics that can affect workers human capital accumulation.

the improvement in future prospects of aggregate productivity. The resulting higher tightness impacts workers' optimal search behavior as the job finding probability increases in all submarkets. As a consequence, workers respond optimally to the productivity increase searching in submarkets that guarantee higher life-time utility promises. Notice that as submarkets that guarantee higher lifetime utility are also populated by firms of better quality, see Proposition 3.4.4, this results shows how, in this framework, aggregate fluctuations modify workers' search incentives.

Firms are willing to commit to higher utility promises as the optimal contract guarantees that the provision of insurance ensures wage paths that can only partially follow productivity realizations. The following propositions provide a clear picture of the wage growth path prescribed by the optimal contract for a continuing firm. First, let us define the productivity threshold that determines whether a worker-firm match does not survive.

Corollary 3.4.2. *There exists a productivity threshold* $a^*(h, \tau, y, W_{y,\Omega})$ *below which firms will not continue the contract.*

The intuition of why this has to be the case is linked to the fact that the Pareto frontier is strictly increasing in a and decreasing in the level of promised utilities to the worker. Hence, once the aggregate state realizes, a firm is able to perfectly predict whether next period it will exit the market or stay in.[13] The choice is taken *before* new realizations of productivity, so the firm might for one period at most end up staying but make negative profits.

Proposition 3.4.3. *For each state in which the firm is willing to continue the contract, the optimal contract delivers a wage path that follows firms profits according to the wage Euler equation:*

$$\frac{\partial \widetilde{p}(\Theta)}{\partial W_{iy,\Omega'}} \frac{J_{t+1}(\Theta)}{\widetilde{p}(\Theta)} = \frac{1}{u'(w_{i,\Omega'})} - \frac{1}{u'(w_i)} \tag{3.14}$$

with $\Theta \equiv (g(h, y), \tau + 1, W_{iy,\Omega'}; \Omega')$ *being the definition of the relevant state and* $w_{i,\Omega'}$ *is the wage paid in the future state.*

Proof. See Proposition 3.B.4 in Appendix 3.B. □

[13]Notice that. given the timing, the decision is based on expected profits, and is thus *not* state-contingent to next period's productivity

The optimal contract links the wage growth to the realization of firms profits. The right hand side of equation (3.14) shows that, in providing insurance to the worker, the firm links wage growth to profits and to the incentive to maximize retention, incorporated in $\frac{\partial \log \widetilde{p}}{\partial W_y}$. As the production stage happens *after* exit choices are taken by the incumbent firms, the wage growth related to the continuation value of the contract is bound to be (weakly) positive, hence workers enjoy an non-decreasing wage profile under the optimal contract.[14]

A feature that the optimal contract derived in our model shares with the literature on long-term contracts with lack of commitment on the worker side is the backloading of wages. Workers in our model make search decisions that are going to affect the survival probability of the match. However, in doing so they do not appropriate the full future value of the match when making these decisions (unless the firm makes zero profits). This makes it optimal for the firm to front-load profits and back-load wages, as already among others noted by Tsuyuhara [2016], and Lamadon [2016]. The reason is that the firm provides insurance and income smoothing to the worker, but given its risk neutrality it prefers to front-load its profits and provide an increasing compensation path to maximize retention. The contract thus optimally balances the consumption smoothing motives (i.e. the insurance provision of the contract) with the commitment problem of the worker.

Notice that the wage Euler equation derived in equation (3.14) implicitly also pins down why, under the optimal contract, wages are downward rigid. To see this, consider that as firms have limited liability but commit to honor the contract as long as its value is weakly positive and that retention probabilities are increasing in promised utilities, i.e. $\frac{\partial \widetilde{p}(\Theta)}{\partial W_{iy,\Omega'}} \geq 0$, $\frac{J_{t+1}(\Theta)}{\widetilde{p}(\Theta)} \geq 0$, then the left-hand side of (3.14) is always weakly positive. This implies that the future wage prescribed by the optimal contract cannot be smaller than the current one, in order for the difference in the marginal utilities on the right-hand side of (3.14) to be positive.[15] The downward rigidity of wages, therefore, is a direct consequence of the incentive structure embedded in the optimal contract.

[14] As the exit decision takes place by considering *expected* profits next period, a firm might continue operating low but positive expected profits and end up, at most for a period, to have a negative continuation value. This would imply that wage growth *can* be negative before a firm's closure, which is actually a common finding in empirical studies (firstly observed in Ashenfelter [1978]).

[15] This last step follows from concavity of the utility function.

The next proposition, instead, confirms our initial conjecture that in equilibrium firm qualities and utility promises are related to a one-to-one mapping.

Proposition 3.4.4. *The mapping defined by the function $f_v : \mathcal{Y} \to \mathcal{V}$ is an injective function for each worker characteristic (h, τ).*

Proof. See Proposition 3.B.2 in Appendix 3.B. □

The intuition for this result is better expressed graphically in Figure 3.2. The figure shows the optimal choice of an entrepreneur that ex-ante has to choose whether to enter posting a utility promise W. Out of the equilibrium path, for a given θ, the optimal choice for the entrepreneur would be to pick the firm quality that guarantees the highest possible return to open a vacancy. Under free entry, however, the possibility of making positive profits however, attracts additional potential entrants. The increase in vacancy posting in the submarket drives the tightness up, lowering the ex-ante return to opening a vacancy to zero. As the firm value function is concave in y, in equilibrium, only one type of firm would be able to fend off the competition and remain active in the submarket. Finally, we provide the alternative recursive formulation for the contracting problem described in the paper. The saddle-point functional equation that can be, alternatively used, to define the recursive contract in equation (3.8) is expressed in the following proposition.

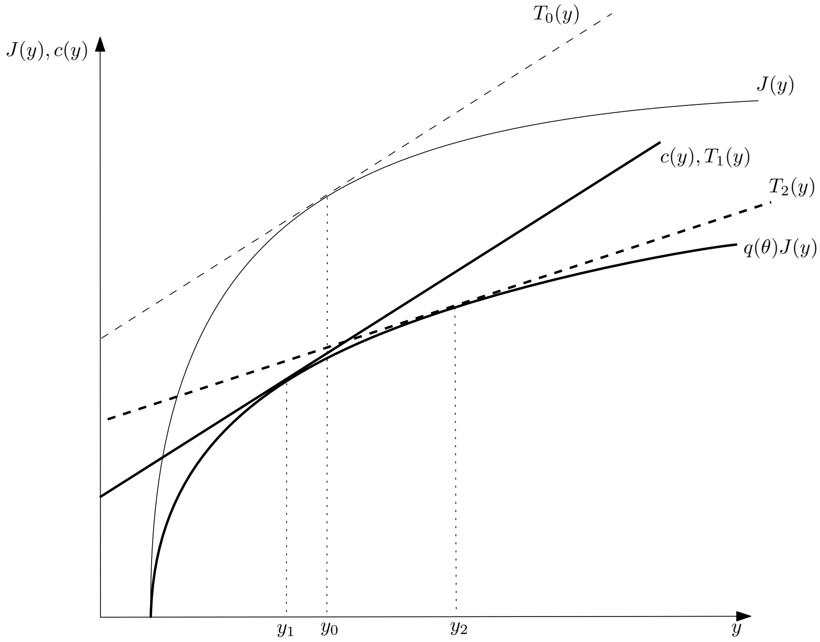
Proposition 3.4.5. *The solution to the contracting problem in equation* (3.8) *is the same as the solution to the following saddle-point functional equation:*

$$\begin{aligned}\mathcal{P}_t(h_t, \tau_t, y_t, a_t, \gamma_t) = & \underset{\gamma_t}{inf} \underset{w_t}{sup} (f(a_t, y_t, h_t) - w_t) + \mu_t^1 W_{y,t} - \gamma_t^1 (W_{y,t} - u(w_t)) + \\ & \beta \mathbb{E}_t(\lambda U_{t+1} + (1-\lambda)\lambda_e p_{t+1} v_{t+1}^*) + \beta \mathbb{E}_t \widetilde{p}_{t+1} \mathcal{P}_{t+1}(h_{t+1}, \tau_{t+1}, y_{t+1}, a_{t+1}, \gamma_{t+1})\end{aligned}$$

with $\mu_t = \gamma_{t_1}$ for some starting γ_0.

Proof. See Appendix 3.C for the details of the derivation of the SPFE following Marcet and Marimon [2019]. □

Figure 3.2: Optimal quality choice for entrepreneurs



Note: The figure shows the intuition behind the proof of Proposition 3.4.4. Given (h, τ, W_y, Ω), an entrepreneur, ex-ante, will ideally select the firm quality that guarantees the maximum possible profits. Free entry in vacancy posting, however, guarantees that, as long as positive profits are available in the submarket, entrepreneurs will continue posting vacancies driving down the vacancy filling probability, $q(\theta)$, and lowering the expected value of a match, progressively pushing firms out. In equilibrium, only entrepreneurs with one firm quality, in the figure's example y_1, will find it profitable to remain in the market.

3.5 Conclusion

In this paper, we develop a rich model of on-the-job search and human capital accumulation that features heterogeneity both on the worker and on the firm side.

In the model, ex-ante heterogeneous workers accumulate on-the-job experience which augments their skills and helps them climb the job ladder. Search frictions in the labor market and the presence of aggregate uncertainty prevent an efficient allocation of workers to firms and expose different cohort of workers to different human capital accumulation paths depending on the aggregate state at the time of entry in the labor force. As workers' lives are finite, investment in human capital early in workers lives are more valuable and initial losses on this front are difficult to recoup as they worsen future labor market prospects, amplifying the effects of transitory aggregate fluctuations and affecting the aggregate behavior of the economy.

In addition, we explicitly model workers-firms relationships by constructing a contractual framework that endogenously accounts for the different incentives between risk-averse workers and risk-neutral entrepreneurs. We characterize how insurance incentives are of paramount importance in shaping the response to business cycles of the labor market, the efficiency of workers-firms matches and the overall dynamic of human capital accumulation. Moreover, we show how, even in absence of institutional frictions, optimal contracts can endogenously generate rigidities in compensation. The intuition relies on the fact that employment relationships are subject to a one-sided limited commitment problem, where the firm can commit to a state-contingent wage path as long the value of the employment relationship is weakly positive, whereby workers can constantly search in the job for better matches. This framework is regulated by a dynamic contract that endogenously determines the optimal provision of insurance to workers. Within this framework, downward wage rigidity emerges as the optimal contract prescribes the firm to pay the worker a (almost) never decreasing compensation path.

The incentives underlying the optimal contract affect also workers behavior. In particular, we show that aggregate fluctuations have the ability to influence workers search decisions. We establish that workers that look for employment in bad economic times direct their search towards less productive firms, this limits their ability to accumulate human capital and imposes a drag on the overall labor productivity of the economy that persists as long as these cohorts of workers are active in the labor force.

Finally, we show that it is possible to represent the contracting problem between workers and firms using both the promised utilities framework and the more computationally feasible Lagrangean approach of Marcet and Marimon [2019]. For future research, we plan to leverage the computational feasibility of this framework and bring the model developed in this chapter to the data to structurally quantify the effects of aggregate fluctuations on worker-firm matches, their consequences for human capital accumulation, workers careers and the overall economic activity.

Appendices

3.A Properties of worker optimal behavior

The following propositions characterize the properties of workers' optimal search strategies that solve the search problem in (3.2), restated here for convenience:

$$R(h, \tau, V; \Omega) = \sup_{\{v_{y,\Omega}\}} \Big[p(\theta(h, \tau, v_{y,\Omega}; \Omega)) \big[v_{y,\Omega} - V \big] \Big]. \tag{3.A.1}$$

Lemma 3.A.1. *The composite function $p(\theta(h, \tau, v; \Omega))$ is strictly decreasing and strictly concave in v.*

Proof. For this proof we follow closely Menzio and Shi [2010], Lemma 4.1 (ii). From the properties of the matching function we know that $p(\theta)$ is increasing and concave in θ, while $q(\theta)$ is decreasing and convex. Consider that the equilibrium definition of $\theta(\cdot)$ is

$$\theta(h, \tau, v; \Omega) = q^{-1} \left(\frac{c(y)}{\beta J(h, \tau, y, v; \Omega)} \right),$$

and that the first order condition for the wage and the envelope condition on V of the optimal contract problem in (3.8) implies

$$\frac{\partial J(h, \tau, y, v; \Omega)}{\partial v} = -\frac{1}{u'(w)}.$$

so that as $u'(\cdot) > 0$, $J(\cdot)$ is decreasing in v.

From the equilibrium definition of $\theta(\cdot)$ and noting that $q^{-1}(\cdot)$ is also decreasing due to the properties of the matching function we have that

$$\frac{\partial \theta(h, \tau, v; \Omega)}{\partial v} = \frac{\partial q^{-1}(\xi)}{\partial \xi} \bigg|_{\xi = \frac{c(y)}{\beta J(h, \tau, y, v; \Omega)}} \cdot \left(- \frac{\partial J(h, \tau, y, v; \Omega)}{\partial v} \right) \cdot \frac{c(y)}{\beta (J(h, \tau, y, v; \Omega))^2} < 0,$$

which, in turn, implies that

$$\frac{\partial p(\theta(h, \tau, v; \Omega))}{\partial v} = \frac{\partial p(\theta)}{\partial \theta} \bigg|_{\theta = \theta(h, \tau, v; \Omega)} \cdot \frac{\partial \theta(h, \tau, v; \Omega)}{\partial v} < 0.$$

Suppressing dependence on the states (h, τ, y, Ω) for readability, to prove that $p(\theta(v))$ is concave, consider that $J(v)$ is concave[16] and a generic function $\frac{c}{v}$ is

[16] $J()$ concave give the two-point lottery in the structure of the contract. See Menzio and Shi [2010] Lemma F.1.

strictly convex in v. This implies that with $\alpha \in [0,1]$ and $v_1, v_2 \in \mathcal{V}$:

$$\frac{c}{J(\alpha v_1 + (1-\alpha)v_2)} \leq \frac{c}{\alpha J(v_1) + (1-\alpha)J(v_2)} < \alpha \frac{c}{J(v_1)} + (1-\alpha)\frac{c}{J(v_2)}.$$

As $p(q^{-1}(\cdot))$ is strictly decreasing the inequality implies that

$$\begin{aligned} p\left(q^{-1}\left(\tfrac{c}{J(\alpha v_1+(1-\alpha)v_2)}\right)\right) &\geq p\left(q^{-1}\left(\tfrac{c}{\alpha J(v_1)+(1-\alpha)J(v_2)}\right)\right) \\ &> \alpha p\left(q^{-1}\left(\tfrac{c}{J(v_1)}\right)\right) + (1-\alpha)p\left(q^{-1}\left(\tfrac{c}{J(v_2)}\right)\right), \end{aligned}$$

and as $\theta(v) = q^{-1}(\frac{c}{J(v)})$:

$$p(\theta(\alpha v_1 + (1-\alpha)v_2)) > \alpha p(\theta(v_1)) + (1-\alpha)p(\theta(v_2))$$

so that $p(\theta())$ is strictly concave in v. □

Proposition 3.A.1. *Given the worker search problem, the following properties hold:*

(i) The returns to search, $p(\theta(h,\tau,v_{y,\Omega};\Omega))\big[v_{y,\Omega} - V\big]$, are strictly concave with respect to promised utility, $v_{y,\Omega}$.

(ii) The optimal search strategy

$$v^*(h,\tau,V;\Omega) \in \arg\max_{v_y} \left\{p(\theta(h,\tau,v_{y,\Omega};\Omega))\big[v_{y,\Omega} - V\big]\right\}$$

is unique and weakly increasing (and Lipschitz continuous) in V.

(iii) For all promised utilities, the search gain $R(h,\tau,V;\Omega)$ is positive, weakly decreasing in V.

(iv) The survival probability of the match, given the optimal choice of the worker, is increasing in the value of promised utilities, so $\widetilde{p}_t(h,\tau,W_{y,\Omega};\Omega)$ is increasing (and Lipschitz continuous) in $W_{y,\Omega}$.

Proof. The proofs follow closely Shi [2009], Lemma 3.1 and Menzio and Shi [2010], Lemma 4.4. More formally, for each triplet (h,τ,Ω) given at each search stage, we can re-define the search objective function as $K(v,V) = p(\theta(v))(v-V)$ and $v^*(V) \in \arg\max_v K(v,V)$ as the function that maximises the search returns (i.e. the optimal search strategy of the worker) and prove the following

(i) To show that $K(v, V)$ is strictly concave in v consider two values for v, (v_1, v_2) such that $v_2 > v_1$ and define $v_\alpha = \alpha v_1 + (1-\alpha)v_2$ for $\alpha \in [0, 1]$.

Then by definition:

$$
\begin{aligned}
K(v_\alpha, V) &= p(\theta(v_\alpha))(v_\alpha - V) \\
&\geq [\alpha p(\theta(v_1)) + (1-\alpha)p(\theta(v_2))][\alpha(v_1 - V) + (1-\alpha)(v_2 - V)] \\
&= \alpha K(v_1, V) + (1-\alpha)K(v_2, V) + \alpha(1-\alpha)[(p(\theta(v_1)) - p(\theta(v_2))](v_2 - v_1) \\
&> \alpha K(v_1, V) + (1-\alpha)K(v_2, V)
\end{aligned}
$$

where the first inequality follows from the concavity of $p(\theta(\cdot))$ (this is true if $J()$ concave with respect to V) and the second inequality stems from the fact that $p(\theta(\cdot))$ is strictly decreasing hence $\alpha(1-\alpha)[(p(\theta(v_1)) - p(\theta(v_2))](v_2 - v_1) > 0$.

(ii) Given that $v \in [\underline{v}, \overline{v}]$, and submarkets are going to open depending on realizations of the aggregate productivity, a, there is only one region in the set of promised utilities where the search gain is positive, conditional on being in a job that pays lifetime utility V. That is $[V, v(a)]$ with $v(a)$ being the highest possible offer that a firm makes in the submarket for the worker (h, τ). As any submarket that promises higher than $v(a)$ is going to have zero tightness, the optimal search strategy for $V \geq v(a)$ is $v^*(V) = V$. For $V \in [V, v(a)]$, instead, as $K(v, V)$ is bounded and continuous, the solution $v^*(V)$ has to be internal and therefore respect the following first order condition

$$
V = v^*(V) + \frac{p(\theta(v^*(V)))}{p'(\theta(v^*(V)) \cdot \theta'(v^*(V))}. \tag{3.A.2}
$$

Now consider two arbitrary values V_1 and V_2, $V_1 < V_2 < \overline{v}$ and their associated solutions $W_i = v^*(V_i)$ for $i = 1, 2$.

Then, V_1 and V_2 have to generate two different values for the right-hand side of (3.A.2). Hence, $v^*(V_1) \cap v^*(V_2) = \emptyset$ when $V_1 \neq V_2$. This also implies that the search gain evaluated at the optimal search strategy is higher than the gain at any other arbitrary strategy so that $K(W_i, V_i) >$

$K(W_j, V_i)$ for $i \neq j$. This implies that

$$\begin{aligned}0 >& [K(W_2, V_1) - K(W_1, V_1)] + [K(W_1, V_2) - K(W_2, V_2)] \\ &= (p(\theta(W_2)) - p(\theta(W_1)))(V_2 - V_1).\end{aligned}$$

Thus, $p(\theta(W_2)) < p(\theta(W_1)))$ and as $p(\theta(\cdot))$ is strictly decreasing (Corollary 3.A.1) $v^*(V_1) < v^*(V_2)$. Uniqueness follows directly from strict concavity shown in (i). Lipschitz continuity still to show but coming from assumption of $J()$ being bi-Lipschitz continuous and $\theta(), p()$ being bounded functions.

(iii) The Bellman equation for the search problem is:

$$R(h, \tau, V; \Omega) = \sup_{\{v_{y,\Omega}\}} \Big[p(\theta(h, \tau, v_{y,\Omega}; \Omega)) \big[v_{y,\Omega} - V \big] \Big]$$

hence a simple envelope argument shows that

$$\frac{\partial R(h, \tau, V; \Omega)}{\partial V} = -p(\theta(h, \tau, v_y; \Omega)) \leq 0,$$

as the job finding probability is weakly positive for all utility promises.

As $p(\theta(\cdot)) \geq 0$, $v^*(\cdot) \in [\underline{v}, \overline{v}]$ then $R(\cdot) \geq 0$.

(iv) Given the optimal search strategy, $v^*(h, \tau, V; \Omega)$, we can define the survival probability of the match as in (3.5):

$$\widetilde{p}(h, \tau, V_{y,\Omega}; \Omega) \equiv (1 - \lambda)(1 - \lambda_e p(\theta(h, \tau, v^*_{y,\Omega}; \Omega))).$$

Then, given (h, τ, Ω)

$$\frac{\partial \widetilde{p}(V)}{\partial V} = -\beta(1 - \lambda)\lambda_e \left.\frac{\partial p(\theta)}{\partial \theta}\right|_{\theta=\theta(v^*)} \left.\frac{\partial \theta(v)}{\partial v}\right|_{v=v^*(V)} \frac{\partial v^*(V)}{\partial V} > 0,$$

because $p(\cdot)$ and $v^*(\cdot)$ are both increasing functions while $\theta(\cdot)$ is a decreasing function in promised utilities.

□

3.B Properties of the optimal contract

Lemma 3.B.1. *The Pareto frontier $J(h, \tau, y, W_{y,\Omega}; \Omega)$ is concave in $W_{y,\Omega}$.*

Proof. This is a direct consequence of using a two-point lottery for $\{w_i, W_{iy,\Omega'}\}$ as shown by Menzio and Shi [2010], Lemma F.1. $\square$

Lemma 3.B.2. *The Pareto frontier $J(h, \tau, y, W_{y,\Omega}; \Omega)$ is increasing in y.*

Proof. The intuition for this proof follows the fact that a higher y firm, once the match exists, can always deliver a certain promise V and have resources left over. Within a dynamic contract, future retention is already optimized as the match is formed. This means that the promise V can be delivered by the greater capacity on the part of producing with respect to a close y firm. In presence of human capital accumulation, the worker is compensated through greater option values in the future, which again means that, even with lower retention, the firm cashes in more profits while decreasing wages (and respecting the V promise). The reason why one does not have to worry about, for instance, variation in retention is that we are evaluating changes in y given the optimal contract, and given that by definition J is maximized, any indirect derivative of controls over y will get to their respective first order conditions and thus have no direct impact on the comparative static.

One can get to the same conclusion by starting from time T, noticing that the function J is trivially increasing in y in the last period, and the stepping back. At $T-1$, given V, any higher y function can make greater profits with the same delivery of value V, given the contract's optimal promise, which is a fortiori true with human capital accumulation (the option value is greater, so the firm can decrease w as a response). $\square$

Proposition 3.B.1. *The Pareto frontier $J(h, \tau, y, W_{y,\Omega}; a, \mu)$ is strictly increasing in the aggregate productivity shock a, while retention probabilities, $\widetilde{p}(h, \tau, y, W_{y,\Omega}; \Omega)$ decrease in aggregate productivity.*

Proof. For a generic period t, a firm matched to a worker in submarket $\{h, T-1, y, W_{y,\Omega}\}$ will face the following Pareto frontier

$$\begin{aligned} J_t(h, T-1, y, W_{y,\Omega}; a, \mu) = \sup_{w_i, \{W_{iy,\Omega'}\}} \Big(& f(y, h; \Omega) - w \\ & + \mathbb{E}_\Omega \left[\widetilde{p}(h', \tau+1, W_{y,\Omega'}; a', \mu')(f(y, h; a') - w') \right] \Big) \end{aligned}$$

The fact that period flows are increasing in a is immediate and follows from the properties of contracts with one-sided lack of commitment, as in Thomas and Worral [1988], Kocherlakota [1996] or Krueger and Uhlig [2006]. At the same time, following the logic of Lemma 3.B.2, the envelope condition on controls guarantees that one does not have to worry about the variation in optimal retention. This proves that J is increasing in a.

For the second part of the statemnt, notice that, in equilibrium,

$$\frac{\partial p(\theta)}{\partial a} = \underbrace{\frac{\partial p(\theta)}{\partial \theta}}_{>0} \cdot \underbrace{\frac{\partial \theta}{\partial J(\cdot)}}_{>0} \cdot \frac{\partial J(\cdot)}{\partial a}$$

where the sign of the second derivative on the right hand side comes from the free entry condition and the properties of vacancy filling probability function $q(\cdot)$. Given this, it has to be that $\frac{\partial p(\theta)}{\partial a}$ and $\frac{\partial J(\cdot)}{\partial a}$ have the same sign in equilibrium. This immediately implies that $\frac{\partial \widetilde{p}}{\partial a} < 0$ according to the optimal contract. □

Corollary 3.B.1. *There exists a productivity threshold $a^*(h, \tau, y, W_{y,\Omega})$ below which firms will not continue the contract.*

Proof. The proof follows immediately from Proposition 3.B.1 and the timing of the shock. Given the timing of the shock, exit is fully determined by the current productivity shock and incumbent firms know in advance whether they are willing to produce in the next period.

Therefore, as the Pareto frontier is strictly increasing in a, firms are willing to continue the contract if $\mathbb{E}_\Omega[J_{t+1}(h', \tau+1, y, W_{y,\Omega'}; a', \mu')|h, \tau, y, W_{y,\Omega}, a, \mu] \geq 0$, so that the threshold that determines exit is

$$a^*(h, \tau, y, W_{y,\Omega}) : \mathbb{E}_\Omega[J_{t+1}(h', \tau+1, y, W_{y,\Omega'}; a', \mu')|h, \tau, y, W_{y,\Omega}, a, \mu] = 0.$$

□

Corollary 3.B.2. *The productivity threshold $a^*(h, \tau, y, W_{y,\Omega})$ below which firm y in match with worker (h, τ) and given promised utility $W_{y,\Omega}$ in the aggregate state Ω is decreasing in y.*

Proof. Consider two firms characterized by y_1, y_2 with $y_1 < y_2$. Consider the threshold for firm y_1, $a_1^* = a^*(h, \tau, y_1, W_{y,\Omega})$. Firm y_1 makes 0 profits if state a_1^* materializes next period. Consider firm y_2 trying to mimic the current contract offered by y_1 to (h,τ). We know that J is increasing in y from Lemma 3.B.2, which implies that the firm is making a profit at a_1^*. This completes the proof. □

Lemma 3.B.3. *The Pareto frontier $J(h,\tau,y,W_{y,\Omega};\Omega)$ is strictly concave in y.*

Proof. The proof follows from the fact that the flow component of the profit function is always a concave function in y.

More formally, start from the last period T. The concavity is trivially given by the concavity of f. Now moving backwards to the problem at $\tau = T-1$, one can still consider the behavior of J given a promise $W_{y,\Omega}$. Again, given the option to search, the flow value is concave in y, retention probability is constant in $W_{y,\Omega}$, and the continuation value is a concave function. By induction, the statement holds for J at all $\tau \in [0,T]$.

□

Corollary 3.B.3. *As $J_t(h,\tau,y,W_{y,\Omega};\Omega)$ is concave, the tangent line at a generic $y_0 \in \mathcal{Y}$ is above the graph of $J_t(h,\tau,y,W_{y,\Omega};\Omega)$ so that*

$$J_t(h,\tau,y_0,W_{y,\Omega};\Omega) + \left.\frac{\partial J_t(h,\tau,y,W_{y,\Omega};\Omega)}{\partial y}\right|_{y=y_0} (y-y_0) \geq J_t(h,\tau,y,W_{y,\Omega};\Omega).$$

Proof. Dropping dependence on $(h,\tau,W_{y,\Omega};\Omega)$, consider two values for firm quality $y_0 < y_1$ both in $\mathcal{Y}$. Then, as $J_t(\cdot)$ is concave in y, taking $\alpha \in [0,1]$ the following inequalities are true:

$$\begin{aligned}
&J(\alpha y_0 + (1-\alpha)y_1) \geq \alpha J(y_0) + (1-\alpha)J(y_1)\\
\Rightarrow &J(\alpha y_0 + (1-\alpha)y_1) - J(y_0) \geq (1-\alpha)(J(y_1) - J(y_0))\\
\Rightarrow &\frac{J(\alpha y_0 + (1-\alpha)y_1) - J(y_0)}{\alpha y_0 + (1-\alpha)y_1 - y_0} \geq \frac{J(y_1) - J(y_0)}{y_1 - y_0}.
\end{aligned}$$

where the third inequality comes from noting that $y_1 > y_0$ and $\alpha y_0 + (1-\alpha)y_1 - y_0 = (1-\alpha)(y_1 - y_0)$.

Taking the limit for $\alpha \to 1$, we have that the left hand side tends to $\left.\frac{\partial J_t(y)}{\partial y}\right|_{y=y_0}$ and hence

$$J(y_0) + \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_0} (y_1 - y_0) \geq J(y_1). \tag{3.B.1}$$

Note that if $y_0 > y_1$ then $\frac{J(\alpha y_0+(1-\alpha)y_1)-J(y_0)}{\alpha y_0+(1-\alpha)y_1-y_0} \leq \frac{J(y_1)-J(y_0)}{y_1-y_0}$ but multiplying again the left hand side and the right hand side for $(y_1 - y_0) < 0$ still delivers (3.B.1). $\square$

Proposition 3.B.2. *Define the mapping between promised values and firm installed capital by the function $f_v : \mathcal{Y} \to \mathcal{V}$. Then f_v is an injective function for each couple of worker characteristics (h, τ).*

Proof. Note: throughout the proof we drop the dependence of the functions to the state (h, τ, Ω) to ease readability.

If the function f_v is an injective function then it defines a one-to-one mapping between $\mathcal{Y}$ and $\mathcal{V}$ so that for $(y_1, y_2) \in \mathcal{Y}$, and $f_v(y_1) = W_1$ and $f_v(y_2) = W_2$, $(W_1, W_2) \in \mathcal{V}$, $f_v(y_1) = f_v(y_2) \Rightarrow y_1 = y_2$.[17] We proceed by contradiction. To begin, assume that $f_v(y_1) = f_v(y_2)$ and $y_1 \neq y_2$.

As the optimal contract is a concave function in firm quality, we know that the tangents at each point are above the graph of the function. Thus, we can define the tangents at the two points y_1, y_2 as

$$T_1(y) \equiv J(y_1) + \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_1} (y - y_1) \quad \text{and} \quad T_2(y) \equiv J(y_2) + \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_2} (y - y_2).$$

Without loss of generality, consider the case in which $y_2 > y_1$. Knowing that $T_i(y) \geq J(y)$ for $i = 1, 2$ due to the concavity of $J(\cdot)$, we can define the following inequalities:

$$T_1(y_2) - J(y_2) \geq 0 \quad \text{and} \quad T_2(y_1) - J(y_1) \geq 0.$$

Using the definitions for the tangents at y_1 and y_2 they imply that

$$\frac{J(y_2) - J(y_1)}{y_2 - y_1} \leq \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_1} \quad \text{and} \quad \frac{J(y_2) - J(y_1)}{y_2 - y_1} \geq \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_2},$$

hence combining the inequalities we get that

$$\left.\frac{\partial J(y)}{\partial y}\right|_{y=y_2} \leq \frac{J(y_2) - J(y_1)}{y_2 - y_1} \leq \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_1}. \tag{3.B.2}$$

However, the free-entry condition in vacancy posting implies that in the sub-market (h, τ, W) both firms must be respecting $c(y_i) = q(\theta)\beta J(y_i)$ for $i = 1, 2$.

[17] As the contrapositive of Definition 2.2 in Rudin [1976], that defines a one-to-one mapping for $(x_1, x_2) \in A$ as $x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$.

As $c(y_i)$ is a linear function of firm quality $\frac{\partial c(y_i)}{\partial y_i} = c$ for $i = 1, 2$ and therefore from the free-entry condition:

$$c = q(\theta)\beta \left. \frac{\partial J(y)}{\partial y} \right|_{y=y_i}$$

which is a contradiction of the slopes of the two tangents being decreasing as shown in equation (3.B.2). Note that if $c(y)$ is convex and twice differentiable, then the derivatives of $c(y)$ are increasing in y while the derivatives of $J(\cdot)$ are decreasing leading again to a contradiction. The proof for the case in which $y_1 > y_2$ follows the same arguments and leads to a similar contradiction on the implied slopes of the optimal contract and those implied by the free entry condition. □

Lemma 3.B.4. *Given a state $(y, h, \tau, W_{y,\Omega})$ the optimal contract implies that*

$$-\frac{\partial J_t(h, \tau, y, W_{y,\Omega}; \Omega)}{\partial W_{y,\Omega}} = \frac{1}{u'(w)}$$

so that promised utilities and wages move in the same direction.

Proof. The proof follows directly from the envelope theorem and the concavity of the utility function $u(\cdot)$, as discussed in the proof of Proposition 3.B.4. □

Corollary 3.B.4. *The Pareto frontier $J(h, \tau, y, W_{y,\Omega}; \Omega)$ is decreasing in promised utilities $W_{y,\Omega}$.*

Proof. The envelope condition in Lemma 3.B.4 and note that $u'() \geq 0$. □

Proposition 3.B.3. *Assume $q(\theta(h, \tau, W_{y,\Omega}; \Omega))$ is not too convex. Then utility promises are unique and increasing in y, $\frac{\partial W}{\partial y} > 0$.*

Proof. Uniqueness follows directly from the concavity of the composite function.

The increasing property follows from the maximization of the entrepreneur in the free entry condition equation (3.12).

Assuming the same (h, τ, y), the entrepreneur has to choose which is the optimal value $W_{y,\Omega}$ to deliver in the contract. We know it is unique by assuming concavity of the composite function (which eventually amounts to assuming that the functional form of $q(\theta(W)))$ is not too convex in W.

For the rest of the proof we consider as given the dependence of the functions on (h, τ) and consider directly the composite function $q(\theta(W))$ as $q(W)$. The optimization involves a trade-off which respects the following first order condition:

$$q_W J(y, W) + q(W) J_W = 0 \tag{3.B.3}$$

For this to be a unique sup, the second order condition must be negative:

$$q_{WW} J + 2q_W J_W + q J_{WW} < 0 \tag{3.B.4}$$

where, as mentioned above, the only element which might lead to a violation is q_{WW} in case it is too convex ($J_{WW} < 0$) by 3.B.1. Notice this hypothesis amounts to assuming that $q(\theta(h, \tau, W_{y,\Omega}; \Omega))J(h, \tau, y, W_{y,\Omega}; \Omega)$ is concave.

By the implicit function theorem, the derivative of equation 3.B.3 is:

$$(q_{WW} J + 2q_W J_W + q J_{WW})W_y + q_W J_y + q J_{Wy} = 0 \tag{3.B.5}$$

The first term in parenthesis is negative, as second order condition. The second term is positive, given Lemma 3.B.2 and the fact that q_W is positive. The third term is 0, as the partial derivative of J in y does not contain V (which is the reason why Lemma 3.B.2 trivially holds). This means that, in order for the equality to be respected, $W_y > 0$.

$\square$

Proposition 3.B.4. *For each state in which the firm is willing to continue the contract, the optimal contract delivers a wage path that follows firms profits according to the wage Euler equation:*

$$\frac{\partial \widetilde{p}(\Theta)}{\partial W_{iy,\Omega'}} \frac{J_{t+1}(\Theta)}{\widetilde{p}(\Theta)} = \frac{1}{u'(w_{i,\Omega'})} - \frac{1}{u'(w_i)},$$

with $\Theta \equiv (g(h, y), \tau + 1, W_{iy,\Omega'}; \Omega')$ *being the definition of the relevant state.*

Proof. Consider the firm problem in equation (3.8), restated here for convenience

$$J_t(h,\tau,y,W_{y,\Omega};\Omega) = \sup_{\pi_i,w_i,\{W_{i,\Omega'}\}} \sum_{i=1,2} \pi_i \Big(f(y,h;\Omega) - w_i$$
$$+\mathbb{E}_{\Omega}\left[\widetilde{p}(h',\tau+1,W_{iy,\Omega'};\Omega')J_{t+1}(h',\tau+1,y,W_{i,\Omega'};\Omega')\right]\Big)$$

$$s.t.\ [\lambda]\ W_{y,\Omega} = \sum_{i=1,2} \pi_i \left(u(w_i) + \mathbb{E}_{\Omega}\widetilde{r}(h',\tau+1,W_{iy,\Omega'};\Omega')\right),$$
$$\sum_{i=1,2} \pi_i = 1, \quad h' = g(h,y).$$

For $i = 1,2$, the first order conditions with respect to the wage and the promised utilities are:

$$[w_i]:\ \lambda = \frac{1}{u'(w_i)} \tag{3.B.6}$$
$$[W_{iy,\Omega'}]:\ \pi_i \frac{\partial\widetilde{p}()}{\partial W_{iy,\Omega'}} J_{t+1}() + \widetilde{p}()\frac{\partial J_{t+1}()}{\partial W_{iy,\Omega'}} + \lambda\frac{\partial\widetilde{r}()}{\partial W_{iy,\Omega'}} = 0. \tag{3.B.7}$$

Note that by definition,

$$\widetilde{r}(h,\tau,V_{y,\Omega};\Omega) \equiv \lambda U(h,\tau;\Omega) + (1-\lambda)\Big[W_{y,\Omega} + \lambda_e \max\{0, R(h,\tau,V_{y,\Omega};\Omega)\}\Big]$$

therefore we can use the envelope theorem as in Benveniste and Scheinkman [1979], Theorem 1 and the definition in equation (3.5) to derive an expression for the derivative of the employment value in $t+1$ as the period ahead of the following:

$$\frac{\partial\widetilde{r}(h,\tau,W_{y,\Omega};\Omega)}{\partial W_{y,\Omega}} = \widetilde{p}(h,\tau,W_{y,\Omega};\Omega).$$

Similarly, using the envelope condition on the firm problem and the first order condition for the wage, we can establish that

$$\frac{\partial J_t(h,\tau,y,W_{y,\Omega};\Omega))}{\partial W_{y,\Omega}} = -\lambda \ \therefore\ \frac{\partial J_t(h,\tau,y,W_{y,\Omega};\Omega))}{\partial W_{y,\Omega}} = -\frac{1}{u'(w_i)}. \tag{3.B.8}$$

Moving these two expressions one period ahead, substituting them in (3.B.7), taking $\pi_i > 0$ and rearranging we have that:

$$\frac{\partial\widetilde{p}(\Theta)}{\partial W_{y,\Omega'}}\frac{J_{t+1}(\Theta)}{\widetilde{p}(\Theta)} = \frac{1}{u'(w_{\Omega'})} - \frac{1}{u'(w)},$$

with $\Theta \equiv (g(h,y), \tau+1, W_{y,\Omega'}; \Omega')$ and where $w_{\Omega'}$ is the wage next period in state Ω'. □

3.C Derivation of recursive contract SPFE

Solving the optimal contract and the overall model given the recursive structure obtained by following the promised utility method of Spear and Srivastava [1987] is computationally infeasible. This is due to the fact that the optimal contract requires to define a valid recursive domain and codomain of promised values that respects all the future forward looking constraints. Known solution methods for these kinds of models [Abreu *et al.*, 1990], although robust, easily become computationally unmanageable as the number of states of the model increases. We thus follow Marcet and Marimon [2019] in deriving a recursive expression for the optimal contract in which the Lagrange multiplier for the promise keeping constraint equation 3.B.8 is added as a co-state of the model, and allows us to circumvent the problem of searching for valid promised values domains altogether.

The reason why the recursive contracts method in Marcet and Marimon [2019] simplifies our problem is simple. As shown in equation 3.B.8, wage growth and levels in any next period and at every node are determined by the state-contingent multiplier on tomorrow's promise keeping constraints. This considerably reduces the complexity of the problem, as by definition Lagrange multipliers are defined over $\mathbb{R}^+$.

We follow Marcet and Marimon [2019] (hereby MM) and their terminology to define how a recursive saddle point functional equation (SPFE) can be obtained from the sequential formulation of the problem. For the present exposition of the constructive method to obtain the SPFE, for simplicity and without loos of generality, we ignore the randomization of the contract over the lotteries and the limited liability constraint. The latter choice, in particular, does not create any problem in terms of thinking about of developing the sequential problem over time: our choice of timing of exit decision is such as that exiting firms know form the start of their period whether the productivity level is below the critical one $a^*_{h,\tau,y,W}$ for the match (h, τ, y, W_y), and thus whether they will exit or not. The lack of uncertainty and optimization over the next periods makes the problem of these firms, at some low states, equivalent to the problem of a firm with a lower maximum length (which is

T, the retirement age, in general). At an exiting state t the firm knows *with certainty* that any $J_j = 0$ for $j > t$, match with a worker of age T.

Consider the problem

$$J_t(h_t, \tau_t, y_t, W_{y_t}, a_t) = \sup_{w_t, \{W_{y,s^{t+1}}\}} \Big(f(a_t, y_t, h_t) - w_t$$
$$+ \mathbb{E}_{s^t} \left[\widetilde{p}(h_{t+1}, \tau_{t+1}, W_{y,s^{t+1}}, a_{s^{t+1}})(J_{t+1}(h_{t+1}, \tau_{t+1}, y_t + 1, W_{y,s^{t+1}}, a_{s^{t+1}})\right] \Big) \tag{3.C.1}$$

$$s.t.\ W_t = u(w_t) + \beta \mathbb{E}_{s^t} \Big(\lambda U_t(h_{t+1}, \tau_{t+1}, a_{t+1}) +$$
$$(1 - \lambda)(\lambda_e p_{t+1}(h_{t+1}, \tau_{t+1}, W_{y,s^{t+1}}, a_{s^{t+1}}) v^*(h_{t+1}, \tau_{t+1}, W_{y,s^{t+1}}, a_{s^{t+1}})$$
$$+ (1 - \lambda_e p_{t+1}(h_{t+1}, \tau_{t+1}, W_{y,s^{t+1}}, a_{s^{t+1}})) W_{y,s^{t+1}}) \Big) \tag{3.C.2}$$

We define as endogenous states $\mathbf{x}_t = [h_t, \tau_t, y_t, W_{y,t}]$, controls $\mathbf{c}_t = [w_t, W_{y,s^{t+1}}]\ \forall t, s^{t+1}$, whereas the only exogenous state is a_t. The endogenous states follow the law of motion

$$\mathbf{x}_{t+1} = \begin{bmatrix} h_{t+1} \\ \tau_{t+1} \\ y_{t+1} \\ W_{y,s^{t+1}} \end{bmatrix} = l(\mathbf{x}_t, \mathbf{c}_t, a_{s^{t+1}}) = \begin{bmatrix} g(h_t, y_t) \\ \tau_t + 1 \\ y_t \\ W_{y,s^{t+1}}) \end{bmatrix} \tag{3.C.3}$$

In the subsequent notation, where appropriate, we omit listing all states on which elements in the equation, and subsume their dependence under just listing the time t. J can be rewritten, by developing forward the recursion until time T, at which the match surely dissolves, as

$$J_t(\{h_t, \tau_t, y_t, W_{y,t}, a_t\}_{t=t_0}^{T-t_0}) = \mathbb{E}_{t_0} \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_o+i} \Big(f(a_t, y_t, h_t) - w_t \Big) \tag{3.C.4}$$

where $\widetilde{p}_{t_0} = 1$. Notice that the forward-looking constraint in equation 3.C.2 is state contingent and an instance of it applies at *every* node of any possible history $s^t\ \forall t$ given the prevailing W_y promised at that node. The equilibrium is an instance of subgame perfect Nash equilibrium in which an agent chooses its strategies while anticipating the best response of the following agent, as common in dynamic games with a leader-follower component introduced by

Von Stackelberg [1934]. The structure of the problem and the solution also shares some commonality with Ramsey optimal policy problems in which a policy maker (in this case the firm) optimizes the utility of all agents according to some weights and taking into account their optimal behavior. [18]

We can redefine the problem:

$$V_{t_0}(\mathbf{x}_t, a_t) = \sup_{\{w_{s^t}, W_{y,s^t}\}} \mathbb{E}_{t_0} \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_o+i} \Big(f(a_t, y_t, h_t) - w_t \Big) \tag{3.C.5}$$

$$s.t.\ [j=0]: \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_o+i} \Big(f(a_t, y_t, h_t) - w_t \Big) - R \geq 0 \tag{3.C.6}$$

$$\begin{aligned}[j=1, s^t]:\ & W_{y,s^t} - u(w_{s^t}) - \beta \mathbb{E}_{s^t} \Big(\lambda U_{s^{t+1}} + \\ & (1-\lambda)(\lambda_e p_{s^{t+1}} v^*_{s^{t+1}}) + \widetilde{p}_{s^{t+1}} W_{y,s^{t+1}}) \Big) \geq 0 \end{aligned} \tag{3.C.7}$$

where the constraint 3.C.13 is a slack participation constraint for a sufficiently small R, so that the principal (the firm) is willing to enter the contract in the first place.

In the terminology of MM we can label

$$h_0^0(\mathbf{x}_t, \mathbf{c}_t, a_t) = f(a_t, y_t, h_t) - w_t \tag{3.C.8}$$

$$h_1^0(\mathbf{x}_t, \mathbf{c}_t, a_t) = f(a_t, y_t, h_t) - w_t - R \tag{3.C.9}$$

$$h_0^1(\mathbf{x}_t, \mathbf{c}_t, a_t) = W_{y,t} \tag{3.C.10}$$

$$h_1^1(\mathbf{x}_t, \mathbf{c}_t, a_t) = W_{y,t} - u(w_t) + \beta \mathbb{E}_t (\lambda U_{t+1} + (1-\lambda)\lambda_e p_{t+1} v^*_{y,t+1}) \tag{3.C.11}$$

and define the Pareto problem ($\mathbf{PP}_\mu$)

[18] In the terminology of MM, we treat constraints coming from equation 3.C.2 as a set of one period ahead forward looking constraint, which makes the analysis of our case akin to their case where one have $j = 1$ forward looking constraints, and $N_1 = 0$. The difference with their problems, however, is that our problem features finite time, and thus each one period ahead forward looking constraint technically applies to a *different* function j_t (indexed by t).

$$\mathbf{PP}_\mu : V_{\mu,t_0}(\mathbf{x}_t, a_t) = \underset{\{w_{s^t}, W_{y,s^t}\}}{sup} \mathbb{E}_{t_0} \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_0+i} \mu^0 \Big(f(a_t, y_t, h_t) - w_t \Big) + \mu^1 W_{y,t_0} \tag{3.C.12}$$

$$s.t.\ [j=0;\ \gamma^0] : \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_o+i} \Big(f(a_t, y_t, h_t) - w_t \Big) - R \geq 0 \tag{3.C.13}$$

$$[j=1, s^t;\ \gamma^1_{s^t}] :\ W_{y,s^t} - u(w_{s^t}) - \beta \mathbb{E}_{s^t} \Big(\lambda U_{s^{t+1}} +$$
$$(1-\lambda)(\lambda_e p_{s^{t+1}} v^*_{s^{t+1}}) + \widetilde{p}_{s^{t+1}} W_{y,s^{t+1}}) \Big) \geq 0 \tag{3.C.14}$$

Still following the notation from Marcet and Marimon [2019], we can define the Saddle Point Problem ($\mathbf{SPP}_\mu$) as:

$$\mathbf{SPP}_\mu : SV_{\mu,t_0}(\mathbf{x}_{t_0}, a_{t_0}) = \underset{\{\gamma \in \mathbb{R}^l_+\}}{inf} \ \underset{\{w_{s^t}, W_{y,s^{t_0}}\}}{sup} \ \mu^0 \Big(f(a_{t_0}, y_{t_0}, h_{t_0}) - w_{t_0} \Big) + \mu^1 W_{y,t_0} +$$
$$+ \beta \mathbb{E}_t \left(\phi(\mu, \gamma) \sum_{i=0}^{T-t_0} \left[\beta^{t_0+i} \prod_{i=0}^{T-t_0-1} \widetilde{p}_{t_0+1+i} \left(f(a_{t_0+i}, y_{t_0+i}, h_{t_0+i}) - w_{t_0+i} \right) + W_{y,t_0+i} \right] \right) +$$
$$+ \gamma^1 \left(u(w_{t_0} + \beta \mathbb{E}_{t_0} \left(\lambda U_{t_0+1} + (1-\lambda) \lambda_e p_{t_0+1} v^*_{y,t_0+1} \right) \right) +$$
$$+ \gamma^0 \left(f(a_{t_0}, y_{t_0}, h_{t_0}) - w_{t_0} - R \right) \tag{3.C.15}$$

The problem can be restated as a saddle-point problem over a Lagrangian equation

$$\underset{\gamma_t}{inf} \ \underset{\{w_{s^t}, W_{y,s^t}\}}{sup} \ \mu^0 \Big(f(a_{t_0}, y_{t_0}, h_{t_0}) - w_{t_0} \Big) + \mu^1 W_{y,t_0} +$$
$$\gamma^0 \Big((f(a_{t_0}, y_{t_0}, h_{t_0}) - w_{t_0} \Big) - R) +$$
$$\gamma^1_{t_0} \Big(- W_{y,t_0} + u(w_{t_0}) + \beta \mathbb{E}_{t_0} (\lambda U_{t_0+1} + (1-\lambda)(\lambda_e p_{t_0+1} v^*_{t_0+1} + \widetilde{p}_{t_0+1} W_{y,t_0+1}) \Big) +$$
$$\beta \mathbb{E}_{t_0} \Bigg[(\mu_0 + \gamma_0) \sum_{t=t_0+1}^{T} \beta^{t-t_0-1} \prod_{i=0}^{T-t_0-1} \widetilde{p}_{t_0+1+i} \Big(f(a_t, y_t, h_t) - w_t \Big) +$$
$$\sum_{t=t_0+1}^{T} \mathbb{E}_t \beta^{t-t_0-1} \prod_{i=0}^{t-t_0-1} \widetilde{p}_{t_0+1+i} \gamma^1_t \Big(- W_{y,t} + u(w_t) +$$
$$\beta(\lambda U_{t+1} + (1-\lambda)(\lambda_e p_{t+1} v^*_{t+1} + \widetilde{p}_{t+1} W_{y,t+1}) \Big) \Bigg] \tag{3.C.16}$$

which, thanks to some algebra and the law of iterated expectations becomes

$$\begin{aligned} \underset{\gamma_t}{inf} \underset{\{w_{s^t}, W_{y,s^t}\}}{sup} & -\gamma^0 R + \mathbb{E}_{t_0} \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_0+i} \Big[\Big(\mu_t^0 + \gamma_t^0\Big)\Big(f(a_t, y_t, h_t) - w_t\Big) + \mu_t^1 W_{y,t} - \\ & \gamma_t^1 \Big(W_{y,t} - u(w_t) - \beta(\lambda U_{t+1} - (1-\lambda)\lambda_e p_{t+1} v_{t+1}^*\Big)\Big] \end{aligned} \tag{3.C.17}$$

where $\mu_t^0 = \mu^0 = 1$, $\gamma_t^0 = \gamma_0 = 0$, $\mu_t^1 = \gamma_{t-1}^1$ for some starting $\gamma_{t_0-1}^1$.

The problem can now be written in recursive form. Define

$$\mathcal{P}_t(h_t, \tau_t, y_t, a_t, \gamma_t) = \underset{W_{y,t}}{sup}\, J_t(h_t, \tau_t, y_t, W_{y,t}, a_t) + \mu_t^1 W_{y,t} \tag{3.C.18}$$

Given equation 3.C.17 the SPFE of the problem can be written as

$$\begin{aligned} \mathcal{P}_t(h_t, \tau_t, y_t, a_t, \gamma_t) =& \underset{\gamma_t}{inf} \underset{w_t}{sup} (f(a_t, y_t, h_t) - w_t) + \mu_t^1 W_{y,t} - \gamma_t(W_{y,t} - u(w_t)) + \\ & \beta \mathbb{E}_t(\lambda U_{t+1} + (1-\lambda)\lambda_e p_{t+1} v_{t+1}^*) + \beta \mathbb{E}_t \widetilde{p}_{t+1} \mathcal{P}_{t+1}(h_{t+1}, \tau_{t+1}, y_{t+1}, a_{t+1}, \gamma_{t+1}) \end{aligned} \tag{3.C.19}$$

One can easily verify that the solution of this equation is the same we found in the maximization of equation (3.8) in the main text. Take the first order conditions and compute the envelope condition:

$$[FOC\; w_t]: \; -1 + \gamma_t u'(w_t) = 0 \tag{3.C.20}$$

$$[ENV\; W_{y,t}]: \; \frac{\partial \mathcal{P}_t}{\partial W_{y,t}} = \mu_t^1 - \gamma_t \tag{3.C.21}$$

$$[FOC\; W_{y,t+1}]: \; -\widetilde{p}_{t+1} W_{y,t+1} \gamma_t + \frac{\partial \widetilde{p}_{t+1}}{\partial W_{y,t+1}} \mathcal{P}_{t+1} + \widetilde{p}_{t+1} \frac{\partial \mathcal{P}_{t+1}}{\partial W_{y,t+1}} = 0 \tag{3.C.22}$$

where equation3.C.22 is obtained by adding and subtracting from equation 3.C.19 $\beta\gamma_t \widetilde{p}_{t+1} W_{y,t+1}$. The reader should also keep in mind that the condition in equation 3.C.22 is actually state contingent and applied to *all* future states next period, with a different set of co-states $\gamma_{s^{t+1}}$ for each realization of a_{t+1}.

Some rearranging of the equation 3.C.22 leads to the following result

$$\frac{\partial \log \widetilde{p}_{t+1}}{\partial W_{y,t+1}} \left(\mathcal{P}_{t+1} - \gamma_t W_{y,t+1} \right) = \gamma_{t+1} - \mu^1_{t+1} \tag{3.C.23}$$

which, given the law of motion of the co-states and the definition in equation 3.C.18 can be re-written as:

$$\frac{\partial \log \widetilde{p}_{t+1}}{\partial W_{y,t+1}} J_{t+1} = \frac{1}{u'(w_{t+1})} - \frac{1}{u'(w_t)} \tag{3.C.24}$$

which is exactly equation (3.14), namely the Euler equation that governs the behavior of wage setting and disciplines the provision of insurance within the contract.

3.D Existence of a Block Recursive Equilibrium

In order to show that a Block Recursive Equilibrium (BRE) exists in our model we need to show that the equilibrium contracts, the workers' and the entrepreneurs value and policy functions do not depend on the distribution of employed and unemployed workers.

Most of the results are tightly linked to our search protocol, directed versus random search, and our contracting structure whereby workers have finite lives and therefore contracts end in finite time. The intuition for why directed search is paramount for the existence of a BRE is linked to the fact that with directed search, workers that are matched with a particular job accept that job with certainty as they are actively looking for it in the labor market. This certainty of acceptance makes the probability of filling a vacancy, and consequently the return of opening it in a particular submarket, independent from the type of worker a firm meets. This implies that the only element of the aggregate state that matters for a firm when making an hiring decision is the state of aggregate productivity but not the distribution of worker types (e.g. employed vs unemployed).

Proposition 3.D.1. *A block recursive equilibrium as defined in Definition 3.3.3 exists.*

Proof. We follow the approach in Menzio *et al.* [2016], Herkenhoff *et al.* [2019] and prove the existence of a BRE using backward induction.

Consider the lifetime values of an unemployed and an employed worker before the production stage in the last period of households lives with $\tau = T$:

$$U(h, T; \Omega) = u(b(h, T)) \tag{3.D.1}$$

$$V(h, T, W; \Omega) = u(w(a)), \tag{3.D.2}$$

their values trivially do not depend on the distribution of types as both valuations are 0 from $T + 1$ onward. Hence, $U(h, T; \Omega) = U(h, T; a)$ and $V(h, T, W; \Omega) = V(h, T, W; a)$.

The optimal contract for agents aged $\tau = T$, instead, solves the following problem

$$J_t(h, T, y, W; \Omega) = \sup_w [f(y, h; a) - w] \quad s.t.\ W = u(w),$$

that clearly does not depend on the distribution of worker types due to the directed search protocol and where the aggregate state only affects the promised utility and the optimal wage through realization of the aggregate productivity processes. Therefore, $J_t(h, T, y, W; \Omega) = J_t(h, T, y, W; a)$.

This also implies that the equilibrium market tightness

$$\theta(h, T, W; \Omega) = q^{-1} \left(\frac{c(y)}{J_t(h, T, y, W; a)} \right)$$

is independent from the distribution of worker types and it is only affected by realization of aggregate productivity, so $\theta(h, T, W; a)$.

This in turn implies that the search problem workers face at the beginning of the last period of their lives depends on the aggregate state only through aggregate productivity a:

$$R(h, T, V; a) = \sup_{\{v_{y,\Omega}\}} \Big[p(\theta(h, T, v_{y,\Omega}; a)) \big[v_{y,\Omega} - V \big] \Big],$$

does not depend on the distribution of worker types.

Stepping back at $\tau = T-1$, the value functions for the unemployed and the employed agents are solutions to the following dynamic programs

$$\sup_{\{v_{y,\Omega'}\}} u(b(h,T-1)) + \beta\mathbb{E}_\Omega\Big(U_{t+1}(h,T;a') + p(\theta(h,T,v_{y,\Omega'};a'))\big[v_{y,\Omega'} - U_{t+1}(h,T;a')\big]\Big)$$

$$u(w) + \beta\mathbb{E}_\Omega\left(\begin{array}{c}\lambda U_{t+1}(g(h,y),T;a') + \beta(1-\lambda)W_{\Omega'}+ \\ +\beta(1-\lambda)\lambda_e \max(0, R(g(h,y),T,W_{\Omega'});a')]\Big]\end{array}\right),$$

where both do not depend on the distribution of worker types.

The optimal contract at this step is a solution to

$$\begin{aligned} J_t(h,T-1,y,V;a) = \sup_{w_i,\{W_{i,\Omega'}\}} \sum_{i=1,2} \pi_i \Big(& f(y,h;a) - w_i \\ & + \mathbb{E}_\Omega\left[\widetilde{p}_{t+1}(h',T,W_{i,\Omega'};a')(J_{t+1}(h',T,y,W_{i,\Omega'};a')\right]\Big) \end{aligned}$$

$$\begin{aligned} s.t.\ V = \sum_{i=1,2} \pi_i \left(u(w_i) + \mathbb{E}_\Omega \widetilde{r}_{t+1}(h',T,W_{i,\Omega'};a')\right),\ h' = g(h,y) \\ \mathbb{E}_\Omega \sum_{i=1,2} \pi_i \left(\mathbb{E}_\Omega J_{t+1}(h',T,y,W_{i,\Omega'};a')\right) \geq 0 \text{ and } t \leq T \end{aligned}$$

which does not depend on types distribution.

Therefore, also the equilibrium tightness and the search gain at $T-1$ are independent from types' distributions, as

$$\begin{aligned} \theta(h,T-1,W;a) &= q^{-1}\left(\frac{c(y)}{J_t(h,T-1,y,W;a)}\right) \\ R(h,T-1,V;a) &= \sup_{\{W_{y,\Omega}\}} \Big[p(\theta(h,T-1,v_{y,\Omega};a))\big[v_{y,\Omega} - V\big]\Big]. \end{aligned}$$

Stepping back from $\tau = T-1,...,1$ and repeating the arguments above completes the proof. $\square$

Bibliography

Abowd, J. M., McKinney, K. L. and Zhao, N. L. (2018). Earnings inequality and mobility trends in the united states: Nationally representative estimates from longitudinally linked employer-employee data. *Journal of Labor Economics*, **36** (S1), S183–S300.

Abreu, D., Pearce, D. and Stacchetti, E. (1990). Toward a Theory of Discounted Repeated Games with Imperfect Monitoring. *Econometrica*, **58** (5), 1041–1063.

Ackerberg, D. A., Caves, K. and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, **83** (6), 2411–2451.

Alp, H. (2019). *Incorporation, Selection and Firm Dynamics: A Quantititative Exploration*. Working paper.

Altomonte, C., Favoino, D., Morlacco, M. and Sonno, T. (2018). *Market Power under Heterogeneous Financial Frictions*. Working paper, Bocconi University.

Altonji, J. G., Kahn, L. B. and Speer, J. D. (2016). Cashier or consultant? entry labor market conditions, field of study, and career success. *Journal of Labor Economics*, **34** (S1), S361–S401.

Anderson, E., Rebelo, S. and Wong, A. (2018). *Markups Across Space and Time*. Working paper, National Bureau of Economic Research.

Arellano-Bover, J. (2020). *Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size*. Working paper.

Argente, D., Lee, M. and Moreira, S. (2018). *How do Firms Grow? The Life Cycle of Products Matters*. Working paper.

Arkolakis, C. (2016). A Unified Theory of Firm Selection and Growth. *Quarterly Journal of Economics*, (1), 1–49.

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, pp. 47–57.

Audoly, R. (2020). Firm dynamics and random search over the business cycle.

Autor, D., Dorn, D., Katz, L. F., Patterson, C. and Van Reenen, J. (2017). *The fall of the labor share and the rise of superstar firms*. Working paper, National Bureau of Economic Research.

—, —, —, — and — (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *The Quarterly Journal of Economics*, **135** (2), 645–709.

Beaudry, P. and Guay, A. (1996). What do interest rates reveal about the functioning of real business cycle models? *Journal of Economic Dynamics and Control*, **20** (9-10), 1661–1682.

Belo, F., Gala, V., Salomao, J. and Vitorino, M. A. (2019). *Decomposing Firm Value*. Working paper, National Bureau of Economic Research.

Benveniste, L. and Scheinkman, J. (1979). On the differentiability of the value function in dynamic models of economics. *Econometrica*, **47** (3), 727–732.

Bernard, A. B., Dhye, E., Magerman, G., Manova, K. and Moxnes, A. (2019). *The origins of firm heterogeneity: a production network approach*. Working paper.

Bernstein, S. (2015). Going Public Affect Innovation ? *The Journal of Finance*, **70** (4), 1365–1403.

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

—, — and — (2019). Monopoly power and endogenous product variety: Distortions and remedies. *Americal Economic Journal: Macroeconomics*, **11** (4).

Bils, M. (1987). The cyclical behavior of marginal cost and price. *The American Economic Review*, pp. 838–855.

— (1989). Pricing in a customer market. *The Quarterly Journal of Economics*, **104** (4), 699–718.

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I. and Terry, S. J. (2018). Really uncertain business cycles. *Econometrica*, **86** (3), 1031–1065.

Blundell, R. and Bond, S. (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric reviews*, **19** (3), 321–340.

Bond, S., Hashemi, A., Kaplan, G. and Zoch, P. (2020). *Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data*. Working paper, National Bureau of Economic Research.

Bonhomme, S., Lamadon, T. and Manresa, E. (2019). A Distributional Framework for Matched Employer Employee Data. *Econometrica*, (87), 1–71.

Bornstein, G. (2018). *Entry and profits in an aging economy: the role of consumer inertia*. Working paper.

Burdett, K., Carrillo-Tudela, C. and Coles, M. (2016). The Cost of Job Loss.

Burstein, A., Carvalho, V. M. and Grassi, B. (2019). *Bottom-up Markup Fluctuations*. Working paper.

Carvalho, V. M. and Grassi, B. (2019). Large Firm Dynamics and the Business Cycle. *The American Economic Review*, **104** (April), 1–64.

Cavenaile, L. and Roldan, P. (2019). *Advertising, Innovation and Economic Growth*. Working paper.

Chemmanur, T. and Yan, A. (2009). Product market advertising and new equity issues. *Journal of Financial Economics*, **92** (1), 40–65.

Chemmanur, T. J. and He, J. (2011). IPO waves, product market competition, and the going public decision: Theory and evidence. *Journal of Financial Economics*, **101** (2), 382–412.

—, —, He, S. and Nandy, D. (2018). Product Market Characteristics and the Choice between IPOs and Acquisitions. *Journal of Financial and Quantitative Analysis*, **53** (2), 681–721.

Chod, J. and Lyandres, E. (2011). Strategic IPOs and product market competition. *Journal of Financial Economics*, **100** (1), 45–67.

Christiano, L. J., Eichenbaum, M. and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, **113** (1), 1–45.

Clementi, G. L. and Palazzo, B. (2016). Entry, exit, firm dynamics, and aggregate fluctuations. *American Economic Journal: Macroeconomics*, **8** (3), 1–41.

Cole, H. and Kubler, F. (2012). Recursive contracts, lotteries and weakly concave pareto sets. *Review of Economic Dynamics*, **15** (4), 479–500.

Cooper, R. and Willis, J. (2014). *Discounting: Investment sensitivity and aggregate implications*. Working paper.

Crouzet, N. and Eberly, J. (2018). *Understanding Weak Capital Investment: the Role of Market Concentration and Intangibles*. Working paper, Prepared for the Jackson Hole Economic Policy Symposium, Federal Reserve Bank of Kansas City.

De Loecker, J. and Eeckhout, J. (2017). *The rise of market power and the Macroeconomic Implications*. Working paper, National Bureau of Economic Research.

—, Eeckhout, J. and Unger, G. (2018). *The rise of market power and the Macroeconomic Implications*. Working paper.

—, — and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, **135** (2), 561–644.

— and Warzynski, F. (2012). American Economic Association Markups and Firm-Level Export Status. *American Economic Review*, **102** (6), 2437–2471.

Deaton, A. (1997). *The analysis of household surveys: a microeconometric approach to development policy*. The World Bank.

Decker, R. A., Haltiwanger, J. C., Jarmin, R. S. and Miranda, J. (2017). Decling Dynamism, Allocative Efficiency, and the Productivity Slowdown. *American Economic Review: Papers & Proceedings*, **107** (5), 322–326.

Dhingra, S. and Morrow, J. (2019). Monopolistic competition and optimum product diversity under firm heterogeneity. *Journal of Political Economy*, **127** (1), 196–232.

Edmond, C., Xu, D. Y. and Midrigan, V. (2018). *How costly are markups?* Working paper, National Bureau of Economic Research.

Fernández-Kranz, D. and Rodríguez-Planas, N. (2017). The perfect storm: Graduating in a recession in a segmented labor market.

Fitzgerald, D. and Priolo, A. (2018). *How do firms build market share?* Working paper, National Bureau of Economic Research.

Fort, T. C., Haltiwanger, J., Jarmin, R. S. and Miranda, J. (2013). How Firms Respond to Business Cycles: The Role of Firm Age and Firm Size. *IMF Economic Review*, **61** (3), 520–559.

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.

—, — and — (2016). The Slow Growth of New Plants: Learning about Demand? *Economica*, **83** (329), 91–129.

Galenianos, M. and Gavazza, A. (2017). A structural model of the retail market for illicit drugs. *American Economic Review*, **107** (3), 858–96.

Garcia-Cabo, J. (2018). The Macroeconomic Effects of Employment Protection on Human Capital and Jobs.

Gilbukh, S. and Roldan, P. (2018). *Firm Dynamics and Pricing under Customer Capital Accumulation*. Working paper.

GILCHRIST, S., SCHOENLE, R., SIM, J. and ZAKRAJSEK, E. (2016). *Financial Heterogeneity and Monetary Union*. Working paper.

—, —, SIM, J. W. and ZAKRAJSEK, E. (2017). Inflation Dynamics during the Financial Crisis. *American Economic Review*, **107** (3), 785–823.

GONZÁLEZ, B. (2020). *Macroeconomics, Firm Dynamics and IPOs*. Working paper.

GOURIO, F. and RUDANKO, L. (2014). Customer capital. *Review of Economic Studies*, **81** (3).

GRULLON, G., KANATAS, G. and KUMAR, P. (2006). The impact of capital structure on advertising competition: An empirical study. *The Journal of Business*, **79** (6), 3101–3124.

GUO, N. (2018). The effect of an early career recession on schooling and lifetime welfare. *International Economic Review*, **59** (3), 1511–1545.

GUTIÉRREZ, G. and PHILIPPON, T. (2019). *The Failure of Free Entry*. Working paper.

HALL, R. (1988). The relation between price and marginal cost in u.s. industry. *Journal of Political Economy*, **96** (5), 921–47.

HALL, R. E. (1986). Market structure and macroeconomic fluctuations. *Brookings papers on economic activity*, **1986** (2), 285–338.

HARRIS, M. and HOLMSTROM, B. (1982). A Theory of Wage Dynamics. *The Review of Economic Studies*, **49** (3), 315.

HECKMAN, J. and ROBB, R. (1985). Using longitudinal data to estimate age, period and cohort effects in earnings equations. In *Cohort analysis in social research*, Springer, pp. 137–150.

HERKENHOFF, K. F., PHILLIPS, G. and COHEN-COLE, E. (2019). How Credit Constraints Impact Job Finding Rates, Sorting & Aggregate Output. pp. 1–78.

HOFFMANN, E. B. (2017). *The Cyclical Composition of Startups*. Working paper.

HONG, S. (2017). *Customer Capital , Markup Cyclicality , and Amplification*. Working paper, Working Paper.

HOTTMAN, C. J., REDDING, S. J. and WEINSTEIN, D. E. (2016). Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics*, **131** (3), 1291–1364.

HUCKFELDT, C. (2018). *Understanding the Scarring Effect of Recessions*. Tech. rep., Society for Economic Dynamics.

JACOBSON, L. J., LALONDE, R. J. and SULLIVAN, D. G. (1993). Earnings Losses of Displaced Workers. *The American Economic Review*, **83** (4), 685–709.

JAIMOVICH, N. and FLOETOTTO, M. (2008). Firm dynamics, markup variations, and the business cycle. *Journal of Monetary Economics*, **55** (7), 1238–1252.

JAROSCH, G. (2015). Searching for Job Security and the Consequences of Job Loss.

KAHN, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, **17** (2), 303–316.

KAPLAN, G. and ZOCH, P. (2020). *Markups, Labor Market Inequality and Nature of Work*. Working paper.

KHAN, A. and THOMAS, J. K. (2008). Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica*, **76** (2), 395–436.

KOCHERLAKOTA, N. R. (1996). Implications of efficient risk sharing without commitment. *The Review of Economic Studies*, **63** (4), 595–609.

KRUEGER, D. and UHLIG, H. (2006). Competitive risk sharing contracts with one-sided commitment. *Journal of Monetary Economics*, **53** (7), 1661–1691.

KUENG, L., YANG, M.-J. and HONG, B. (2014). *Sources of Firm Life-Cycle Dynamics: Differentiating Size vs. Age Effects*. Working paper, National Bureau of Economic Research.

LACHOWSKA, M., MAS, A. and WOODBURY, S. A. (2017). Sources of Displaced Workers' Long-Term Earnings Losses.

LAMADON, T. (2016). Productivity Shocks , Long-Term Contracts and Earnings Dynamics. pp. 1–86.

LISE, J. and POSTEL-VINAY, F. (2019). Multidimensional Skills , Sorting , and Human Capital Accumulation.

— and ROBIN, J.-M. (2017). The macrodynamics of sorting between workers and firms. *American Economic Review*.

LJUNGQVIST, L. and SARGENT, T. J. (1998). The European Unemployment Dilemma. *Journal of Political Economy*, **106** (3), 514–550.

LUTTMER, E. G. J. (2007). Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, **122** (3), 1103–1144.

MARCET, A. and MARIMON, R. (2019). Recursive contracts. *Econometrica*, **87** (5), 1589–1631.

MELE, A. (2014). Repeated moral hazard and recursive lagrangeans. *Journal of Economic Dynamics and Control*, **42**, 69–85.

MENZIO, G. and SHI, S. (2010). Block recursive equilibria for stochastic models of search on the job. *Journal of Economic Theory*.

—, TELYUKOVA, I. A. and VISSCHERS, L. (2016). Directed Search over the Life Cycle . *Review of Economic Dynamics*, **124** (3), 771–825.

MESSNER, M., PAVONI, N. and SLEET, C. (2012). Recursive methods for incentive problems. *Review of Economic Dynamics*, **15** (4), 501–525.

MOREIRA, S. (2015). *Firm Dynamics, Persistent Effects of Entry Conditions, and Business Cycles*. Working paper.

MOSCARINI, G. and POSTEL-VINAY, F. (2012). The contribution of large and small employers to job creation in times of high and low unemployment. *American Economic Review*, **102** (6), 2509–39.

NAKAMURA, E. and STEINSSON, J. (2011). Price setting in forward-looking customer markets. *Journal of Monetary Economics*, **58** (3), 220–233.

NEIMAN, B. and VAVRA, J. S. (2019). *The Rise of Niche Consumption*. Working paper, National Bureau of Economic Research.

NEKARDA, C. J. and RAMEY, V. A. (2013). *The cyclical behavior of the price-cost markup*. Working paper, National Bureau of Economic Research.

OECD (2014). *Employment Outlook 2014*.

OREOPOULOS, P., VON WACHTER, T. and HEISZ, A. (2012). The short- and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, **4** (1), 1–29.

OTTONELLO, P. and WINBERRY, T. (2019). *Financial Heterogeneity and the Investment Channel of Monetary Policy*. Working paper.

PACIELLO, L., POZZI, A. and TRACHTER, N. (2019). Price dynamics with customer markets. *International Economic Review*, **60** (1), 413–446.

PAKES, A. (1994). Dynamic structural models: Problems and prospects. mixed continuous discrete controls and market interactions. In C. Sims (ed.), *Advances in Econometrics*, Cambdridge University Press.

PERLA, J. (2019). *Produc Awarness, Industry LifeCycles and Aggregate Profits*. Working paper.

PETERS, M. (2018). *Heterogeneous Markups, Growth and Endogenous Miscallocation*. Working paper.

PHELPS, E. S. and WINTER, S. G. (1970). Optimal price policy under atomistic competition. *Microeconomic foundations of employment and inflation theory*, pp. 309–337.

PUGSLEY, B., SEDLÁČEK, P. and STERK, V. (2019). *The Nature of Firm Growth*. Working paper.

RAMEY, V. A. (2016). Macroeconomic shocks and their propagation. In *Handbook of macroeconomics*, vol. 2, pp. 71–162.

RAVN, M., SCHMITT-GROHÉ, S. and URIBE, M. (2006). Deep habits. *The Review of Economic Studies*, **73** (1), 195–218.

ROTEMBERG, J. J. and WOODFORD, M. (1999). The cyclical behavior of prices and costs. *Handbook of macroeconomics*, **1**, 1051–1135.

RUDIN, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill Book Company.

SCHMIEDER, J. F., VON WACHTER, T. and HEINING, J. (2018). The Costs of Job Displacement over the Business Cycle and Its Sources : evidence from Germany. *Working Paper*.

SCHULHOFER-WOHL, S. (2018). The age-time-cohort problem and the identification of structural parameters in life-cycle models. *Quantitative Economics*, **9** (2), 643–658.

SCHWANDT, H. and VON WACHTER, T. (2019). Unlucky cohorts: Estimating the long-term effects of entering the labor market in a recession in large cross-sectional data sets. *Journal of Labor Economics*, **37** (S1), S161–S198.

SEDLÁČEK, P. and STERK, V. (2017). The Growth Potential of Startups over the Business Cycle. *American Economic Review*, **107** (10), 1–39.

SHI, S. (2009). Directed search for equilibrium wage-tenure contracts. *Econometrica*, **77** (2), 561–584.

SPEAR, S. E. and SRIVASTAVA, S. (1987). On Repeated Moral Hazard with Discounting. *The Review of Economic Studies*, **54** (4), 599.

STOUGHTON, N. M., WONG, K. P. and ZECHNER, J. (2001). Ipos and product quality. *The Journal of Business*, **74** (3), 375–408.

SYVERSON, C. (2019a). Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives*, **33** (3), 23–43.

— (2019b). Macroeconomics and market power: Facts, potential explanations and open questions. *Economic Studies at Brookings*.

THOMAS, J. and WORRAL, T. (1988). Self-enforcing wage contracts . *The Review of Economic Studies*, **55** (4), 541–553.

TRAINA, J. (2018). *Is Aggregate Market Power Increasing? Production Trends Using Financial statements*. Working paper.

TSUYUHARA, K. (2016). Dynamic Contracts With Worker Mobility Via Directed on-the-Job Search. *International Economic Review*, **57** (4), 1405–1424.

VAN REENEN, J. (2018). *Increasing differences between firms: market power and the macro-economy*. Working paper, Prepared for the Jackson Hole Economic Policy Symposium, Federal Reserve Bank of Kansas City.

VARDISHVILI, I. (2018). *Entry Decision, Option Value of Delay and Business Cycles*. Working paper.

VON STACKELBERG, H. (1934). *Marktform und gleichgewicht*. J. springer.

WEE, S. L. (2016). Delayed Learning and Human Capital Accumulation : The Cost of Entering the Job Market During a Recession.

WIES, S. and MOORMAN, C. (2015). Going public: How stock market listing changes firm innovation behavior. *Journal of Marketing Research*, **52** (5), 694–709.

WINBERRY, T. (2020). *Lumpy Investment, Business Cycles,and Stimulus Policy*. Working paper.

XIAOLAN, M. Z. (2014). *Who Bears Firm-Level Risk? Implications for cash-flow volatility*. Working paper.