Heterogeneous Markups Cyclicality and Monetary Policy*

Andrea Chiavari† Marta Morazzoni‡ Danila Smirvov§

First draft: July 2021. This draft: October 14, 2021

Preliminary and Incomplete

Abstract

Firms' markups cyclicality is at the heart of monetary policy transmission in the New Keynesian model. Using US Compustat data and employing local projection techniques, we uncover a novel empirical fact: dominant firms have a more countercyclical markup response after an unexpected contractionary monetary policy shock. Using a heterogeneous firms New Keynesian model with demand accumulation and endogenous markups that evolve over the life-cycle of producers, we show that this is due to the different demand elasticities faced by the firms. Dominant firms face a more inelastic demand, which implies a lower pass-through rate from costs to prices. Therefore, after a contractionary monetary policy shock, dominant firms pass less the reduction in marginal costs to prices compared to competitors, and increase their markups by more, as documented empirically. After calibrating the model to US micro-level data, we find that considering firms' heterogeneous demand elasticities has important implications for monetary policy amplification.

Keywords: Markups, Heterogeneous Firms, Firm Life-Cycle, Monetary Policy Shocks

JEL Codes: D4, E2, E52, L1, O4

*We are grateful to our advisors Isaac Baley, Davide Debortoli and Edouard Schaal for their invaluable support. We particularly thank Andrea Caggese, Jan Eeckhout, Andrea Fabiani, Jordi Galí and Priit Jeenas, as well as all the participants at the CREI Macro Lunch. All errors are our own.

†Department of Economics, Universitat Pompeu Fabra. Email: andrea.chiavari@upf.edu.

‡Department of Economics, Universitat Pompeu Fabra. Email: marta.morazzoni@upf.edu.

§Department of Economics, Universitat Pompeu Fabra. Email: danila.smirnov@upf.edu.

1 Introduction

Far for being a closed topic of investigation, the discussion around the cyclicality of the aggregate markup and its response to monetary policy shocks still fosters a significant volume of macroeconomic research. Parallel to that, recent contributions have brought attention on companies' heterogeneous market power, as the availability of firm-level datasets has made it easier to estimate markups from balance sheet data. However, the empirical evidence of the heterogeneity in the behavior of firm-level markups after interest rate movements is scarce, and any related quantitative analysis has not yet been provided. This project is a first step towards filling this gap. In particular, we document crucial differences in the response of markups to monetary policy shocks by firm age, and assess their macroeconomic relevance into a novel New Keynesian framework enriched with firms' heterogeneity, demand accumulation and endogenous markups that evolve along firms' life-cycle.

We begin by estimating the behavior of markups at the company level conditional on interest rate movements, and document a significant degree of heterogeneity across old and young firms. Combining together quarterly data from Compustat with two different and exogenously identified series of monetary policy (hereafter MP) shocks for the US, we employ state-of-the-art local projection techniques to establish that the markups of firms above the median age respond more countercyclically to negative MP shocks, while for young firms the response is either mildly procyclical or insignificant. Controlling for commonly-used measures of aggregate economic activity and horse-racing our regression specifications with other firm-level characteristics, we are able to confirm that corporate age in particular influences the differential trajectory of markups upon a negative change in the interest rate. Moreover, we provide evidence that this result could indeed be related to a latent process of demand (or customer base) accumulation, for which dominant firms that are more established in their markets may have to change by less their prices in response to MP shocks, thereby leading to the stronger countercyclical response in old firms' markups documented in the data.

Next, we embed our findings into a New Keynesian (NK) framework that we enrich with firm heterogeneity, demand accumulation and endogenous markups. The model features imperfect competition among heterogeneous intermediate firms that produce using labor and choose prices to maximize profits subject to price adjustment costs à la Rotemberg. New firms can entry every period, while the exit of incumbent firms is exogenous. Our framework presents therefore two main characteristics: on the one hand, intermediate firms face a process of demand accumulation, characterized by some persistence and idiosyncratic shocks, along with a long-run mean that allows for the demand faced by companies to increase along their life-cycle. On the other hand, we assume that the final good producer combines together the intermediate inputs by means of a Kimball aggregator. As in Klenow and Willis (2016) for example, this specific choice introduces in a tractable way endogenous markups in the model, as the elasticity of substitution across intermediate goods become decreasing in their relative quantity. Dominant firms will face lower elasticities of demand and, since demand is accumulated with age, older businesses will hence be able to charge higher markups.

The model is then calibrated on the US economy, following standard strategies in the literature and making use of the richness of Compustat data. In particular, the validation analysis shows that our quantitative framework is able to replicate several untargeted features of the data, such as the

increasing profile of markups along the life-cycle of the firms, the fat right tail in the distribution of markups, and the growth rates of sales and employment. Moreover, the model can get realistic steady state distributions of businesses and employment shares by firm age, and replicate the elasticity of wages to firms' sales shares that we estimate in the data. This latter moment is tightly linked to the fact that dominant (and hence old) companies can increase their profits by cutting quantities and raising prices, thereby suppressing labor demand and hence wages in equilibrium.

Importantly, our NK framework enriched with firm heterogeneity, demand accumulation and endogenous markups can deliver the differential response of markups by firm age that we document in Compustat. As previously mentioned, old firms in our model economy face a lower passthrough from costs to prices due to the presence of the Kimball aggregator. When hit by a contractionary MP shock that decreases wages and puts a downward pressure on prices, dominant firms can cut prices by relatively less compared to young ones. Since markups depend on the ratio between firm prices and marginal costs, this mechanism is in turn responsible for the stronger countercyclical response in old firms' markups. In particular, we can match up to 20% of the empirically estimated relative difference in old and young firms' markups responses to a negative MP shock. In our analysis, we also show that the differential response of old firms in the model can be quantitatively decomposed to highlight the contribution of changes in aggregate variables to the overall general equilibrium impact on markups. In particular, the movements in real wages generated by a negative shock to the interest rate are found to be key in shaping the differential behavior of dominant firms' markups.

Finally, we conclude our quantitative analysis with an investigation of the shock amplification mechanisms at play in our framework, comparing our set up to a standard one-firm NK model with price rigidities. Both the presence of the Kimball aggregator and the heterogeneity of firms are shown to affect the way and extent to which MP shocks transmit in the economy, with output decreasing on average by roughly 20 percentage points (p.p.) more after a negative movement in the interest rate. Focusing on the role of the Kimball aggregator, since intermediate firms – especially old ones – temper their price drops after a negative MP shock due to the increase in their desired markup, the shock itself propagates more through quantities than through prices in our set up as opposed to the standard constant elasticity NK framework. At the same time, the Kimball aggregator alone is not sufficient to generate the observed amplification of MP shocks, as its effects on the elasticity of demand faced by firms kick in when firms are indeed heterogeneous and hence have a different passthrough from costs to prices. Firm heterogeneity is therefore key in affecting and amplifying the movements in the macroeconomic aggregates in the economy following a negative MP shock.

We therefore see the contribution of this paper as twofold: on the one hand and to the best of our knowledge, we bring novel evidence on the remarkable heterogeneity in the response of firm markups to MP shocks based on corporate age. Specifically, while several empirical macroeconomic studies have focused on the different response of investment conditional on movements in the interest rates, we take a different perspective and explore the heterogeneous behavior of markups, the most direct measure of firms' market power. On the other hand, enriching a NK model with firm heterogeneity, demand accumulation and endogenous markups, we attempt to quantitatively study the role of firms' life-cycle in shaping the differential response of markups to changes in the interest rates, and then analyse how aggregate shocks propagate (and get amplified) in our model economy.

Related Literature. Our work builds on several macroeconomic contributions to the study of markups cyclicality. With respect to papers that have analysed the *aggregate* markup (see Gali et al. (2007), Hall (1988), Bils et al. (2018) and Nekarda and Ramey (2020)), we focus on the heterogeneous response of *firm-level* markups to changes in the interest rate, both from an empirical and quantitative point of view. Second, in comparison to recent research on firm-level markups by Hong (2017), Burstein et al. (2020), Meier and Reinelt (2020) and Alati (2020), we do no investigate markups response to business cycle movements, but rather markups behavior conditional on monetary policy.

On the other hand, we attempt to contribute to the theoretical and quantitative macroeconomic literature that has started incorporating micro-level heterogeneity into NK frameworks and understand its implications for the transmission of monetary policy. Recent studies in this field have focused on how household-level heterogeneity affects the consumption channel of monetary policy (see, for example, McKay et al. (2016), Kaplan et al. (2018), Auclert (2019), or Wong (2019)). More in line with the spirit of Ottonello and Winberry (2020)'s investigation of firm investment, we explore the role of firm-level heterogeneity in determining differences in the response of markups to monetary policy shocks. In so doing, we also relate our work to several analyses of supply-side heterogeneities in NK set ups, such as studies on price-setting behavior (see Golosov and Lucas (2007)), market power (see Klenow and Willis (2016) and Mongey (2017)), and product life-cycle (see Bilbiie et al. (2007) and Bilbiie et al. (2012)). With respect to these papers, we present a model of firm life-cycle behavior in order to examine the endogenous response of markups to monetary policy by firm age.

Our work is also related to Gilchrist et al. (2017), who study how financial distortions can create an incentive for firms to raise prices in response to adverse financial or demand shocks. While in Gilchrist et al. (2017) the rise in markups reflects firms' decisions to preserve internal liquidity and avoid accessing external finance, the endogenous response to markups in our set up is related to the differential demand elasticities faced by firms in their life-cycle. In this sense, we see our work as closely related to Baqaee et al. (2021) from a theoretical perspective: the authors explore the first-order effect on aggregate TFP caused by the reallocation of resources triggered by a demand shock across firms with non-uniform markups. While we answer a different research question, also in our model dominant firms tend to have both higher markups and lower pass-through from marginal costs to prices. When faced with an increase (decrease) in nominal marginal costs, high-markup firms raise (lower) their prices by less than low-markup firms in order to remain competitive.

Finally, our work is related to the macroeconomic literature pioneered by Gertler and Gilchrist (1994) that empirically documents how the effect of monetary policy can vary across firms of different characteristics. To the best of our knowledge, existing studies in this area have focused on firm-level investment, and assess how firm default risk (Ottonello and Winberry (2020)), liquidity (Jeenas (2019)) or age (Cloyne et al. (2018)) may shape the response of investment to monetary policy shocks. In a similar spirit, Fabiani et al. (2020) examine how monetary policy can influence the maturity structure of corporate debt. We also use state-of-the-art local projection techniques in our empirical analysis and explore the heterogeneous response of firms' markups to monetary policy shocks. In fact, our core contribution is to document that old firms' markups react more counter-cyclically to contractionary interest rate shocks than young ones, and then embed our finding into a heterogeneous firms NK framework augmented with endogenous markups formation.

The paper is organized as follows: in Section 2, we report and discuss the empirical evidence on the heterogeneous cyclicality of firms' markups after monetary policy shocks. Section 3 lays down our theoretical framework, characterized by heterogeneous firms in a NK setting with endogenous markups. Then, in Section 4, we illustrate the calibration and fit of the model, while in Section 5 we present steady state results and firm-level impulse responses to monetary policy shocks, and also discuss amplification mechanisms. In Section 6 we finally conclude and present the way ahead.

2 Empirical Analysis

In what follows, we study the heterogeneous cyclicality of firms' markups in response to monetary policy shocks. We begin by describing the sample of US firms and the monetary policy shock series on which we draw our evidence, and then illustrate how we estimate markups at the firm-level. Secondly, we document that old firms show a more countercyclical markups response after a monetary policy tightening. Finally, we briefly analyse the behavior of markups over firm's life-cycle and motivate why firm age could be a source of heterogeneity in markups responses to demand shocks.

2.1 Sample Construction

As previously mentioned, we make use of firm-level data from Compustat, which contains quarterly balance sheet information for North-American listed companies between 1975 and 2016. Compustat constitutes a panel of US corporations that is sufficiently high-frequency to be used to study monetary policy, and long enough to exploit within-firm variation. However, it comes at the expenses of representing the universe of publicly-listed incorporated firms only, even though these companies are estimated to make up for 30% of private sector employment. In terms of coverage, Compustat reports details on firm performance indicators and outcomes, including sales, liquid assets, financing sources, total assets, and production costs. It also reports the industry sector (SIC codes) where the business operates and firm age, which is the crucial dimension of heterogeneity in our analysis. Importantly, the age variable contained in the original dataset counts the years since incorporation, but we provide a robustness considering the establishment year for each firm in our sample.

Table 1: Summary Statistics

	Sales	Cogs	Assets	Leverage	Liquidity	Age
mean	447.69	303.17	4919.69	0.45	0.17	9.46
p25	6.06	3.31	37.83	0.04	0.02	4
p50	31.01	17.18	229.50	0.18	0.07	8
p75	164.58	100.60	1118.33	0.39	0.22	14
N	715,874	715,874	685,784	641,316	683,696	715,874

Notes: the first three columns are measured in millions of real 2012 $, while column (4) and (5) are ratios and column (6) is measured in years. *Cogs* is the cost of good sold, which includes production expenditures.

Following standard practices in the literature, we restrict our attention to firms that are incorporated in the US and our final sample excludes utilities companies (SIC codes 4900-4999), financial

entities (SIC codes 6000-6999), as well as corporations for which the industry code, or the information on sales, assets and production costs is missing. Whenever applicable, we deflate variables using a GDP-deflator from the NIPA tables. Table 1 reports summary statistics for the variables of interest.

Our final sample of firms is then merged with two different interest rates datasets: first, we take the quarterly monetary policy shock series from Gürkaynak et al. (2005), who build a measure of interest rate surprises based on the % change in the FED Funds Futures rate in 30-minute windows around the policy announcement. Secondly, we also and primarily make use of the quarterly monetary policy shocks from Jarociński and Karadi (2020), a "pure" interest rate surprises series that removes from the estimation any component attributed to the provision of private FED information on the state of the economy to private agents through policy announcements. The common identifying assumption on the exogeneity in the variation of the policy rate is that nothing else occurs within this 30-minutes time window that could drive both private sector behavior and monetary policy decisions. Both series are available for the years and quarters between 1990Q1 and 2016Q4.

Before describing the estimation of markups at the firm level, we briefly recap on other important variables that we further employ as controls in our regressions. First, using balance sheet data, we compute the leverage and the holdings of liquid assets for the companies in our sample. With respect to the former, we take the ratio of corporate total debt divided by total assets in each period, both measured at book values and where debt is the sum of short term and long term debt. Parallel to that and to provide a measure of corporate liquidity, we compute the ratio of cash and short-term investments to total assets. Our main regression specifications also include firm size as a control, which is measured as the log of total assets (at book value).[1] Finally, we complement our firm-level data with general indicators of economic activity at quarterly level. In particular, we include the GDP growth rate, the Consumer Price Index (CPI) growth rates, the Excess Bond Premium (EBP), and the 1-Year Treasury rate change, all taken from the Federal Reserve of St.Louis (FRED) series.[2]

2.2 Markups Estimation

Firm-level markups are a common measure of whether companies are able to set their prices above marginal costs. To estimate them, we follow recent works by De Loecker and Warzynski (2012) and De Loecker et al. (2020), which are based on the production function approach pioneered by Hall (1988) on industry-level data. Their estimation strategy is grounded on firm's optimizing behavior with respect to production costs-minimization, and delivers an estimate of markups at the firm-level without specifying an explicit demand system. In fact, consider a firm i that employs a production technology given by:

$$Q_{i,t} = F_{i,t}(\boldsymbol{X}_{i,t}, K_{i,t}, \omega_{i,t})$$

where $\boldsymbol{X}$ is a vector of variable inputs, K is the predetermined input and ω is firm-specific productivity. The cost minimization problem for each producer can be hence expressed as follows:

$$\min_{\{\boldsymbol{X}_{i,t}, K_{i,t}\}} \{\boldsymbol{P}'_{i,t}\boldsymbol{X}_{i,t} + R_t K_{i,t} + \lambda_{i,t}(Q_{i,t} - Q(\cdot)\}$$

[1] To eliminate seasonality, variables can be measured as the rolling means in the previous 4 quarters as in Jeenas (2019).
[2] Even if the identified monetary policy shock series are exogenous, macro controls are typically included for robustness.

where $\boldsymbol{P}_{i,t}$ is the vector of prices for variable inputs, R_t is the price of the predetermined input, and $\lambda_{i,t}$ is the Lagrangian multiplier associated to the firm's cost minimization problem. One can then compute the first order condition (FOC) for a generic variable input $X^v \in \boldsymbol{X}$, which is given by:

$$\frac{\partial \mathcal{L}(\cdot)}{\partial X^v_{i,t}} = P^v_{i,t} - \lambda_{i,t} \frac{\partial Q(\cdot)}{\partial X^v_{i,t}} = 0 \tag{1}$$

Notice that the Lagrangian multiplier $\lambda_{i,t}$ can be also interpreted as the marginal cost of producing at a given level of output. Equation 1 can be further rearranged as:

$$\frac{\partial Q(\cdot)}{\partial X^v_{i,t}} \frac{X^v_{i,t}}{Q_{i,t}} = \frac{1}{\lambda_{i,t}} \frac{P^v_{i,t} X^v_{i,t}}{Q_{i,t}}$$

Defining the markup as price over marginal costs, $\mu_{i,t} \equiv \frac{P_{i,t}}{\lambda_{i,t}}$, it is possible to rearrange the FOC for a generic variable input $X^v \in \boldsymbol{X}$ such that it yields:

$$\mu_{i,t} = \theta^v_{s,t} \frac{P_{i,t} Q_{i,t}}{P^v_{i,t} X^v_{i,t}} \tag{2}$$

where $\theta^v_{s,t}$ is the elasticity of output with respect to the variable input X^v. The computation of markups can hence be implemented using firms' financial statements only. To estimate this theoretical expression in Compustat, we make use of both sales and cost of good sold data for each firm and in each quarter, which map to the denominator and numerator of Equation 2 according to:

$$\hat{\mu}_{i,t} = \widehat{\theta}^v_{s,t} \frac{\texttt{Sales}_{i,t}}{\texttt{Cogs}_{i,t}}$$

where we use the estimates of the sectoral output-input elasticity $\hat{\theta}^v_{s,t}$ from De Loecker et al. (2020).

2.3 Heterogeneous Markups Cyclicality

We then proceed to use US listed firms' quarterly balance-sheets to investigate cross-sectional differences in the response of markups to interest rate policies. The main goal of our analysis is to estimate how firm i's markup $\mu_{i,t+h}$, at horizon $h \geq 0$, behave in response to a monetary policy shock at time t conditional on firm i's age just before the shock. To this end, we borrow empirical strategies from Jeenas (2019) and Ottonello and Winberry (2020), and use a panel version of the Jordà (2005)'s local projections (hereafter: LP) to regress the cumulative difference in firm markups at different horizons on the interaction term between firm age at time $t-1$ and the monetary policy shock at time t, alongside a set of control variables. This flexible specification enables us to estimate impulse response functions on our firm-level panel data using the identified monetary shocks as instruments for changes in the policy interest rate. In particular, we estimate by OLS the following set of equations:

$$\begin{aligned} \Delta_h \log \mu_{i,t+h} = & \sum\nolimits_{x \in \mathcal{X}} \left(\alpha_{x,h} + \beta_{x,h} \Delta Y_{t-1} + \sum\nolimits_{k=-\kappa}^{h} \gamma^k_{x,h} \varepsilon^m_{t+k} \right) \times \mathbb{1}_{i \in \mathcal{I}^x} \\ & + \sum\nolimits_{\ell=1}^{L} \boldsymbol{\delta}'_h \boldsymbol{X}_{i,t-\ell} + \varphi_{i,h} + \varphi_{s,t,h} + \boldsymbol{\vartheta_h t} + u_{i,t+h} \end{aligned} \tag{3}$$

with horizons $h = 0, 1, \ldots, H$ and $H = 20$ quarters. The dependent variable is the cumulative change in markups for any firm i at horizon h, given by:

$$\Delta_h \log \mu_{i,t+h} \equiv \log \mu_{i,t+h} - \log \mu_{i,t-1}$$

Focusing on our regressors, $\mathbb{1}_{i \in \mathcal{I}^x}$ is an indicator that takes a value of 1 if $i \in \mathcal{I}^x$, namely if the firm i is above the *median* in one or more dimensions of the vector $x \in \{age, leverage, liquidity, assets\}$. The main coefficient of interest is $\gamma_{age,h}$, which captures the relative response of old companies (compared to young ones) to a variation in the FED short-term policy rate.[3] Note that we also horse-race our main regressor – corporate age – against other layers of heterogeneity which have been found to generally influence firms' response to monetary policy shocks (albeit not specifically markups) such as leverage and liquidity (see Jeenas (2019) and Ottonello and Winberry (2020)).

Furthermore, $\boldsymbol{\varepsilon_t^m} \equiv \sum_{k=-4}^{h} \varepsilon_{t+k}^m$ is the series of monetary policy shocks from Jarociński and Karadi (2020), while $\boldsymbol{X_{i,t}}$ is a vector of controls that includes firm-level variables such as sales growth and overhead costs to sales, and macro-level controls like GDP and CPI growth, 1-year treasury rate change, EBP, and fiscal quarter dummies to account for seasonality. Following standard practices in the literature, we include control variable lags (up to 4) and measure the controls and the variables in x at the end of the quarter before the arrival of the monetary policy shock to ensure exogeneity with respect to it. We then allow for firm ($\varphi_{i,h}$), and sector-time ($\varphi_{s,t,h}$) fixed effects (FE) to control for the unobserved time-invariant heterogeneity at the level of the firm and to absorb time-varying shocks that are common to all firms in a given industry. We also include a linear and quadratic trend ($\boldsymbol{\vartheta_h t}$). Saturating the regression in Equation 3 with these FE implies that, first, our coefficients of interest are identified by within-firm variation over time, namely by changes in the markups response of an otherwise identical firm when it is old compared to when it was young. Secondly, the estimation fully exploits the cross-sectional variation across firms in a given industry. Finally, we cluster the standard errors at the firm and quarter level to account for correlation in the error term.[4]

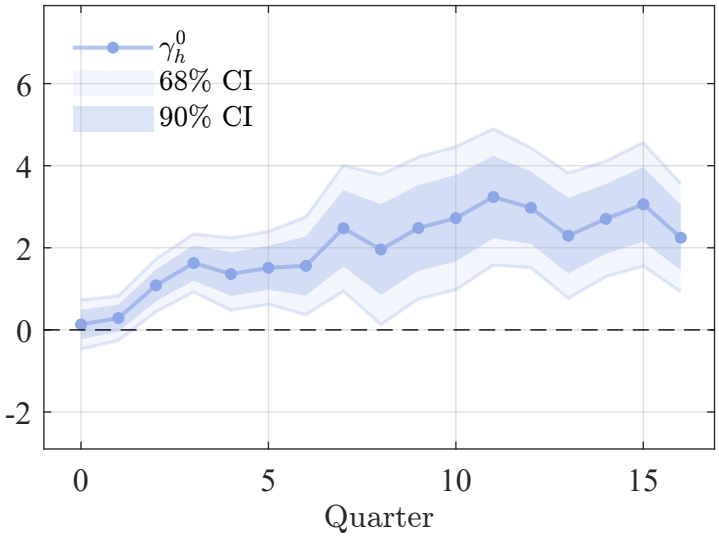
As mentioned in the previous paragraph, our main coefficient of interest is given by $\gamma_{age,h}$, which captures the differential h-quarter growth of markups for firms above the median age after a 25 basis point hike in the interest rate (which corresponds to a rise of a quarter of a percent). Since we are including quarter FE and hence controlling for the time variation of the shock, the coefficient $\gamma_{age,h}$ can precisely identify the excess cyclicality of older firms' markups. In particular, Figure 1 reports the impulse response function obtained from the OLS-estimation of $\gamma_{age,h}$ in Equation 3, along with standard confidence intervals around the point estimates. The magnitude of the $\gamma_{age,h}$ coefficient suggests that being above the median age before a contractionary monetary policy shock hits can imply up to a +3% statistically significant difference in the response of firm markups.

Interestingly, older firms' markups respond more countercyclically to a monetary policy tightening, with the cumulative effect lasting for at least 16 quarters after the shock. It is important to

[3] In our preferred specification, we therefore adopt a non-parametric estimation approach by using dummies instead of linear interactions. We show robustness checks following instead a parametric approach at the end of the section.

[4] Clustering at the firm level allows for a fully flexible dependence in the error terms across time within each company. Clustering by time is necessary whenever firm-level shocks are correlated within a quarter and if this effect may go potentially above the co-movement caused by industry-level shocks already captured by the sector-quarter dummies. We note that the confidence intervals on estimates would be significantly lower without clustering at the quarter-level.

Figure 1: Markups Response to a Monetary Policy Tightening



Notes: Within each quarter, firms' markups are winsorized at the 1% and 99% cutoff, to avoid any outlier to drive our results. Confidence intervals at 90% and 68%, which approximates one standard deviation.

stress that we control for firm's FE and for other crucial determinants of between-firms heterogeneity studied by the literature, namely size, leverage and liquidity. Yet, none of the interactions between these three firm-level variables and the monetary policy shock are statistically significant predictors of markups heterogeneous response to interest rate changes, as further reported in the Appendix. Moreover, as in Cloyne et al. (2018), we note that firm age is pre-determined and cannot vary as a result of changes in monetary policy. In contrast, size, leverage and liquidity endogenously respond to shocks and vary over the business cycle, which can in turn affect the ranking of firms in the distribution of these variables. In this sense, even if there was any, it would be hard to interpret markups (ex-post) heterogeneity as being driven by ex-ante differences in these specific firm characteristics. Contrary to that, we can establish that firm age can significantly determine the differential response of producers' markups to MP shocks, above and beyond other relevant firm characteristics.

The relative response of markups of old companies estimated through Equation 3 does not allow to understand the separate response of markups of firms in different age categories to monetary policy shocks. In particular, the regression specification in Equation 3 is saturated with industry and time FE that span out completely the time-series variation common across all firms. Hence, we proceed to estimate the following regression specification for firms above and below the median age:

$$\Delta_h \log \mu_{i,t+h} = \varphi_{i,h} + \boldsymbol{\vartheta_h t} + \sum_{\ell=1}^{L} \boldsymbol{\delta_h' X_{i,t-\ell}} + \sum_{k=-\kappa}^{h} \gamma_{x,h}^{k} \varepsilon_{t+k}^{m} + u_{i,t+h} \tag{4}$$

with horizons $h = 0, 1, \ldots, H$ and $H = 20$ quarters. Note that the dependent variable is the cumulative change in markups for any firm i at horizon h, which is defined as:

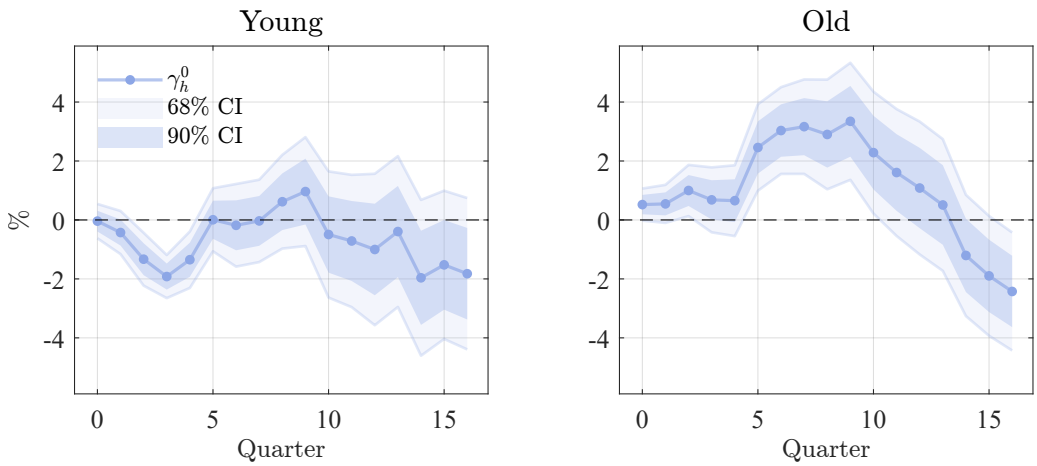
$$\Delta_h \log \mu_{i,t+h} \equiv \log \mu_{i,t+h} - \log \mu_{i,t-1}$$

Hence, in this second specification, we simply exploit the time-variation and look at the absolute change in markups after a change in the interest rate set by the FED for firms of different age cat-

egories, while the coefficient of interest γ_h is estimated for each age group separately. Note that $\boldsymbol{\varepsilon_t^m} \equiv \sum_{k=-4}^{h} \varepsilon_{t+k}^m$ is again the series of monetary policy shock from Jarociński and Karadi (2020). Moreover, $\boldsymbol{X_{i,t}}$ is a vector of controls that include firm-level variables such as sales growth and overhead costs to sales, leverage, liquidity and assets, as well as macro-level controls like GDP growth, CPI growth, 1-year treasury rate change, EBP, and fiscal quarter dummies. Importantly, we also include control variable lags (up to 4). We also allow for firm's FE ($\varphi_{i,h}$) to control for time-invariant firm-heterogeneity, as well as for a linear and quadratic trend ($\boldsymbol{\vartheta_h t}$). Finally, we cluster our robust standard errors at the firm and quarter level to account for correlation in the error term.

The results of our estimation are shown in Figure 2: more specifically, the left panel documents the cumulative response of markups for firms below the median age to a negative movement in the FED interest rate, while the right panel focuses on the markups response of companies above the median age. This second estimation strategy further strengthens the insight from Figure 1, by documenting that older firms present a pronounced and statistically significant countercyclical response in their markups after a monetary policy tightening, while young firms' markups move procyclically, albeit the statistical significance of the estimated coefficient is much lower. Importantly, note that this second specification estimates a dynamic regression without the sector-time fixed effects and still shows that the above-median age firms' response in markups is nonetheless persistent, peaking 8 to 10 quarters after the shock. Taken together, these findings seem to suggest that only old companies do adjust upwards their markups, whereas young firms' markups are generally less sensitive to monetary policy or tend to be adjusted downwards following a negative change in the FED interest rates. Finally,

Figure 2: Firms' Markups Response to a Monetary Policy Shock by Age Category



Notes: Within each quarter, firms' markups are winsorized at the 1% and 99% cutoff, to avoid any outlier to drive our results.

we check that our results hold more generally when we focus on a different partitioning of the age distribution, by for example considering as "old" those firms that are above the third quartile and as "young" firms all the others. Moreover, to have a further understanding of the possible interaction between corporate age and other firm-level characteristics, we also split old and young companies according to their position in the distribution of leverage, liquidity and size, which are other dimensions of firm's heterogeneity typically investigated in the literature that we have always controlled for in our regression analysis. As reported in Figure A.3, Figure A.4 and Figure A.5, corporate age is the crucial dimension determining the heterogeneous response of markups to monetary policy

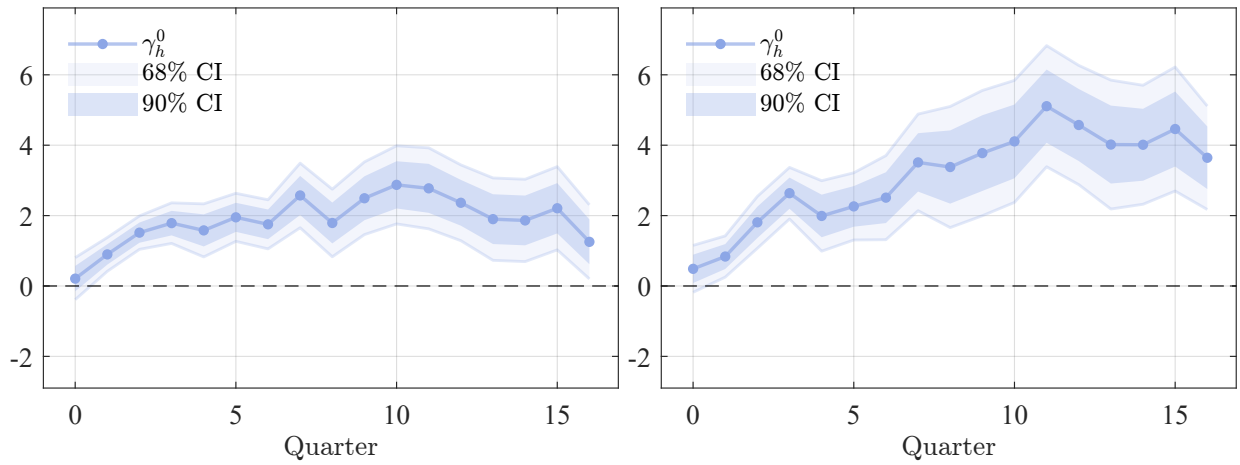
shocks, while other firm's characteristics – such as leverage, liquidity or size – are less powerful or even insignificant predictors of the differential behavior of markups at the company-level.

2.4 Discussion of Results

In what follows, we discuss our main robustness checks to further confirm the evidence on the role of age in shaping markups response to MP shocks at the firm-level. First, we run alternative regression specifications that present minor differences with respect to our baseline case. In particular, relative to the way we define the regressor of interest – namely corporate age – our main result is robust to consider age groups by industry and quarter, and also to interact the interest rate shock series with firm's age in a linear fashion, thereby adopting a parametric estimation strategy (similarly, we also linearly interact the MP shock series with the leverage, size and liquidity of the firms). The results of these alternative specifications are reported in Figure A.1 in the Appendix and both confirms that older firms present a stronger countercyclical response of markups to a monetary policy tightening.

Secondly, we check that our insights are not driven by the specific time span considered, in particular by running again our estimation procedure on a sub-sample of the dataset that extends until the 2009 crisis. This is due to the fact that the Great Recession was indeed a period of exceptional financial conditions and, at the same time, the post-2009 era was characterized by a lower variation in the interest rate policy, with the federal funds rate often hitting the zero lower bound. However, as reported in the right panel of Figure 3, our results acquire a stronger statistical significance when the post-2009 era is excluded, and are hence not driven by specific period conditions only. Moreover,

Figure 3: Using GSS Shocks (left) and Focusing on pre-2009 period (right)



Notes: Within each quarter, firms' markups are winsorized at the 1% and 99% cutoff, to avoid any outlier to drive our results.

we also replicate our estimation using the monetary policy shock series from Gürkaynak et al. (2005) (hereafter GSS), which does not remove the informational component when measuring the interest rate surprises based on the 30-minute windows around FED policy announcements. As it is possible to check from the left panel of Figure 3, the coefficient on the interaction between the MP shock and firm's age – $\gamma_{age,h}$ – is economically relevant and significant, confirming that firms above the median age present an excess counter-cyclicality in their markups response to a monetary policy tightening, and that this differential effect lasts for an horizon of 16 quarters after the shock.

Finally, our findings are robust to excluding future shocks from the estimation, as well as sector-quarter fixed effects, as reported in Figure A.2. In particular, future shocks were included among the regressors to control for the presence of auto-correlation and to increase the estimation precision, despite of the fact that the monetary policy shock series we have used should already be clear from confounding factors of this sort. Taking our results together, we argue that corporate age is a robust driver of firm's heterogeneity in the cyclicality of markups response to a monetary policy shock, and we hence proceed to briefly analyse the behavior of markups over firms' life-cycle.

2.5 Markups and Firm's Life-Cycle

After having estimated the heterogeneous response of firms' markups to MP shocks, we provide a further discussion on why old and young companies may possibly show such stark differences in cyclical behavior of their respective markups. As mentioned before, corporate age has been investigated to be an important element of firm's employment and leverage dynamics over the business cycle by the works of Haltiwanger et al. (2013), Dinlersoz et al. (2018), and Pugsley et al. (2019). Interestingly, Cloyne et al. (2018) have studied how corporate age can determine investment heterogeneity across firms, especially in response to interest rate changes. Specifically, by documenting that the investment and the borrowing of younger firms paying no-dividends exhibit a large and significant decline in response to a tightening of the monetary policy, the authors argue that such companies are more likely to face financial frictions. In their view, this can also rationalize why the borrowing of young and non-dividend paying firms is far more sensitive to fluctuations in collateral values compared to other businesses, for which their results turn less significant.

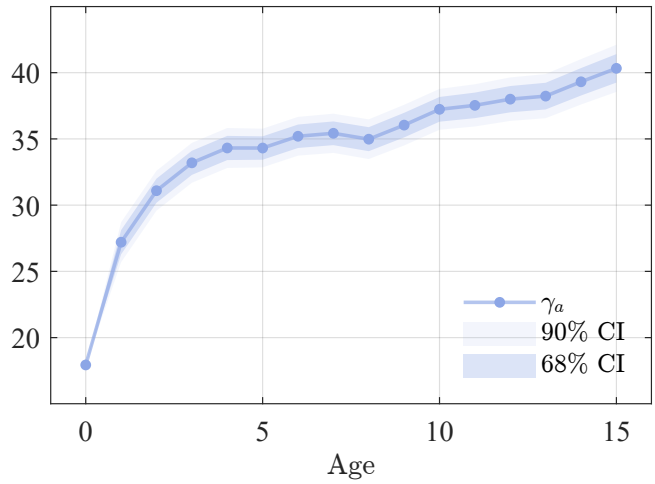
In a similar in spirit, we argue instead that firm age matters significantly for the profile of markups and their response to a MP shock. In particular, it is reasonable to assume that corporate age may capture how *established* is a firm in her (unobservable) product market. Older firms, by means of having competed and produced in their given markets for a longer period, may be able to charge higher prices to consumers and hence be less subject to the downwards pressure exerted on prices by a monetary policy tightening. In fact, according to the theoretical expression of firm-level markups, a negative interest rate shock puts a negative pressure on both input costs and prices. However, if older firms are able to decrease their prices by relatively less by taking advantage of their established position within a market, this may rationalize a more countercyclical markups response to a monetary tightening. To provide suggestive evidence of how corporate age is related to firm's established position in a given market, we examine the profile of markups ($\mu_{i,t}$) and selling expenditure ($sr_{i,t}$) over the business life-cycle using Compustat data. In particular, we run the following regressions:

$$\log \mu_{i,t} = \alpha + \sum_{a=2}^{\mathcal{A}} \gamma_a \mathbb{1}_{\{age_{i,t}=a\}} + \varphi_{s,t} + \varepsilon_{i,t}$$

where $\varphi_{s,t}$ are sector and quarter fixed effects. Not only old firms are on average big, but they most importantly tend to have higher markups, as documented in Figure 4. The main takeaway from Figure 4 is that firms are able to charge higher markups (hence higher prices) as they grow older. We interpret this suggestive evidence as an indicator that older companies may have already secured their customer base enough to be less inclined to drastically reduce prices in response to a negative

monetary policy shock, resulting in the stronger markups counter-cyclicality documented in the previous paragraphs. Since we argue that firm age is a proxy for how established a production unit is in her given market, we will then rationalize our findings into a NK model with heterogeneous firms, demand accumulation and endogenous markups that not only will generally differ across firms but that will be further allowed to grow according to firm's life-cycle.

Figure 4: Markups over Firm's Life-Cycle



3 The Model

In this section, we outline our theoretical framework and discuss how each assumption relates to and can deliver qualitative predictions in line with the evidence from the data. In particular, we enrich a relatively standard NK to accommodate three main novelties: first, we allow for full heterogeneity on the supply side of the economy, by including heterogeneous intermediate firms producing a different variety of input used in the final good sector. Secondly, we introduce a simple form of demand accumulation in our economy that makes the demand for the good of a given firm increase with the time that the firm survives on the market. Third, we embed endogenous and variable markups in the economy, which differ across companies according to the quantity produced, and that also evolve with the life-cycle of the firms. We now proceed to present the model in full details below.

3.1 Household's Side

Time is continuous. The model features a representative household that optimizes the discounted flow of utility from consumption and labor over an infinite lifetime horizon, where we indicate the discount factor as $\rho \geq 0$. We assume that the utility of the agent is strictly increasing and concave in consumption, and strictly decreasing and convex in the amount of hours worked respectively. Preferences are time-separable and the infinite stream of household's utility is hence given by:

$$\int_0^\infty e^{-\rho t} \left(\frac{C_t^{1-\nu}}{1-\nu} - \frac{L_t^{1+\gamma}}{1+\gamma} \right)$$

where ν represents the risk aversion in the CRRA utility function over consumption, whereas γ is the inverse of the Frisch labor elasticity. Moreover, $L_t \in [0,1]$ are the hours supplied as a fraction of the time endowment (normalized to 1), while C_t denotes the aggregate consumption good. In each period, the household can borrow in bonds B_t at the real interest rate r_t. Finally, the household owns all the firms in the economy, while labor supply, aggregate consumption and bond investment paths are chosen as a result of a value maximization problem subject to a standard budget constraint:

$$\begin{aligned} \mathcal{V} &= \max_{\{C_t, L_t, \dot{B}_t\}} \int_0^\infty e^{-\rho t} \left(\frac{C_t^{1-\nu}}{1-\nu} - \frac{L_t^{1+\gamma}}{1+\gamma} \right) dt \\ \text{s.t.} \quad & C_t + \dot{B}_t = W_t L_t + r_t B_t + D_t \end{aligned}$$

where we denote by D_t the dividends from the firms and by W_t the wage earned by the household in real terms. As we will explain below, r_t will be determined by the monetary policy and Fisher equation, while W_t is determined by market clearing conditions for labor. Solving for the optimal value of consumption and labor, we get the following standard Euler and labor supply equations:

$$r = \rho + \nu \frac{\dot{C}}{C} \tag{5}$$

$$L^\gamma C^\nu = \frac{W}{\varphi} \tag{6}$$

3.2 Final Good Producer

A competitive representative final-good producer aggregates a continuum of intermediate inputs indexed by $i \in [0,1]$ according to the following expression:

$$\int_0^1 \mathcal{K}\left(a_{i,t} \frac{y_{i,t}}{Y_t} \right) di = 1 \tag{7}$$

where we assume that intermediate inputs denote by y_t are aggregated using the Kimball aggregator $\mathcal{K}$, with $\mathcal{K}'(\cdot) > 0$, $\mathcal{K}''(\cdot) < 0$, and $\mathcal{K}(1) = 1$. Notice that the CES aggregator obtains as a special case of the Kimball aggregator, and namely when $\mathcal{K}(q) = q^{\frac{\sigma-1}{\sigma}}$ for an elasticity of substitution $\sigma > 1$. Importantly, $a_{i,t}$ is a stochastic demand process that will be explained in due details in the next paragraph. For the moment, taking the prices $p_{i,t}$ of any intermediate input i as given and normalizing the price of the final good to 1, the final good producer minimizes production costs subject to Equation 7. The optimality condition of this problem gives rise to the *inverse demand* function for good i:

$$p_{i,t} = \mathcal{K}'\left(a_{i,t} \frac{y_{i,t}}{Y_t} \right) a_{i,t} \mathcal{D}_t \tag{8}$$

where:

$$\mathcal{D}_t = \left(\int_0^1 \mathcal{K}'\left(a_{i,t} \frac{y_{i,t}}{Y_t} \right) a_{i,t} \frac{y_{i,t}}{Y_t} di \right)^{-1} \tag{9}$$

is a *demand index*. In the CES case $\mathcal{K}(q) = q^{\frac{\sigma-1}{\sigma}}$ this index is a constant $\mathcal{D}_t = \frac{\sigma}{\sigma-1}$ so that Equation 8 would reduce to the familiar constant elasticity demand curve given by $p_{i,t} = \left(a_{i,t}\frac{y_{i,t}}{Y_t}\right)^{\frac{-1}{\sigma}}$. To keep the exposition concise, further derivations related to Equation 8 and Equation 9 are contained in the Appendix. Moreover, we use the Klenow and Willis (2016) specification for $\mathcal{K}(q)$ given by:

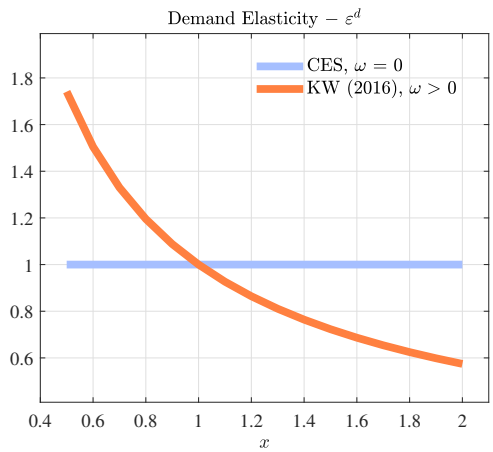
$$\mathcal{K}(q) = 1 + (\sigma - 1)\exp\left(\frac{1}{\omega}\right)\omega^{\frac{\sigma}{\omega}-1}\left[\Gamma\left(\frac{\sigma}{\omega}, \frac{1}{\omega}\right) - \Gamma\left(\frac{\sigma}{\omega}, \frac{q^{\omega/\sigma}}{\omega}\right)\right] \tag{10}$$

where $\sigma > 1$, $\omega \geq 0$ and $\Gamma(s, x)$ is the upper incomplete Gamma function such that $\Gamma(s, x) := \int_x^\infty t^{s-1}e^{-t}dt$. In particular, ω is the *super elasticity*, which is 0 in the CES aggregator. Finally, we can derive an analytical expression for the elasticity of demand ε_i^d that is faced by a producer of any good variety i as a function of the relative quantity of good i in the economy, which is given by:

$$\varepsilon_i^d = \sigma\left(a_{i,t}\frac{y_{i,t}}{Y_t}\right)^{-\frac{\omega}{\sigma}}, \quad \omega \geq 0$$

As already pointed out, the standard CES case is recovered when $\omega = 0$ and hence the elasticity of demand $\varepsilon_i^d = \sigma$ is constant across producers. In contrast, in the case of Kimball aggregator the elasticity of substitution is lower for firms with higher relative quantity $x = a\frac{y}{Y}$, implying that larger firms can choose higher markups, in a similar spirit to the different set up adopted in Atkeson and Burstein (2008) and as further made clear in Figure 5. When $\omega > 0$, the extent to which a firm's markup increases with its relative size is determined by the ratio σ/ω, which will also be shown to quantitatively matter in shaping how markups change with monetary policy later on in the analysis.

Figure 5: Kimball Aggregator



3.3 Intermediate Good Producers

Each intermediate good i is produced by a monopolistically competitive firm using effective units of labor $\ell_{i,t}$ in the production process according to the technology given by:

$$y_{i,t} = \ell_{i,t}^{1-\alpha} \tag{11}$$

with $\alpha \in [0,1]$. In each time t, firms hire labor at wage W_t in a competitive labor market. As already mentioned, intermediate producers are monopolistic competitors on their respective markets and each one of them faces a demand function which can be written explicitly from Equation 8 as:

$$y_{i,t} = \left(1 - \omega \log\left(\frac{\sigma}{\sigma - 1}\frac{1}{a_{i,t}}\frac{p_{i,t}}{\mathcal{D}_t}\right)\right)^{\sigma/\omega} \frac{Y_t}{a_{i,t}} \tag{12}$$

Moreover, each intermediate firm i is characterized by a process of demand accumulation given by a_i, which shows some persistence ρ_a and idiosyncratic shocks given by $\xi_a d\mathcal{W}$, where $d\mathcal{W}$ is a standard Wiener process. We also include a drift $\bar{a}$ that allows for the demand to grow over time, generating a realistic life-cycle profile. It is important to stress that we load the heterogeneity across firms that we see in the data in this specific process, which is meant to capture in a reduced-form the fact that markups and size increase with the firm's life-cycle. Such demand process may actually rationalize some underlying form of customer accumulation or, alternatively, of consumers habit formation. In other words, one can think about it in the sense that the more consumers experience the good of a given firm i, the more inelastic their demand for that specific item would consequently be.

Intermediate firms in this economy, characterized by the demand process a, maximize the discounted stream of profits by choosing prices. Hence, at each instant in time, the state of the economy is given by the joint distribution $\lambda_t(da, dp)$. Finally, intermediate producers discount future profits at the rate $r_t + \delta$, where δ is the exogenous Poisson intensity that determines firm's exit. Exiters are replaced by new firms with an initial a_0 drawn from a log-normal distribution of mean $\bar{a}_{entry}$ and standard deviation $\xi_{a,entry}$, which will be further discussed in the calibration exercise. Moreover, intermediate firms bear Rotemberg adjustment costs when changing prices, which we assume to be proportional to their sales and quadratic. We can summarize the problem of a given firm i as follows:

$$\begin{aligned}
\mathcal{J}_{i,0} = &\max_{\{\dot{p}_{i,t}, \ell_{i,t}, y_{i,t}\}_{t\geq 0}} \mathbb{E}_0 \int_0^\infty e^{-\int_t^\infty (r_t+\delta)dt} \left\{ p_{i,t}y_{i,t} - W_t\ell_{i,t} - \frac{\vartheta}{2}\left(\pi_t + \frac{\dot{p}_{i,t}}{p_{i,t}}\right)^2 p_{i,t}y_{i,t} \right\} dt \\
\text{s.t.} \quad & y_{i,t} = \left(1 - \omega \log\left(\frac{\sigma}{\sigma - 1}\frac{1}{a_{i,t}}\frac{p_{i,t}}{\mathcal{D}_t}\right)\right)^{\sigma/\omega} \frac{Y_t}{a_{i,t}} \\
& y_{i,t} = \ell_{i,t}^{1-\alpha} \\
& \dot{a}_{i,t} = \rho_a(\bar{a} - a_{i,t})dt + \xi_a d\mathcal{W}_{i,t} \\
& p_{i,0} \text{ and } a_{i,0} \text{ given}
\end{aligned}$$

Importantly, the initial price set by entrant firms p_0 is the one that maximizes the expected value $\mathcal{J}_{i,0}$ for a given initial value of firm's productivity $a_{i,0}$. Note that, in the solution process, the demand process given by $\dot{a}_{i,t}$ is exponentiated. Intermediate firms take as given equilibrium paths for the real wage $\{W_t\}_{t\geq 0}$ and the interest rate $\{r_t\}_{t\geq 0}$. In steady state, the recursive solution to this problem consists of decision rules for labor $\ell(a, p; \mathcal{S})$ and output $y(a, p; \mathcal{S})$, with $\mathcal{S} := (r, W, Y, \mathcal{D}, \pi)$. These rules in turn also imply optimal drifts for prices, and together with a stochastic process for a, induce a stationary joint distribution of firms given by $\lambda(da, dp; \mathcal{S})$ and characterized by a standard Kolmogorov forward equation. Out of the steady state, each of these objects is time-varying and depends on the time path of prices and policies: $\{\mathcal{S}_t\}_{t\geq 0} := \{r_t, W_t, Y_t, \mathcal{D}_t, \pi_t\}_{t\geq 0}$.

3.4 Monetary Authority

Our model economy features a monetary authority that sets the nominal interest rate according to a standard Taylor rule, penalizing deviations from the optimal inflation rate π^* in the following way:

$$i_t = \phi_\pi(\pi_t - \pi^*) + \rho + \varepsilon_t^m$$

where $\phi_\pi > 1$, ρ is the discount factor and ε_t^m is the monetary policy shock that can be mapped directly to the series from either Jarociński and Karadi (2020) or Gürkaynak et al. (2005) that we have used in the empirical analysis of the paper. Note that $\varepsilon_t^m = 0$ in steady state: one of our main quantitative exercises will be precisely to study the economy's adjustment after an unexpected temporary monetary shock, namely after a change in ε_t^m. Finally, given inflation π_t and the nominal interest rate i_t, the real return on bonds r_t is determined by the Fisher equation $r_t = i_t - \pi_t$.

3.5 Equilibrium Condition

An equilibrium in this economy is defined as a set of paths for individual household's $\{C_t, L_t\}_{t\geq 0}$ and firm's decisions $\{\dot{p}_{i,t}, \ell_{i,t}, y_{i,t}\}_{t\geq 0}$, input prices $\{W_t\}_{t\geq 0}$, the return on bonds $\{r_t\}_{t\geq 0}$, the inflation rate $\{\pi_t\}_{t\geq 0}$, the distribution of firms $\{\lambda_t\}_{t\geq 0}$, the demand index $\{\mathcal{D}_t\}_{t\geq 0}$, and aggregate quantities such that, at every t: (i) the household and the firms maximize their objective functions taking as given equilibrium prices and aggregate quantities; (ii) the sequence of distributions satisfies aggregate consistency conditions; (iii) all markets clear. There are three markets in our economy: the bond market, the labor market, and the goods market. The bond market clears when the following holds:

$$B_t = 0 \tag{13}$$

Moreover, the labor market clears when:

$$L_t = \int \ell_t(a, p) d\lambda_t$$

Finally, the goods market clears according to:

$$C_t = Y_t - \int \frac{\vartheta}{2}\left(\pi_t + \frac{\dot{p}_t(a,p)}{p}\right)^2 py(a,p)d\lambda_t$$

where C_t is the total real expenditure in consumption, Y_t is aggregate output, and the last term is the sum of adjustment costs to prices paid by intermediate firms.

4 Quantification

In what follows, we proceed to explain the quantification of our model, including the calibration strategy and the overall fit of both targeted and untargeted moments computed from available US data. In particular, we discuss the ability of our theoretical framework to replicate salient features of the markups and firms' distribution, which is a crucial property needed to provide a link with the empirical analysis of the previous sections. Once quantified, the model is then used in Section 5 to

study and analytically decompose the impulse response functions of firms' markups after a negative monetary policy shock. Moreover, in Section 5, we also compare the amplification mechanism implied in our framework with respect to a standard representative firm New Keynesian model.

4.1 Calibration

A model period in one quarter. Of the 14 parameters we need to calibrate, 8 are fixed outside of the model, for which we pick common values used in the literature. In particular, we set the risk aversion $\nu = 2$ and the disutility of labor $\gamma = 2$, while the discount factor $\rho = 0.012$ is specified to deliver a yearly interest rate of 5% in equilibrium. With respect to the parameters related to firm's life-cycle, technology and pricing behavior, we fix the quarter exit rate $\delta = 0.024$ to imply that 10% of the firms exit each year, and the returns to scale $\alpha = 0.33$ such that the labor share is around 0.6 in equilibrium. Moreover, it is important to specify that we normalize at 1 the mean demand at entry $\overline{a}_{entry}$, while the demand dispersion $\xi_{a,entry}$ faced by entrant firms is set to be equal to the dispersion of the demand process faced by incumbents.[5] Finally, the monetary policy coefficient $\phi_\pi = 1.5$ in the Taylor rule is chosen to replicate a similar strategy as in Taylor (1999) and Galí (2015). The full list of both fixed and fitted parameter, as well as targeted moments, is presented in Table 2.

Table 2: Estimated Parameters and Targeted Moments

Fixed	Value	Description			
ρ	0.012	Discount factor			
ν	1	Risk aversion			
γ	2	Inverse Frisch elasticity			
α	0.33	Production function curvature			
δ	0.024	Exit rate			
$\overline{a}_{entry}$	1	Mean demand entrants			
$\xi_{a,entry}$	0.11	Demand dispersion entrants			
ϕ_π	1.5	Taylor rule coefficient			
Fitted	**Value**	**Description**	**Moments**	**Model**	**Data**
θ	20	Price adjustment cost	Avg. cost change prices over sales	0.11	0.09
σ	4	Elasticity of demand	Avg. markup	1.68	1.68
ω	5.1	Superelasticity of demand	Elasticity markups to sale shares	0.11	0.10
$\overline{a}$	2	Mean demand	Median markup	1.37	1.30
ξ_a	0.11	Demand dispersion	Markups standard deviation	1.23	1.22
ρ_a	0.02	Demand mean reversion	Markups growth between age 0-5	0.24	0.22

Note: Empirical estimates for fitted parameters from Compustat Data (1990Q1-2016Q4). For the fixed parameters, see text.

In addition to that, we need to endogenously assign values to the remaining 7 parameters, for which we match as many salient moments from available US data. To begin with, we set the price

[5] Our results do not depend on this choice, which is just a simplification for the sake of the estimation procedure.

adjustment cost factor $\theta = 20$ such that the average ratio between the cost paid by firms to change prices and their sales is the same in the model and in the data.[6] As standard in the literature, we set the elasticity of demand $\sigma = 4$ to match an average markup of 1.68 computed in Compustat data,[7] as such parameter determines the level of substitutability across the output of the different firms in the model and hence influences the average market power in the economy. Moreover, the superelasticity of demand ω is fitted such that the elasticity of markups to sale shares in the model is the same as in the data. In particular, our choice is motivated by the fact that the parameter ω in the Kimball aggregator is tightly linked to the relationship between the relative size of the firms and their markups: if ω was to be 0, such relationship would be null because all firms would have the same markup independently of their size. On the contrary, for $\omega > 0$, the higher the ω, the higher the dependence of markups on sales shares. To this end, using Compustat firm-level data, we empirically estimate the elasticity of (log) markups to (log) sales shares according to:

$$\log \mu_{i,t} = \beta * \log(\textit{sales shares})_{i,t} + \varphi_{s,t} + \varepsilon_{i,t} \tag{14}$$

where $\varphi_{s,t}$ are sector-time FE and the coefficient β precisely informs by how much markups are linked to firms' sales shares. In the model, we use the theoretical definitions of markups and sales shares.

Finally, turning to the parameters related to the demand accumulation process, the mean demand is set to match the median markup in the US economy, as $\bar{a}$ identifies the distance between the average demand faced by entrants and incumbents, and hence relates to the skewness of the markup distribution. Furthermore, the dispersion in the demand process faced by incumbent firms ξ_a is identified from the standard deviation of markups, while the mean reversion in the demand process ρ_a is picked to match the growth of markups for firms between age 0 to 5. In particular, a higher demand mean reversion impacts how fast firms grow, and therefore relates to the trajectory of markups over the firm's life-cycle.

4.2 Quantitative Fit

In the following paragraphs, we present and discuss our main validation exercises, which provide a overview of the quantitative fit of our framework with respect to empirical moments and data features that have not been targeted in the calibration. In particular, we first discuss the cross-sectional and life-cycle characteristics of firms in our model, and how they compare to their empirical counterparts from Compustat. Secondly, we dig into the properties of the markup distribution and then analyse markups dynamics over the firm's life-cycle. Finally, we conclude with a note on the model and data-implied elasticity of wages to sales and relate it to the behavior of markups under the Kimball aggregator case and in imperfect competition, following similar lines as in Edmond et al. (2018).

4.2.1 Implications for Markups in Steady State

One of our main validation exercises is to look at the properties of markups in the data and compare them with the ones implied by our quantitative framework. Importantly, in Section 2, we have shown

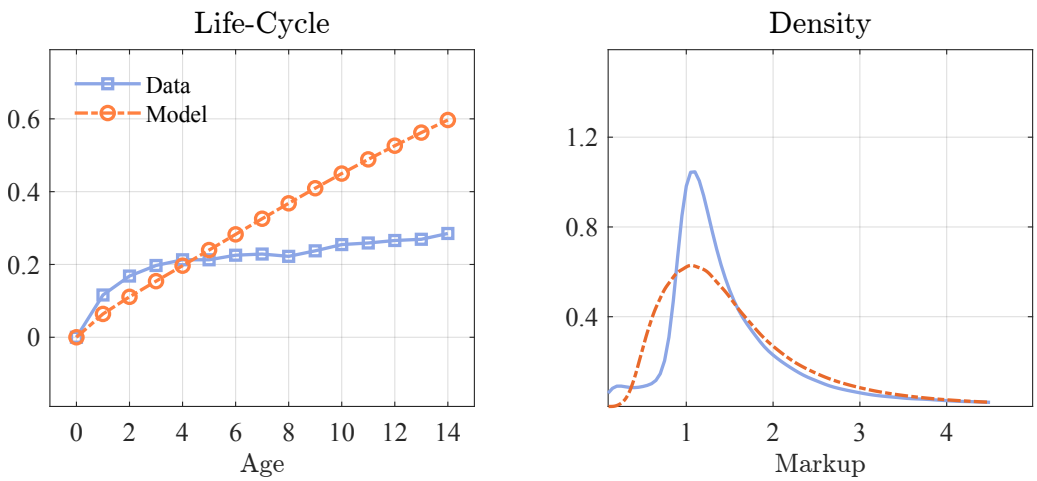
[6] Estimates for vary between 0.04 for physical costs and 0.09 for customer costs, see for example Levy et al. (1997) and Zbaracki et al. (2004). As in Golosov and Lucas (2007) and Baley and Blanco (2019), we choose a value in between those.

[7] We use Compustat Data between 1990Q1 and 2016Q4. For the empirical definition of markups, see Section 2.

that markups increase with the age of the firm, and argued that such behavior may in principle be due to the fact that, as businesses advance over their life-cycle, they are also able to establish their position in their respective markets and progressively accumulate demand for their products. This in turn allows producers to progressively charge higher prices and hence set higher markups. Accordingly, the left panel in Figure 6 reports the pattern of markups over firm's life-cycle both in the model and in the data. In particular, we remind the reader that the empirical series has been computed using Compustat data between 1990Q1 and 2016Q4 and netting out sector and time FE.

On the one hand, the model slightly underestimates the rapid increase of markups in the first 5 years of a firm's life, whereas it tends to modestly overestimate their subsequent growth in the next years.[8] On the other hand, our calibrated framework can replicate qualitatively the growth of markups over firm age and match more than half of the quantitative features of the relationship between markups and the life-cycle of producers. Importantly, it needs to be stressed that the ability of the model to imply life-cycle markups' properties consistent with the empirical observations will prove crucial when assessing the differential response of firms to interest rate shocks. In fact, as documented in Section 2, old firms' markups show a more countercyclical response after a negative MP shock: absent the fit of the life-cycle profile of markups, our model would then not be able to replicate the heterogeneous response of markups to a MP shock according to firms' relative age.

Figure 6: Markups Steady State Properties



Secondly, as illustrated in the right panel of Figure 6, we can reasonably match the entire distribution of markups estimated from Compustat data. In particular, while a couple of distributional properties have been indeed targeted in the calibration, the model itself delivers a fat right tail in the distribution of markups consistent with our empirical observations and with the analysis of De Loecker et al. (2020). As reported in Table 3, our quantitative framework implies that the bottom 25% firms in the distribution have an average markup of 1.15, against an empirical value of 1.03 computed in the data, while a similar fit holds for the top 75% firms. Matching the right proportions of high and low-markup firms' will prove crucial when comparing the response of firms' markups to a monetary policy shock across companies that are below or above the median age.

[8]The fit is very precise during the first years of business operations which is due to the fact that, in our calibration, we target the mean reversion in the demand process ρ_a to match the growth of markups for firms aged 0 to 5.

Table 3: Distributional Properties of Markups

	Model	Data
Bottom 25% Firms	1.15	1.03
Top 75% Firms	1.79	1.86

4.2.2 The Link between Wages and Sales

While the superelasticity parameter ω has been identified by computing the elasticity of markups to sales shares, our calibrated model has also a testable prediction on the relationship between the wage bill and the sales of firms, which we can match as an untargeted dimension. In particular, recall that markups are a measure of whether firms can set prices above their marginal costs. Similarly to Edmond et al. (2018), in our theoretical set up the salaries paid by firm i hence depend on its sales and markup according to a simple expression given by:

$$\text{wage bill} = \frac{\text{sales}}{\text{markup}}$$

Moreover, if the superelasticity ω in the Kimball aggregator was equal to zero as in the standard NK model, markups would not increase with firm sales and, in turn, the wage bill shares would increase one-for-one with sales shares. But if ω is positive, as in our framework, markups do increase with firms sales implying that the wage bill increases less than one-for-one with sales. In this sense, both empirically and quantitatively, the extent to which the wage bill share of firms increases with their sales shares can therefore be linked to the extent to which markups increase with producers' size. A small caveat to keep in mind is that Compustat does not report a precise measure for firms' wage bills but only a balance sheet item related to the cost of goods sold. This variable comprises the cost of all variable inputs used in production, included (but not exclusively) labor. Nevertheless, we exploit the available data to run the following regression:

$$\log(\textit{wage bill shares})_{i,t} = \beta * \log(\textit{sales shares})_{i,t} + \varphi_{s,t} + \varepsilon_{i,t} \tag{15}$$

where $\varphi_{s,t}$ are sector-time FE and the coefficient β precisely informs by how much variable input costs are linked to firms' sales shares. A value of the elasticity $\beta < 1$ confirms the fact that, absent perfect competition – as in our model –, firms increase sales by increasing prices, thereby suppressing produced quantities. In turns, this mechanism implies that growing firms also demand less employment, which creates a wedge such that wage bill shares do not move one to one with sales shares. The results of the empirical estimation and quantitative fit are reported in Table 4.

4.2.3 Cross-sectional and Life-Cycle Properties

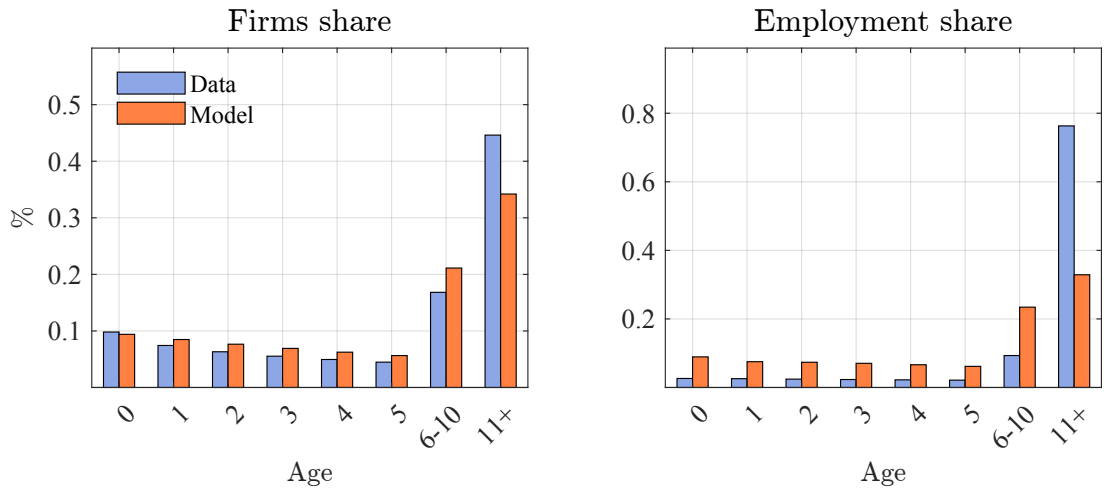
In our last exercise, we analyse the distribution of firms by age and the life-cycle profile of both employment and sales growth rates for the businesses in our model economy. In Figure 7, we report the distribution of firms and employment shares by age, comparing the empirical ones from Compustat

Table 4: Estimated Relationship between Wages and Sales

	Model	Data
Elasticity of Wage Bill Shares to Sales Shares	0.87	0.88

(1990Q1-2016Q4) to the ones obtained in our quantified framework. Note that none of these distributions was targeted in the calibration of the model, and hence both comparisons are to be considered as a pure validation exercise. First, focusing on the left panel, one can observe that our framework succeeds in replicating the distribution of firms over their age, and only partially underestimates the share of businesses that are 11+ years old. In this sense, as most of our empirical analysis is highly focused on markups' properties over the life-cycle of firms, it is remarkably important that we are able to capture the correct number of firms per age bin. In fact, the share of companies in each age bin influences the heterogeneous response of markups' to monetary policy shocks, and hence is relevant to get a correct quantitative fit of the empirically estimated dynamics of markups by firms' age.

Figure 7: Distributions of Firms and Employment Shares by Age

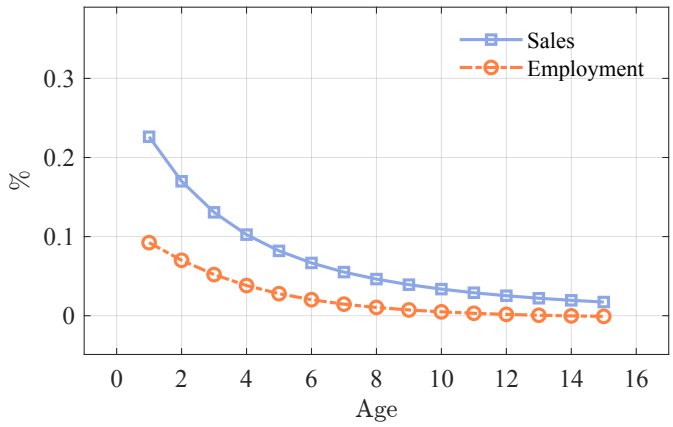


Secondly, in the right panel of Figure 7, we plot the distribution of employment shares over firm age, comparing the empirical ones with their model-implied counterparts. As it becomes clear from the graph, our framework performs moderately worse and is able to match up to half of the right tail in the employment share distribution. This is precisely due to the fact that, in the model economy, big firms (and hence old firms), find optimal to increase sales by increasing prices, thereby suppressing produced quantities and employment demand. This mechanism is a key characteristic of our set up in which companies operate in an environment with imperfect competition, and it is hence responsible for the fact that 11+ years old firms in the model generate a lower employment share compared to their empirical counterpart. Nonetheless, from this particular validation exercise we are still able to get a satisfactory fit of both firms and employment share distributions over the age of the businesses.

As a final note, in Figure 8 we plot the average employment and sales growth rates over the life-cycle of firms. Understandably, both measures decrease over time, as companies become old and hence slow down in their growth processes: this means that growth rates are unconditionally

negatively correlated with age, as empirically noted in Dunne et al. (1989). However, sales grow relatively more than employment, which is indeed consistent with the early discussion related to the employment share distribution depicted in the right panel of Figure 7. In particular, as argued in the previous paragraphs and due to the presence of the Kimball aggregator, markups do increase with firms sales, implying that the wage bill increases less than one-for-one with sales, depressing the labor demand by firms and resulting in lower employment growth rates compared to the growth rate of firm's sales. In other words, due to market power, companies can increase sales by raising prices and decreasing output, which lowers their demand of labor and hence employment growth.

Figure 8: Employment and Sales Growth Rates



5 Results

In the following section, we begin by discussing the response of firm markups to interest rate shocks, and compare the relative response of old and young firms in the model with the ones obtained in the data and reported in Section 2. Secondly, having assessed how much of the heterogeneity in markups response to interest rates by age our model is able to replicate, we also illustrate by how much the changes of aggregate variables such as output and wages after a MP shock contribute to the differential response of markups of old firms with respect to young ones. Finally, we conclude by analysing the amplification of shocks at work in our framework compared to a one-firm NK model.

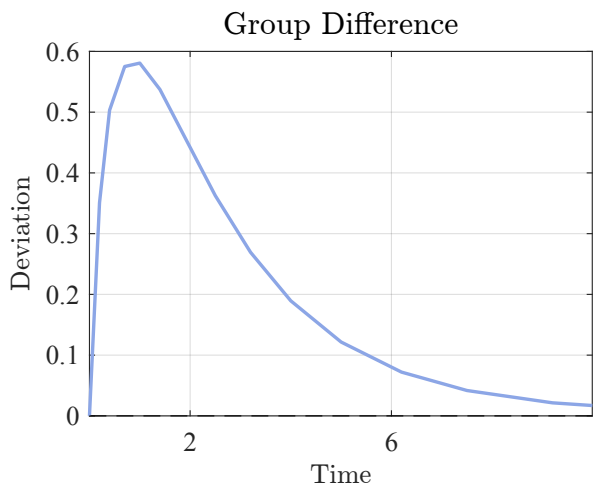
5.1 Response of Markups to Monetary Policy Shocks

We proceed to illustrate the dynamics of the economy after the arrival of a negative MP shock and to compare the relative response of old firms and young firms' markups to the ones obtained from the data and discussed in Section 3. As standard in frameworks characterized by nominal rigidities, a negative MP shock features an increase in the nominal interest rate and implies a downwards pressure on the labor cost W. Parallel to that, both employment, consumption and output decrease on impact and slowly recover as the shock fades away, while the downwards pressure on prices determines a deflationary episode. Moreover, the aggregate markup increases as a result of decreasing labor costs, and hence shows a countercyclical behavior in response to negative shocks to the nominal interest rate. The aggregate response of our calibrated economy hence resembles qualitatively the

one of a standard NK textbook model, as in Galí (2015). However, the aggregate pattern of markups masks a noticeable degree of heterogeneity at the firm-level which we explore in what follows.

To obtain a comparable set up to our empirical analysis, we first categorize firms in our model economy by their age decile and then classify all businesses above the median age as "old" and below the median age as "young". We then simulate the hit of a negative MP shock, otherwise defined as an exogenous increase in the nominal interest rate. Similarly to the empirical analysis in Figure 1, we then compute the differential response of markups to a MP shock for firms above the median age compared to firms below the median age. Figure 9 plots the differential response of markups by firm age over a horizon of several quarters and in deviation from the mean response. Clearly, firms above the median age respond more countercyclically to a negative MP shock compared to businesses below the median age, consistent with the empirical evidence from Compustat data.

Figure 9: Markups IRFs After a Negative MP Shock



Moreover, the differential response of old firms markups upon a negative MP shock in the model peaks at a value of 0.6%, while empirically it goes up to 3%. Importantly then, our quantitative framework is able to replicate 20% of the excess counter-cyclicality of old firms' markups to MP shocks that we have estimated in Compustat. In turn, this represent a satisfactory quantitative validation of our framework, which is hence able to replicate both qualitatively and quantitatively the heterogeneity in the response of markups by firm age to MP shocks that has been documented in the data.

5.2 Decomposing the Differential Response of Markups

Analytical Result under Flexible Prices. In what follows, we show that the combination of total derivatives of the demand function and the desired markup respectively gives the opportunity to understand the heterogeneous response of firm prices to changes in aggregate output Y, wage W and the demand index $\mathcal{D}$. For the sake of analytical tractability, we first carry out such decomposition in a version of the model without price adjustment costs. As derived in Section 3, the demand function in our theoretical framework is given by:

$$y = \left(1 - \omega \log\left(\frac{\sigma}{\sigma - 1}\frac{1}{a}\frac{p}{\mathcal{D}}\right)\right)^{\sigma/\omega}\frac{Y}{a}$$

while the desired markup can be written as follows:

$$\frac{\alpha p}{W y^{\frac{1}{\alpha}-1}} = \frac{\sigma \left(\frac{y}{Y} a\right)^{-\omega/\sigma}}{\sigma \left(\frac{y}{Y} a\right)^{-\omega/\sigma} - 1} \equiv \mu(a)$$

where $\mu(a)$ denotes the markup and increases in the demand faced by the firm. From these two equations, we can derive a set of expressions linking the change in firm prices (and similarly markups) to changes in aggregates W, Y and $\mathcal{D}$ and model parameters (see the derivations in the Appendix):

$$\frac{\partial \log p}{\partial \log Y} = \frac{\frac{1}{\alpha} - 1}{1 + \left(\frac{1}{\alpha} - 1\right) \frac{\mu(a)}{\mu(a)-1} + \frac{\omega}{\sigma} \mu(a)}$$

$$\frac{\partial \log p}{\partial \log W} = \frac{1}{1 + \left(\frac{1}{\alpha} - 1\right) \frac{\mu(a)}{\mu(a)-1} + \frac{\omega}{\sigma} \mu(a)}$$

$$\frac{\partial \log p}{\partial \log \mathcal{D}} = \frac{\left(\frac{1}{\alpha} - 1\right) \frac{\mu(a)}{\mu(a)-1} + \frac{\omega}{\sigma} \mu(a)}{1 + \left(\frac{1}{\alpha} - 1\right) \frac{\mu(a)}{\mu(a)-1} + \frac{\omega}{\sigma} \mu(a)}$$

where the standard CES equivalent, in the case of perfect competition, can be obtained setting the Kimball superelasticity parameter $\omega = 0$. Notice that all derivatives are positive, which means that the negative MP shock negatively affects W, Y and $\mathcal{D}$. Moreover, the first two derivatives are decreasing in $\mu(a)$ and the third one increases in $\mu(a)$. The same observations hold true if we were to write the derivative of firm's markup with respect to Y, W and $\mathcal{D}$. At the same time, the second derivatives with respect to the demand faced by the firm are given by:

$$\frac{\partial^2 \log p}{\partial \log Y \partial a} < 0, \quad \frac{\partial^2 \log p}{\partial \log W \partial a} < 0, \quad \frac{\partial^2 \log p}{\partial \log \mathcal{D} \partial a} > 0$$

The signs of these second derivatives imply that the prices of firms facing higher demand decline less after the MP shock due to the effects coming from the decline in W and Y, whereas prices decline more due to the decline in $\mathcal{D}$. The aggregate effect prevailing in GE will then depend on the specific parametrization. It is important to stress that, while these derivatives have been taken with respect to the demand a face by firms, there is a strong correlation and direct mapping between the accumulation of demand and firm age progression. This ensures that we can safely interpret the above results as the effects of the changes in Y, W, $\mathcal{D}$ after a MP shock on the prices of relatively older firms.

Benchmark Economy. The same decomposition is then carried out in practice in the quantitative model with nominal rigidities. In particular, we first compute numerically the general equilibrium response of the economy to a negative MP shock. Then, taking as given the equilibrium paths for the aggregate variables $Y, W, \mathcal{D}, r, \pi$ we look at the partial responses of old firms' markups to each of the shocks separately. Before commenting on the quantitative results, we follow the same spirit as in Kaplan et al. (2018) and provide intuition for the channels at play in our fully-fledged economy with heterogeneous firms and endogenous markups. Let us first write the difference between the average markups of old firms and the average markups of young firms as a function of the equilibrium prices, quantities, and inflation. We collect these terms in the vector $\{\mathcal{S}_t\}_{t \geq 0}$, with $\mathcal{S}_t = \{r_t, W_t, Y_t, \mathcal{D}_t, \pi_t\}$,

and define the above-mentioned difference $\widehat{\mathcal{M}}(\{\boldsymbol{\mathcal{S}}_t\}_{t\geq 0})$ induced by the path of the monetary shock $\{\varepsilon_t\}_{t\geq 0}$ from its initial hit until its fully reverts to zero as:

$$\widehat{\mathcal{M}}(\{\mathcal{S}_t\}_{t\geq 0}) := \int \mu_t(p,a;\{\mathcal{S}_t\}_{t\geq 0})\mathbb{1}\{g_t(p,a) \geq \bar{a}\}d\lambda_t - \int \mu_t(p,a;\{\mathcal{S}_t\}_{t\geq 0})\mathbb{1}\{g_t(p,a) < \bar{a}\}d\lambda_t. \quad (16)$$

where $\mu_t(p,a;\{\mathcal{S}_t\}_{t\geq 0})$ is the firm markup, $g_t(p,a)$ is a mapping between firm's states and its age, $\bar{a}$ is the median firms' age, and $d\lambda_t(p,a;\{\mathcal{S}_t\}_{t\geq 0})$ is the joint distribution of prices and idiosyncratic demand. Totally differentiating Equation 16, we decompose the difference in the average markup response between old and young firms at time $t = \tau$ as:

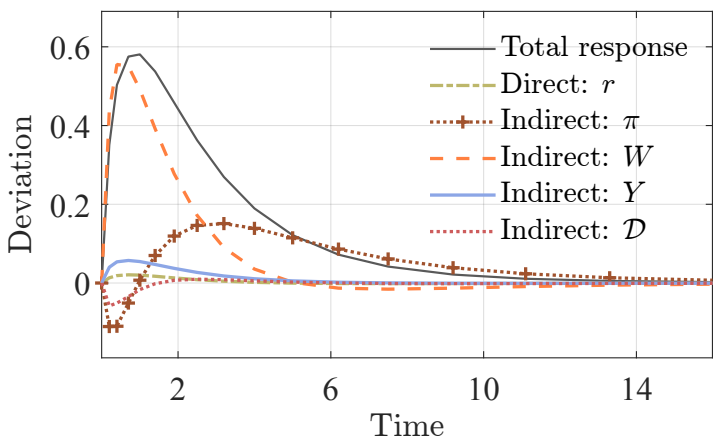
$$d\widehat{\mathcal{M}}_\tau = \underbrace{\int_\tau^\infty \frac{\partial \widehat{\mathcal{M}}_\tau}{\partial r_t} dr_t dt}_{\text{direct effect}} + \underbrace{\int_\tau^\infty \left(\frac{\partial \widehat{\mathcal{M}}_\tau}{\partial W_t} dW_t + \frac{\partial \widehat{\mathcal{M}}_\tau}{\partial Y_t} dY_t + \frac{\partial \widehat{\mathcal{M}}_\tau}{\partial \mathcal{D}_t} d\mathcal{D}_t + \frac{\partial \widehat{\mathcal{M}}_\tau}{\partial \pi_t} d\pi_t \right) dt}_{\text{indirect effect}} \quad (17)$$

where the first term reflects the direct effect of a change in the interest rate, which enters the Euler equation of the agents, holding the other variables of interest constant. The remaining terms in the decomposition reflect the indirect effects of changes in inflation, the real wage, real output and the demand index that arise in general equilibrium after the hit of the MP shock. In practice, we need to compute each of these components numerically. For example, the formal definition of the first term in Equation 17, which is the direct effect of changes in the real interest rate $\{r_t\}_{t\geq 0}$, is:

$$\int_\tau^\infty \frac{\partial \widehat{\mathcal{M}}_\tau}{\partial r_t} dr_t dt = \int_\tau^\infty \frac{\partial \widehat{\mathcal{M}}(\{r_t, \overline{W}, \overline{Y}, \overline{\mathcal{D}}, \overline{\pi}\}_{t\geq 0})}{\partial r_t} dr_t dt. \quad (18)$$

This term is the *partial-equilibrium* response of the difference in the average markups between old and young firm that face a time-varying real interest rate path $\{r_t\}_{t\geq 0}$, but holding the paths for the real wage $\overline{W}$, the real output $\overline{Y}$, the demand index $\overline{\mathcal{D}}$, and nominal inflation rate $\overline{\pi}$ constant at their steady-state values. We calculate this term from the model by feeding these time paths into the firms' (and household's) optimization problem, computing the policy function and their markups for each firms, and aggregating across firms using the corresponding distribution. The other terms in the decomposition are computed in a similar fashion.

Figure 10: Decomposing the Differential Response of Markups

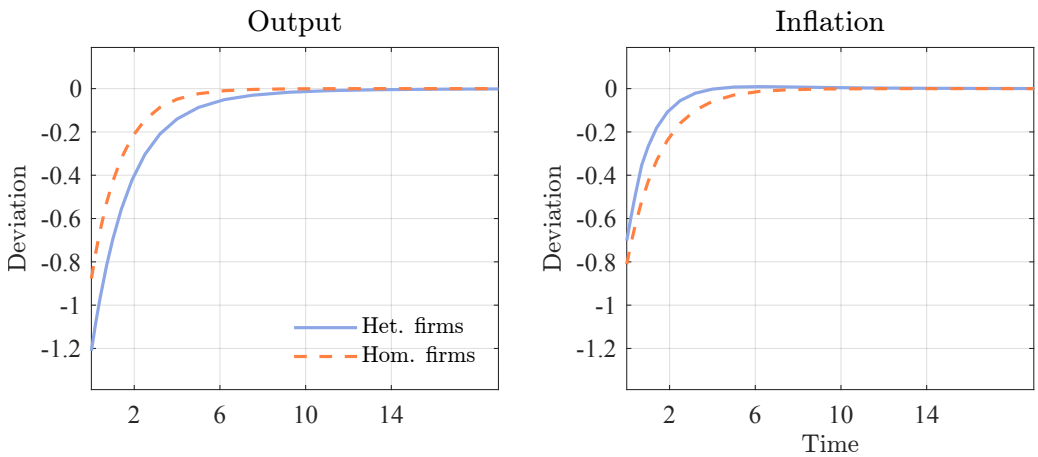


The results of the quantitative exercise are depicted in Figure 10 below. All effects are to be intended as p.p. deviations from the mean response across all firms in the economy, and we also note that the GE effect is not a direct sum of the partial effects due to non-linearities in aggregation. The outer dark line represents the total GE effect of a negative MP shock on the differential impulse response of markups for old firms compared to young ones. Most of the resulting effect on the differential response of old firms' markups is to be attributed to changes in the aggregate W, hence to changes in the cost of labor after a negative shock to the interest rate. Since our model features heterogeneous and endogenous markups in the presence of a Kimball aggregator, big firms (and hence old firms) have a lower passthrough from production costs to prices. In this sense a negative MP shock in the economy puts a downward pressure on the labor input cost W, but old firms' sales react less than proportionally, as dominant companies do not decrease prices as much. Since markups are the ratio between business sales and costs, the resulting effect on markups is positive, leading to the observed stronger countercyclical response of old firms' markups to a negative MP shock.

5.3 Amplification Mechanism

In what follows, we conclude the quantitative analysis by studying the shock amplification mechanism at work in our economy, comparing our calibrated framework with a standard one-firm NK model. As pointed out in Mongey (2017), in economies where real rigidities are present, shocks have a strong propagation through quantities, which we set to verify in our case. Moreover, we proceed to also explain to which extent both firm heterogeneity and the Kimball aggregator that characterize our model can be responsible for greater swings in macro aggregates after a negative MP shock.

Figure 11: Comparing Output and Inflation Responses



To ensure we are working with two comparable economies, we first calibrate the one-firm NK economy to have the same size as in our heterogeneous firms framework (hereafter the FDNK) in terms of overall output produced. Moreover, since the standard NK model features perfect competition, we set the elasticity of substitution σ in its CES aggregator to match an aggregate markup of 1.68, which is the value targeted in our FDNK model under the Kimball aggregator. With the two models at hand, we simulate a negative MP shock in both economies and solve for the response of the main macroeconomic aggregates in the two economies. In particular, we analyse the trajectories of

inflation π and output Y over an 16-quarters period and compare their relative percentage deviations from steady state values. The results of this exercise are depicted in Figure 11.

Comparing output and inflation responses across the two models, it is clear that the negative MP shock produces a bigger drop in output and a milder decline in prices in our FDNK set up compared to a standard one-firm NK model. The negative change in the interest rate decreases output by on average 20 p.p. more in the economy characterised by heterogeneous firms and endogenous markups, with the effect lasting for more than 10 quarters after the shock hits. At the same time, prices and hence inflation drop by relatively more in the one-firm NK model, which implies that the presence of the Kimball aggregator and the differential passthrough that characterize our model economy mitigate the downward pressure exerted by the negative MP shock on firm prices.

On the one hand, as argued in Klenow and Willis (2016), the presence of the Kimball aggregator adds a source of real frictions in the NK model, represented by a higher degree of concavity in the firm's profit function with respect to its relative price. Under the Kimball aggregator, sellers face a price elasticity of demand that is increasing in their good's relative price. For instance, when a repricing producer faces a lower wage after a negative MP shock, it will temper its price drop because of the endogenous increase in its desired markup, and this effect would be stronger the lower the elasticity of demand faced by the producer. Since the presence of a real rigidity makes firms more reluctant to change prices, firms do not pass marginal cost shocks as fully onto their prices as they would in a standard NK model with a CES aggregator. Hence, in our FDNK set up, MP shocks propagate more through quantities than prices, and decrease aggregate output by relatively more.

On the other hand, without heterogeneity on the firm's side, the presence of the Kimball aggregator alone does not automatically imply the amplification of shocks in our setup: in fact, the effects of the real rigidity introduced by the Kimball aggregator kick in only when businesses are indeed heterogeneous and hence present a different passthrough from costs to prices with respect to one another. If all firms were to be equal (as in the representative-firm NK model), they would also be equal to the mean and have identical sales shares. Specifically, focusing on Equation 8, the elasticity of demand faced by producers would not vary across firm, and their response to MP shocks would be identical. On the contrary, in our FDNK set up, since big firms (hence old firms) respond more countercyclically than small ones and decrease their prices by less, the propagation of a negative shock gets strengthened. Hence, the heterogeneity of firms, combined with the real rigidity introduced by the Kimball aggregator set up, delivers the amplification mechanism at work in the present model.

6 Conclusion

In this paper, we have taken an empirical and theoretical approach to the study of firm heterogeneity in the response of markups to MP shocks. In order to carry out our data analysis, we have merged exogenously-identified monetary policy shocks series with a rich quarterly dataset comprising publicly-listed companies based in the US between 1990Q1 and 2016Q4. Next, we have documented that old firms' markups tend to increase after a monetary policy tightening, while young firms' markups show a mildly procyclical behavior after a negative interest rate shock. Moreover, our empirical investigation seems to also suggest that the differential response of markups by firm's

age could be related to the accumulation of customers and demand over time, which enables older firms to change by relatively less their prices thanks to an established position in their markets.

In our quantitative analysis, we have embedded our findings into a NK model, augmented with heterogeneous firms and a process of demand accumulation, and in which markups arise endogenously and evolve over the life-cycle of the companies. Our calibrated framework can replicate the life-cycle profile of firms' markups and growth rates, and the distribution of companies and employment shares by corporate age. Moreover, we were able to explain up to a fifth of the empirically estimated excess counter-cyclicality in the markups of firms above the median age after a negative monetary policy shock. Finally, we have shown that both firms' heterogeneity and endogenous markups generate amplification in the response of aggregate quantities to contractionary interest rate movements, which further distinguishes our set up from standard frameworks with nominal rigidities. In the future, we aim to further study optimal monetary policies in the presence of imperfect competition, demand accumulation, and heterogeneity in the passthrough from costs to prices.

References

Andrea Alati. *Essays on firms heterogeneity and business cycles*. PhD thesis, The London School of Economics and Political Science (LSE), 2020.

Andrew Atkeson and Ariel Burstein. Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, 98(5):1998–2031, 2008.

Adrien Auclert. Monetary policy and the redistribution channel. *American Economic Review*, 109(6):2333–67, 2019.

Isaac Baley and Andrés Blanco. Firm uncertainty cycles and the propagation of nominal shocks. *American Economic Journal: Macroeconomics*, 11(1):276–337, 2019.

David Baqaee, Emmanuel Farhi, and Kunal Sangani. The supply-side effects of monetary policy. Technical report, National Bureau of Economic Research, 2021.

Florin O Bilbiie, Fabio Ghironi, Marc J Melitz, Virgiliu Midrigan, and Julio J Rotemberg. Monetary policy and business cycles with endogenous entry and product variety [with comments and discussion]. *NBER Macroeconomics Annual*, 22:299–379, 2007.

Florin O Bilbiie, Fabio Ghironi, and Marc J Melitz. Endogenous entry, product variety, and business cycles. *Journal of Political Economy*, 120(2):304–345, 2012.

Mark Bils, Peter J Klenow, and Benjamin A Malin. Resurrecting the role of the product market wedge in recessions. *American Economic Review*, 108(4-5):1118–46, 2018.

Ariel Burstein, Vasco M Carvalho, and Basile Grassi. Bottom-up markup fluctuations. Technical report, National Bureau of Economic Research, 2020.

James Cloyne, Clodomiro Ferreira, Maren Froemel, and Paolo Surico. Monetary policy, corporate finance and investment. Technical report, National Bureau of Economic Research, 2018.

Jan De Loecker and Frederic Warzynski. Markups and firm-level export status. *American economic review*, 102 (6):2437–71, 2012.

Jan De Loecker, Jan Eeckhout, and Gabriel Unger. The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2):561–644, 2020.

Emin Dinlersoz, Sebnem Kalemli-Ozcan, Henry Hyatt, and Veronika Penciakova. Leverage over the life cycle and implications for firm growth and shock responsiveness. Technical report, National Bureau of Economic Research, 2018.

Timothy Dunne, Mark J Roberts, and Larry Samuelson. The growth and failure of us manufacturing plants. *The Quarterly Journal of Economics*, 104(4):671–698, 1989.

Chris Edmond, Virgiliu Midrigan, and Daniel Yi Xu. How costly are markups? Technical report, National Bureau of Economic Research, 2018.

Andrea Fabiani, Luigi Falasconi, and Janko Heineken. Monetary policy and corporate debt maturity. 2020.

Jordi Galí. *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press, 2015.

Jordi Gali, Mark Gertler, and J David Lopez-Salido. Markups, gaps, and the welfare costs of business fluctuations. *The review of economics and statistics*, 89(1):44–59, 2007.

Mark Gertler and Simon Gilchrist. Monetary policy, business cycles, and the behavior of small manufacturing firms. *The Quarterly Journal of Economics*, 109(2):309–340, 1994.

Simon Gilchrist, Raphael Schoenle, Jae Sim, and Egon Zakrajšek. Inflation dynamics during the financial crisis. *American Economic Review*, 107(3):785–823, 2017.

Mikhail Golosov and Robert E Lucas. Menu costs and phillips curves. *Journal of Political Economy*, 115(2): 171–199, 2007.

Refet S Gürkaynak, Brian Sack, and Eric Swanson. The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models. *American economic review*, 95(1):425–436, 2005.

Robert E Hall. The relation between price and marginal cost in us industry. *Journal of political Economy*, 96(5): 921–947, 1988.

John Haltiwanger, Ron S Jarmin, and Javier Miranda. Who creates jobs? small versus large versus young. *Review of Economics and Statistics*, 95(2):347–361, 2013.

Sungki Hong. Customer capital, markup cyclicality, and amplification. *FRB St. Louis Working Paper*, (2017-33), 2017.

Marek Jarociński and Peter Karadi. Deconstructing monetary policy surprisesâ€"the role of information shocks. *American Economic Journal: Macroeconomics*, 12(2):1–43, 2020.

Priit Jeenas. Firm balance sheet liquidity, monetary policy shocks, and investment dynamics. 2019.

Òscar Jordà. Estimation and inference of impulse responses by local projections. *American economic review*, 95 (1):161–182, 2005.

Greg Kaplan, Benjamin Moll, and Giovanni L Violante. Monetary policy according to hank. *American Economic Review*, 108(3):697–743, 2018.

Peter J Klenow and Jonathan L Willis. Real rigidities and nominal price changes. *Economica*, 83(331):443–472, 2016.

Daniel Levy, Mark Bergen, Shantanu Dutta, and Robert Venable. The magnitude of menu costs: direct evidence from large us supermarket chains. *The Quarterly Journal of Economics*, 112(3):791–824, 1997.

Alisdair McKay, Emi Nakamura, and Jón Steinsson. The power of forward guidance revisited. *American Economic Review*, 106(10):3133–58, 2016.

Matthias Meier and Timo Reinelt. Monetary policy, markup dispersion, and aggregate tfp. 2020.

Simon Mongey. Market structure and monetary non-neutrality. *Job market paper. http://www. simonmongey. com/uploads/6/5/6/6/65665741/mongey_market_structure_monetary_nonneutrality_draft.pdf*, 2017.

Christopher J Nekarda and Valerie A Ramey. The cyclical behavior of the price-cost markup. *Journal of Money, Credit and Banking*, 52(S2):319–353, 2020.

Pablo Ottonello and Thomas Winberry. Financial heterogeneity and the investment channel of monetary policy. *Econometrica*, 88(6):2473–2502, 2020.

Benjamin W Pugsley, Petr Sedlacek, and Vincent Sterk. The nature of firm growth. *Available at SSRN 3086640*, 2019.

John B Taylor. The robustness and efficiency of monetary policy rules as guidelines for interest rate setting by the european central bank. *Journal of Monetary Economics*, 43(3):655–679, 1999.

Arlene Wong. Refinancing and the transmission of monetary policy to consumption. *Unpublished manuscript*, 20, 2019.

Mark J Zbaracki, Mark Ritson, Daniel Levy, Shantanu Dutta, and Mark Bergen. Managerial and customer costs of price adjustment: direct evidence from industrial markets. *Review of Economics and statistics*, 86(2): 514–533, 2004.

Appendix

A Data Appendix

Figure A.1: Alternative Specification for Corporate Age

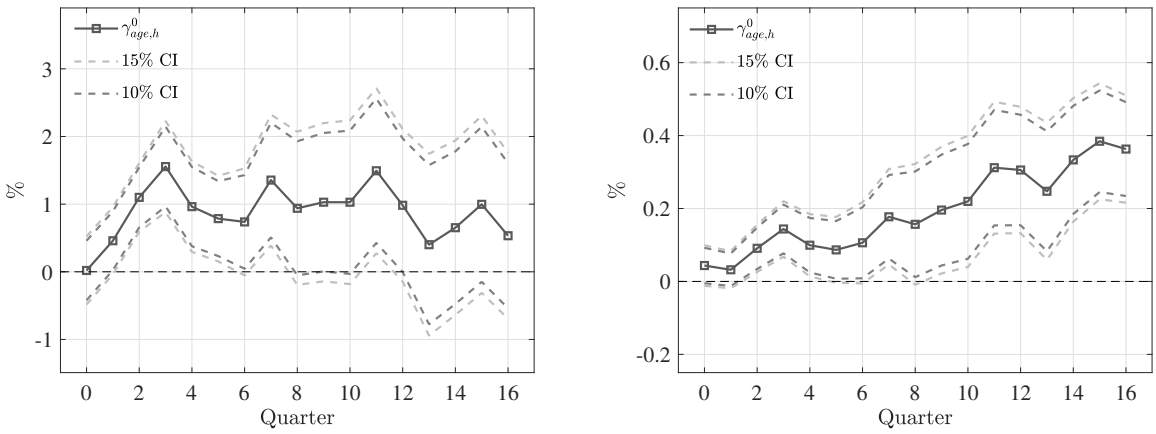


Figure A.2: Excluding Future Shocks (left) and Sector-Quarter FE (right)

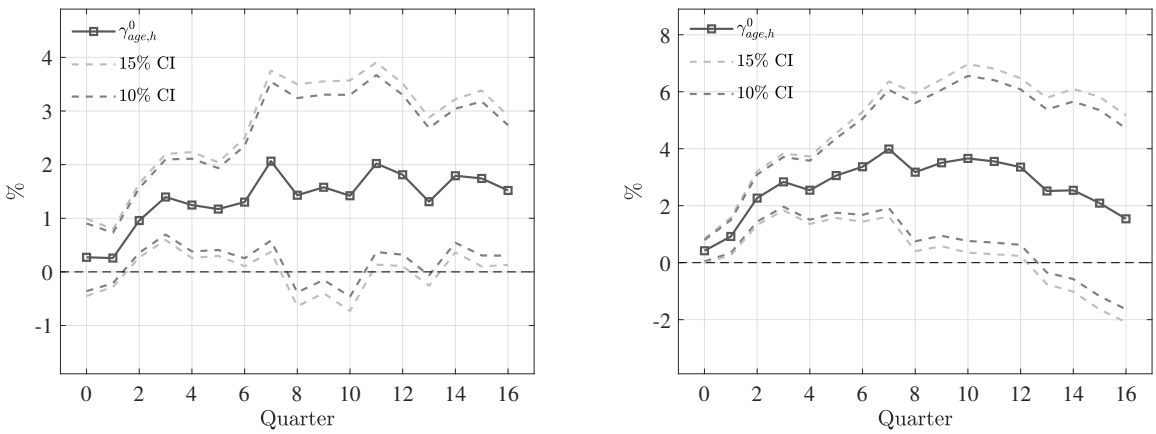


Figure A.3: Age and Leverage

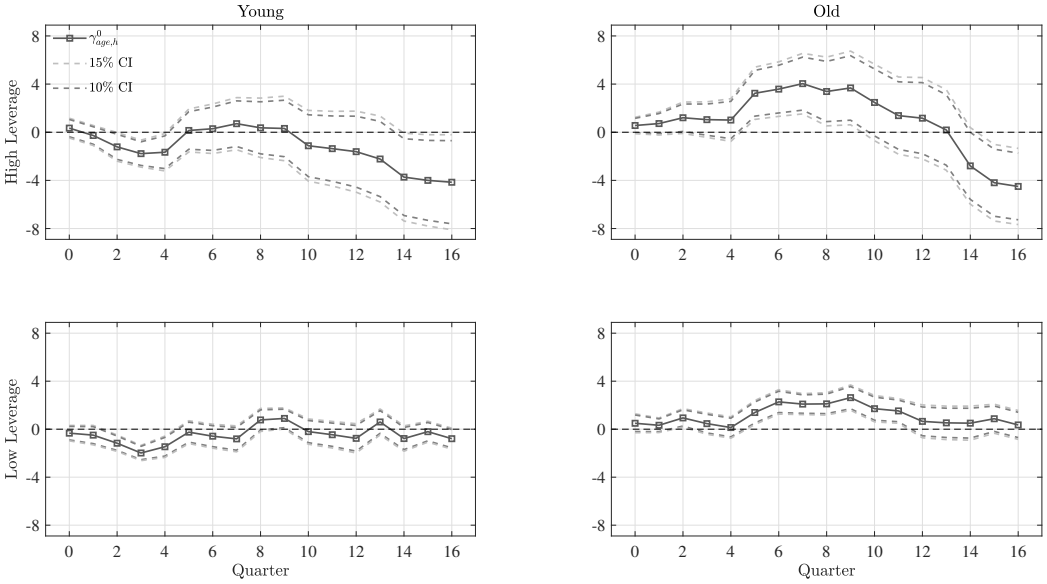


Figure A.4: Age and Liquidity

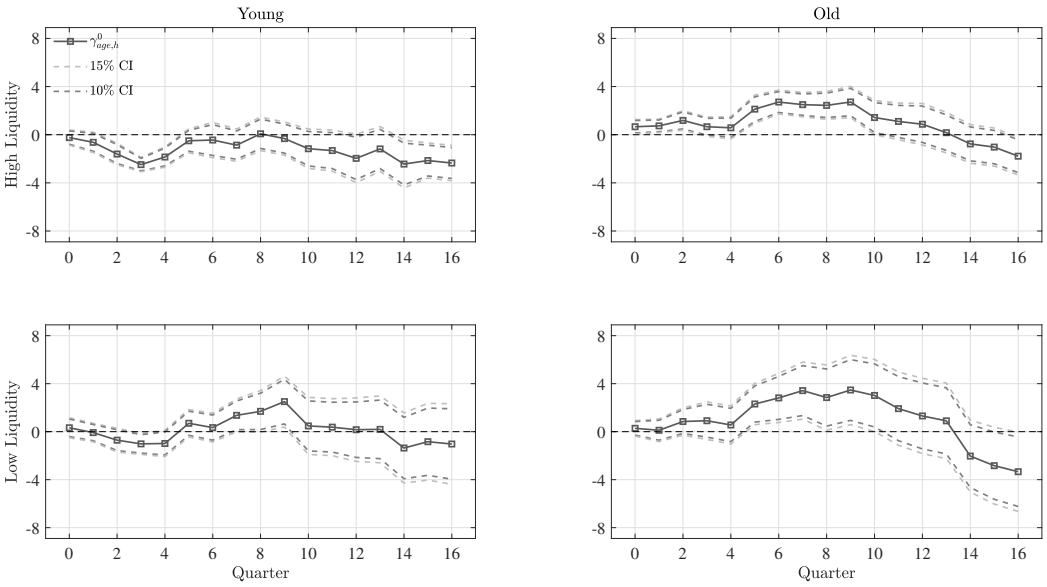
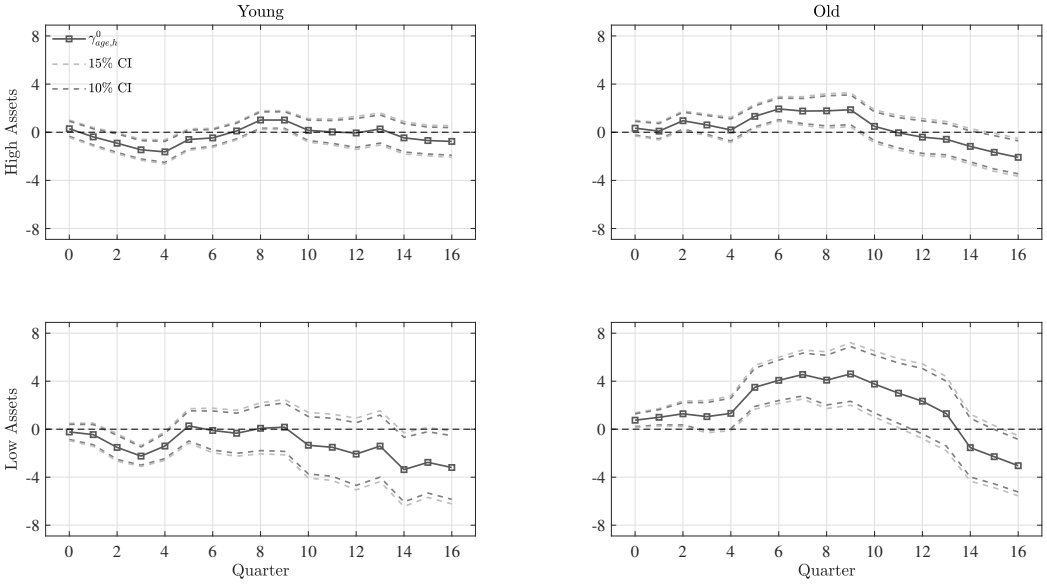


Figure A.5: Age and Size



B Quantitative Appendix

B.1 Decomposition Exercise: Derivations

The demand function in our model is given by:

$$y = \left(1 - \omega \log\left(\frac{\sigma}{\sigma - 1}\frac{1}{\xi(a)}\frac{p}{\mathcal{D}}\right)\right)^{\sigma/\omega} \frac{Y}{\xi(a)}$$

Which has the total derivative:

$$dlog\frac{y}{Y}\xi(a) = -\frac{\sigma(dlogp - dlog\mathcal{D})}{1 - \omega\log\left(\frac{\sigma}{\sigma-1}\frac{1}{\xi(a)}\frac{p}{\mathcal{D}}\right)} = -\sigma\left(\frac{y}{Y}\xi(a)\right)^{-\omega/\sigma}(dlogp - dlog\mathcal{D})$$

The desired markup is instead defined as follows:

$$\frac{\alpha p}{Wy^{\frac{1}{\alpha}-1}} = \frac{\sigma\left(\frac{y}{Y}\xi(a)\right)^{-\omega/\sigma}}{\sigma\left(\frac{y}{Y}\xi(a)\right)^{-\omega/\sigma} - 1}$$

By taking the total derivative it is possible to get:

$$\frac{\alpha}{Wy^{\frac{1}{\alpha}-1}}dp - \frac{\alpha p}{Wy^{\frac{1}{\alpha}-1}}dlogW + \left(1 - \frac{1}{\alpha}\right)\frac{\alpha p}{Wy^{\frac{1}{\alpha}}}dy =$$
$$-\frac{1}{\left(\sigma\left(\frac{y}{Y}\xi(a)\right)^{-\omega/\sigma} - 1\right)^2}\sigma\left(-\frac{\omega}{\sigma}\right)\left(\frac{y}{Y}\xi(a)\right)^{-\omega/\sigma-1}d\frac{y}{Y}\xi(a)$$

Substituting in the above expression $dy = \frac{Y}{\xi(a)} d\frac{y}{Y}\xi(a) + \frac{y}{Y}dY = y(dlog\frac{y}{Y}\xi(a) + dlogY)$ it is possible to obtained the following equation:

$$\frac{\alpha p}{Wy^{\frac{1}{\alpha}-1}}dlogp - \frac{\alpha p}{Wy^{\frac{1}{\alpha}-1}}dlogW + \left(1 - \frac{1}{\alpha}\right)\frac{\alpha p}{Wy^{\frac{1}{\alpha}-1}}(dlog\frac{y}{Y}\xi(a) + dlogY) =$$
$$-\frac{1}{\left(\sigma\left(\frac{y}{Y}\xi(a)\right)^{-\omega/\sigma} - 1\right)^2}\sigma\left(-\frac{\omega}{\sigma}\right)\left(\frac{y}{Y}\xi(a)\right)^{-\omega/\sigma} dlog\frac{y}{Y}\xi(a)$$

which in turn implies:

$$dlogp - dlogW + \left(1 - \frac{1}{\alpha}\right)(dlog\frac{y}{Y}\xi(a) + dlogY) = \frac{1}{\left(\sigma\left(\frac{y}{Y}\xi(a)\right)^{-\omega/\sigma} - 1\right)}\left(\frac{\omega}{\sigma}\right) dlog\frac{y}{Y}\xi(a)$$

$$dlogp - \left(\left(\frac{1}{\alpha} - 1\right) + \frac{1}{\left(\sigma\left(\frac{y}{Y}\xi(a)\right)^{-\omega/\sigma} - 1\right)}\left(\frac{\omega}{\sigma}\right)\right) dlog\frac{y}{Y}\xi(a) = dlogW + \left(\frac{1}{\alpha} - 1\right) dlogY$$

Substituting $dlog\frac{y}{Y}\xi(a)$ from the total derivative of the demand function we get:

$$dlogp + \left(\left(\frac{1}{\alpha} - 1\right) + \frac{1}{\left(\sigma\left(\frac{y}{Y}\xi(a)\right)^{-\omega/\sigma} - 1\right)}\left(\frac{\omega}{\sigma}\right)\right)\sigma\left(\frac{y}{Y}\xi(a)\right)^{-\omega/\sigma}(dlogp - dlog\mathcal{D}) =$$
$$dlogW + \left(\frac{1}{\alpha} - 1\right) dlogY$$
$$dlogp + \left(\left(\frac{1}{\alpha} - 1\right)\frac{\mu(a)}{\mu(a) - 1} + \frac{\omega}{\sigma}\mu(a)\right)(dlogp - dlog\mathcal{D}) = dlogW + \left(\frac{1}{\alpha} - 1\right) dlogY$$
$$dlogp = \frac{dlogW + \left(\frac{1}{\alpha} - 1\right) dlogY + \left(\left(\frac{1}{\alpha} - 1\right)\frac{\mu(a)}{\mu(a)-1} + \frac{\omega}{\sigma}\mu(a)\right) dlog\mathcal{D}}{1 + \left(\frac{1}{\alpha} - 1\right)\frac{\mu(a)}{\mu(a)-1} + \frac{\omega}{\sigma}\mu(a)}$$

Where $\mu(a)$ is the markup. The above expression can be rearranged to get the relative contributions of Y, W and D to the change in prices and markups, reported in the main text.