Heterogeneous Markups Cyclicality and Monetary Policy[*]

Andrea Chiavari[†] Marta Morazzoni[‡] Danila Smirnov[§]

May 27, 2025

This paper revisits the question on the conditional cyclicality of the aggregate markup using a micro-to-macro approach, which highlights the role of firm-level heterogeneous cyclicality, the reallocation of economic activity across firms, and aggregation methods. Using US firm-level data from 1990 to 2016, we find that young firms have procyclical markups conditional on monetary shocks, while older firms show countercyclical markups. Moreover, economic activity reallocates from old to young firms after monetary shocks. Aggregating these responses, we find that the aggregate markup is countercyclical to monetary shocks. Over time, firm aging has changed the distribution of firms, altering the aggregate markup cyclicality, which help reconcile part of the conflicting findings in the literature.

Keywords: Heterogeneous Firms, Markups Cyclicality, Monetary Policy Shocks

JEL Codes: E32, E52, D22, L11

[*]We are deeply in debt to our advisors Isaac Baley, Davide Debortoli, and Edouard Schaal for their invaluable advice and support. We thank Andrea Caggese, James Cloyne, Maarten De Ridder, Andrea Fabiani, Andrea Ferrero, Jordi Galí, Basile Grassi, Priit Jeenas, François Le Grand, Virgiliu Midrigan, Benjamin Moll, Simon Mongey, Neil Mehrotra, Emi Nakamura, Galo Nuño, Franck Portier, Morten O Ravn, Jon Steinsson, and Gianluca Violante, as well as seminar participants at CREi, University of Barcelona, University of Kent, University of Bristol, University of Warwick, Federal Reserve Bank of Minneapolis, LSE, UCL, UC Berkeley, the 2022 BSE Summer Forum, and the 2022 conference on Individual Risks and the Macroeconomy at Sciences Po. The views expressed herein are those of the authors and should not be attributed to the IMF, its Executive Board, or its management.

[†]University of Oxford. Email: andrea.chiavari@economics.ox.ac.uk.

[‡]University College London, CEPR and Institute for Fiscal Studies. Email: m.morazzoni@ucl.ac.uk.

[§]International Monetary Fund. Email: dsmirnov@imf.org.

1 Introduction

The cyclicality of markups conditional on economic shocks plays a central role in shaping business cycles and for the transmission of monetary policy in sticky-prices New Keynesian models (e.g., Galí, 2015). In the logic of this framework, a countercyclical markup reflects firms' tendency to reduce prices during downturns and increase them during booms. This behavior moderates inflation during recessions and amplifies it during expansions, aligning inflation's cyclicality with GDP, as observed in the data. Despite the importance of this concept and the substantial attention received in the literature, consensus on the cyclicality of markups remains elusive (e.g. Bils, Klenow and Malin, 2018; Nekarda and Ramey, 2020).

In addition, due to the prevalence of standard representative-agent models and the lack of readily available data on firms' marginal costs, much of this literature has primarily focused on the behavior of different aggregate markup proxies. Yet, most price-setting decisions — and consequently, markup adjustments — occur at the firm level (e.g., Mongey, 2021). Therefore, this paper revisits the question of the cyclicality of the aggregate markup by adopting a micro-to-macro approach. We focus specifically on its cyclicality conditional on monetary policy (MP) shocks, and our strategy emphasizes the role of individual firms' markup cyclicality, the reallocation of economic activity across firms after a MP shock, and aggregation methods.

This approach yields three novel insights. First, after estimating firm-level markups using state-of-the-art IO measures on US data, we find that markups of old firms are countercyclical to MP shocks, whereas those of young firms are procyclical. Second, by properly aggregating these heterogeneous firm-level responses, we observe that the aggregate markup is countercyclical to MP shocks across our sample period. Third, we demonstrate that its cyclicality in response to MP shocks has evolved over time due to changes in the age distribution of US firms. This finding helps reconcile some of the seemingly conflicting results in the literature.

In practice, we begin by reviewing the conventional wisdom of sticky-price New Keynesian models regarding the role of aggregate markup cyclicality in response to demand shocks. We concentrate on demand shocks, as they are a prominent source of business cycle fluctuations in the literature. We then link this conventional view to the role of individual firms and propose a micro-to-macro approach that highlights the contributions of (i) firm-level markup cyclicality, (ii) the reallocation of economic activity across firms, and (iii) aggregation methods. In so doing, we give specific consideration to three dimensions of heterogeneity suggested by the literature that could directly inform the source and direction of firm-level markup cyclicality: financial frictions, demand frictions, and nominal pricing frictions.

We implement our strategy using quarterly Compustat data. Despite comprising only a subset of large firms in the US, this dataset offers several advantages. First, it allows the measurement of high-frequency markups at the firm level using state-of-the-art IO methods, as in De Loecker, Eeckhout and Unger (2020).[1] Second, firms in Compustat are among the largest in the economy and account for most of business cycle fluctuations (Crouzet and Mehrotra, 2020). Finally, it allows us to follow best practices in the literature by using high-frequency identified MP shocks from Jarociński and Karadi (2020) as a measure of demand shocks.

To identify the response of firm-level markups to MP shocks, we estimate local linear projections following Jordà (2005). Moreover, to analyze the heterogeneous response of markups across firms, we interact the MP shocks series with variables that capture various dimensions of firm-level heterogeneity emphasized by the literature. In particular, we account for financial frictions using common proxies such as firm age (Cloyne, Ferreira, Froemel and Surico, 2023), assets (Gertler and Gilchrist, 1994), leverage (Ottonello and Winberry, 2020), and liquidity (Jeenas, 2024). We also assess the role of demand frictions by controlling for firm relative size and markups. Finally, since nominal price frictions are only measured at the sector level, our empirical strategy includes sector-time fixed effects to control for this source of variation.

Our empirical findings reveal significant differences in the response of firm-level markups to MP shocks, particularly along the age dimension, while we find limited evidence supporting additional layers of heterogeneity. Specifically, old firms increase markups in response to a contractionary MP shock, indicating a countercyclical behavior, while young firms lower their markups, reflecting a procyclical response. We also estimate a small reallocation of economic activity from older to younger firms in response to MP shocks, consistent with the theoretical results in Baqaee, Farhi and Sangani (2024). Our evidence suggests that financial frictions may be playing a role, insofar as young firms tend to be more financially constrained (Cloyne, Ferreira, Froemel and Surico, 2023) and may need to reduce prices more during contractionary

[1] This measure is our baseline as it captures markup changes, our primary focus, as shown by De Ridder, Grassi and Morzenti (2024). However, to address potential identification concerns raised by Bond, Hashemi, Kaplan and Zoch (2021), we verify our findings using the accounting markup measure by Baqaee and Farhi (2020).

shocks, aligning with Kim (2021)'s findings on firms' prices. Yet, we observe that, while the literature has primarily interpreted age as a proxy for financial frictions, the significance of age for the heterogeneous cyclicality of firm markups could also be consistent with the existence of demand frictions. Specifically, old firms – more established in their markets – may not lower markups as much during a contractionary MP shock, as they face a more inelastic demand.

We then apply a micro-to-macro approach and use our firm-level empirical findings to study the aggregate markup cyclicality in response to MP shocks. This involves aggregating firm-level markups using theoretically consistent variable-costs weights, as in Grassi (2017) and Edmond, Midrigan and Xu (2023). Our findings indicate that, over the entire sample, the aggregate markup responds positively to contractionary MP shocks, demonstrating a countercyclical behavior consistent with the predictions of the sticky-price New Keynesian model.

Then, we show that changes in the age distribution of firms may indeed shape the cyclicality of the aggregate markup. A growing body of literature (e.g., Hathaway and Litan, 2014; Decker, Haltiwanger, Jarmin and Miranda, 2016) documents a significant decline in firm formation and in the exit rate of older firms, which contributes to the progressive aging of the US firm population. In our micro-to-macro approach, the relative sizes of young and old firm groups play a central role for aggregation. Assessing the impact of firm aging on the aggregate markup cyclicality over time, we find that at the beginning of our sample, when young, procyclical firms were more prevalent, the aggregate markup was cyclical leaning to procyclical. Conversely, in the latter part of the sample, when older, countercyclical firms were more prevalent, the aggregate markup became countercyclical — slightly more so than in the results based on the entire sample described earlier. To confirm our findings, we directly estimate the response of the aggregate markup to MP shocks over time, which leads to similar conclusions.

While this evidence is significant in itself, demonstrating that firm aging affects business cycle fluctuations beyond its impact on factor shares (Hopenhayn, Neira and Singhania, 2022) and long-term trends, it also offers a potential reconciliation for some of seemingly conflicting findings in the literature. Recent papers that analyze long data series (e.g., Nekarda and Ramey, 2020; Cantore, Ferroni and León-Ledesma, 2021), which predate the effects of firm aging, tend to find mildly procyclical markups. In contrast, studies focusing on shorter time series (e.g., Bils, Klenow and Malin, 2018), closer to the period when firm aging became more pronounced, identify a significant role for countercyclical markups. Therefore, our micro-to-macro approach — by explicitly accounting for aggregation — reveals the key role of the firm distribution in driving aggregate markup cyclicality, helping reconcile previous divergent findings.

Literature review. This paper contributes to empirical studies on the aggregate markup cyclicality. Key works that provide differing evidence based on aggregate data include Bils (1987), Galeotti and Schiantarelli (1998), Rotemberg and Woodford (1999), Bils and Kahn (2000), Gali, Gertler and Lopez-Salido (2007), Hall (2014), Bils, Klenow and Malin (2018), Cantore, Ferroni and León-Ledesma (2021), and Nekarda and Ramey (2020). Note that all of these studies focus on aggregate markup proxies, overlooking the role of across-firms reallocation and firm-level markups aggregation in shaping the overall response of this aggregate measure.

At the firm level, studies such as Hong (2017), Alati (2020), Burstein, Carvalho and Grassi (2020) and Afrouzi and Caloi (2022) focus on the unconditional cyclicality of markups. However, this tends to complicate comparisons with model predictions, which are instead conditional on the specific shock driving the cycle. For example, the sticky-price NK model predicts countercyclical markups to demand shocks, but procyclical markups to productivity shocks. Particularly relevant to our study are Meier and Reinelt (2022), which empirically demonstrates that firm-level markup dispersion increases following monetary policy shocks, and Santos, Costa and Brito (2022), which examines the response of the average firm-level markup to demand and productivity shocks.

Our paper bridges these two strands of research by providing evidence of *heterogeneous* firm-level markup cyclicality in response to identified MP shocks, a key demand shock in the literature. We demonstrate that properly aggregating these firm-level responses reveals a countercyclical aggregate markup. Furthermore, our micro-to-macro approach, which highlights the importance of properly aggregating heterogeneous firm-level markups, helps reconcile part of the seemingly conflicting findings in the literature regarding the cyclicality of the aggregate markup, by attributing them to shifts in the firm distribution caused by the aging of firms.

This paper also contributes to the literature on the heterogeneous firm-level effects of monetary policy. Studies examining the impact of monetary policy on firm-level investment responses include Gertler and Gilchrist (1994), Crouzet and Mehrotra (2020), Ottonello and Winberry (2020), Deng and Fang (2022), Cloyne, Ferreira, Froemel and Surico (2023), Anderson and Cesa-Bianchi (2023), González, Nuño, Thaler and Albrizio (2024), Jeenas (2024), and Jungherr, Meier, Reinelt and Schott (2024). Research on the heterogeneous responses of firm stock prices to monetary policy includes Ippolito, Ozdagli and Perez-Orive (2018), Darmouni, Giesecke and Rodnyansky (2022), and Gürkaynak, Karasoy-Can and Lee (2022), while Fabiani, Falasconi and Heineken (2020) focuses on the heterogeneous effects of monetary policy on debt maturity.

We contribute to this literature by providing the first evidence – to the best of our knowledge – of heterogeneity in the firm-level response of markups to MP shocks, which are a key factor in the business cycle dynamics of sticky-price NK models. We also highlight the dominant role of firm age in driving the bulk of the cross-sectional variation in markup responses.

Outline. Section 2 presents the conceptual framework used throughout the paper. Section 3 details our data and empirical measures. Section 4 outlines our empirical strategy and firm-level results. Section 5 discusses the aggregate results. Section 6 concludes the paper.

2 Conceptual Framework

This section outlines the conceptual framework guiding our empirical analysis throughout the paper. First, we review the centrality of the cyclicality of the aggregate markup in standard NK frameworks. Next, we demonstrate how the concept of an aggregate markup cyclicality is indeed connected to the cyclicality of markups at the firm level. Finally, we discuss various theories on the macroeconomic role of the heterogeneous cyclicality of firm-level markups.

Theoretical role of markups in NK models. The markup of price over marginal cost is central to sticky-price New Keynesian (NK) macroeconomic models (Galí, 2015). In these models, current inflation is directly connected to present and future markups through the NK Phillips Curve (NKPC). The NKPC posits that current inflation is negatively related to the current and expected future deviations of markups from their steady-state values, or desired markups. This relationship stems from a straightforward logic: if future markups are anticipated to be below the desired level, firms will raise prices now to achieve the higher desired markup. Because prices are sticky and adjust slowly, firms will act preemptively rather than waiting for the future. This mechanism directly links markups and current inflation.

In the standard NK model, demand shocks – a class of shocks that also include MP shocks – increase (or decrease) both output and firms' marginal costs. However, when prices cannot adjust freely, markups over marginal costs decrease (or increase). Consequently, while a positive demand shock raises output, it also reduces markups below their desired levels. It is the stickiness of prices, together with the NKPC logic described above, that leads to higher inflation after a positive demand shock. Such mechanism generates also the expected co-movement between output and inflation, and Debortoli and Galí (2022, 2024) explain that the same logic is also at the heart of two-agent NK models and heterogeneous-agent NK models.

Micro-to-macro link. While much of the empirical literature has focused on aggregate markups due to data constraints (e.g., Nekarda and Ramey, 2020), most price-setting decision – and therefore markup movements – are likely to occur at the firm level. Here, we present a conceptual framework for linking firm-level markups to aggregate markups empirically.

In a range of models, including those within the NK framework, Grassi (2017), Edmond, Midrigan and Xu (2023), and Baqaee, Farhi and Sangani (2024) illustrate that the relationship between aggregate markups and firm-level markups can be captured by the following:

$$\mathcal{M} = \sum_i \omega_i \mu_i, \tag{1}$$

where $\mathcal{M}$ denotes the aggregate markup, ω_i represents the firm-level variable-cost weight, and μ_i is the firm-level markup. From Equation (1), the impulse response function (IRF) of the aggregate markup in response to a demand shock at horizon h can be expressed as follows:

$$\frac{\partial \Delta_h \log \mathcal{M}_h}{\partial \varepsilon^m} = \underbrace{\sum_i \frac{\omega_i \mu_i}{\mathcal{M}} \frac{\partial \Delta_h \log \mu_{i,h}}{\partial \varepsilon^m}}_{\text{Firm-level effect}} + \underbrace{\sum_i \frac{\mu_i}{\mathcal{M}} \frac{\partial \Delta_h \omega_{i,h}}{\partial \varepsilon^m}}_{\text{Reallocation effect}}, \tag{2}$$

with $\sum_i \frac{\partial \Delta_h \omega_{i,h}}{\partial \varepsilon^m} = 0$ and where ε^m is the demand shock, $\frac{\partial \Delta_h \log \mu_{i,h}}{\partial \varepsilon^m}$ is the firm-level IRF of markups to the demand shock, and $\frac{\partial \Delta_h \omega_{i,h}}{\partial \varepsilon^m}$ is the firm-level IRF of the variable-cost weights to the demand shock. Equation (2) consists of two components. The first is a *firm-level* component, which captures the effect of demand shocks on firm-level markups. The second is a *reallocation* component, which summarizes the impact of demand shocks on the redistribution of costs across firms, and empirically corresponds to the theoretical argument made by Baqaee, Farhi and Sangani (2024). Moreover, the presence of firm-level weights in Equation (2) highlights the importance of a correct, theory-consistent aggregation of firm-level markup dynamics for determining the aggregate markup response to shocks, as argued by Grassi (2017), Edmond, Midrigan and Xu (2023), and Baqaee, Farhi and Sangani (2024).

Indeed, Equation (2) guides our subsequent empirical analysis. First, it underscores the relevance of firm-level markups' response to demand shocks. Thus, we empirically investigate how firm-level markups react to identified demand shocks and whether there are variations in responses across different groups of firms. Second, conditional on observing differences in responses across firms, our conceptual framework emphasizes the importance of investigating the role of reallocation. This involves studying how costs are redistributed among the identified groups in response to demand shocks. The rest of the paper implements this logic.

Dimensions of firm-level heterogeneity in markup responses to shocks. Theoretical contributions to the literature have identified several candidate factors that could lead to heterogeneity in firm-level markup responses to shocks. We review these factors and categorize them into three main groups: financial frictions, demand frictions, and pricing frictions.

Financial frictions. Recent literature has emphasized the role of financial frictions in shaping firms' pricing decisions over the business cycle (e.g., Gilchrist, Schoenle, Sim and Zakrajšek, 2017; Kim, 2021; Meinen and Soares, 2022), which is why we consider financial frictions as a potential driver of firm-level differences in markup responses to MP shocks. While direct measures of financial frictions are often difficult to implement using available firm data, we resort to commonly accepted proxies. In particular, we exploit firm age (Cloyne, Ferreira, Froemel and Surico, 2023), assets (Gertler and Gilchrist, 1994), leverage (Ottonello and Winberry, 2020), and liquidity (Jeenas, 2024) to explore the potential impact of financial frictions on the differences in the cyclicality of firm-level markups conditional on identified MP shocks.

Demand frictions. Another explanation for heterogeneous firm-level markup responses to shocks has been proposed by a growing body of literature suggesting the role of oligopolistic-like forces (e.g., Gopinath and Itskhoki, 2010; Klenow and Willis, 2016; Mongey, 2021; Wang and Werning, 2022; Baqaee, Farhi and Sangani, 2024; Alvarez, Lippi and Souganidis, 2022) or customer capital considerations (e.g., Ravn, Schmitt-Grohé and Uribe, 2006; Nakamura and

Steinsson, 2011; Argente and Yeh, 2022). Both strands of this literature indicate that a fundamental characteristic explaining differential firm-level markup responses to shocks is relative firm size. To investigate this possible explanation in relation to MP shocks, we proxy relative firm size by (i) their sales shares within narrowly defined industries and by (ii) their markups.[2] Note that both measures closely correlate with relative firm size in this class of models.

Pricing frictions. A recent significant contribution by Meier and Reinelt (2022) demonstrates theoretically that heterogeneous nominal price rigidities can also lead to differential markup responses to shocks. However, to the best of our knowledge, existing measures of nominal rigidities are available only at the sector level. Given our focus on firm-level heterogeneity, and in the absence of firm-level proxies for this factor, our empirical analysis will consistently control for it using sector dummies, i.e., we identify heterogeneity in firm-level markup responses while accounting for existing sectoral differences in nominal rigidities.

3 Firm-Level Data, MP Shock, and Markups Measurement

Here, we outline the data utilized in the empirical analysis, detailing our primary data sources and the series of MP shocks. Following that, we describe our measure of firm-level markups.

3.1 Firm-Level Data and MP Shock

Firm-level data. We exploit firm-level data from the quarterly version of Compustat, a dataset that provides balance sheet information for North American publicly listed companies. Our focus is on the period from 1991q1 to 2016q4, aligning with the availability of empirically identified MP shocks series discussed below. We next review the strengths and limitations of this dataset. Additional details on the sample construction are found in Appendix A.

The advantage of Compustat is twofold: first, it provides high-frequency firm-level data, i.e., quarterly data, which is crucial to study questions related to the business cycle. Second, it contains extensive information on firms' age and financial statements, including measures of sales, input expenditures, capital stock, and liabilities, as well as a detailed industry activity

[2]It is worth noting that firm age could, in principle, also serve as a proxy for demand frictions, although the literature primarily uses it as a proxy for financial frictions. Age may reflect how established a firm is in setting its own markups. We explore this idea in more detail in Section 4.2.3 when discussing our main results.

classification. Finally, although publicly traded firms are relatively few compared to the total number of firms, they are often among the largest in the economy, accounting for approximately 30% of US employment (see Davis, Haltiwanger, Jarmin, Miranda, Foote and Nagypal, 2006) and for the bulk of business cycle fluctuations (e.g., Crouzet and Mehrotra, 2020).[3]

To measure firm-level production, we exploit information on sales (SALEQ), while we use the cost of goods sold (COGSQ) to determine the variable inputs used in production, and gross capital (PPEGTQ) to measure tangible capital. In line with the literature, we use selling, general, and administrative expenses (XSGAQ) as a measure of overhead costs, and total assets (ATQ) for a measure of firm assets. We compute liquidity and leverage using cash and short-term investments (CHEQ) and short- and long-term liabilities (DLCQ and DLTTQ), respectively, with liquidity calculated as $\mathtt{CHEQ}/\mathtt{ATQ}$ and leverage as $\mathtt{DLCQ}/\mathtt{ATQ} + \mathtt{DLTTQ}/\mathtt{ATQ}$. Additionally, we assess a firm's relative size by its sales share within a 4-digit NAICS sector, which represents the finest level of disaggregation available. Summary statistics are provided in Appendix A.

Compustat presents two key limitations for our analysis. First, as noted by Cloyne, Ferreira, Froemel and Surico (2023), it provides information on firm age based on corporate age, which reflects the time since incorporation into the dataset. However, we are primarily interested in the firm's true age, based on its founding year. To address this, we cross-verify our findings using Jay Ritter's database, which contains founding years for many US firms. Since this database covers only a subset of Compustat firms, we use it as a validation measure rather than the baseline age measure. Second, like most firm-level datasets, Compustat does not offer firm-level deflators, restricting our analysis to revenues rather than output quantities. We discuss the implications of this limitation for the measurement of markups in Section 3.2.

MP shock series. In our empirical analysis, we use MP shocks as identified demand shocks, as they are extensively studied and widely used in the empirical monetary literature. Particularly, as a benchmark measure for MP shocks, we use the series provided by Jarociński and Karadi (2020). This series refines the identification of high-frequency MP shocks and has been applied successfully in firm-level studies. Specifically, it is constructed from interest rate surprises, based on the percentage change in FED Funds Futures within 30-minute windows around policy announcements. A key feature of the estimation procedure in Jarociński and Karadi (2020) is its correction for biases related to the FED's provision of economic information

[3]The extensive literature on granularity, pioneered by Gabaix (2011), supports this observation.

to private agents through policy announcements. For robustness, we also test our results using an alternative MP shock series proposed by Gürkaynak, Sack and Swanson (2005).

Additional data sources. After merging Compustat and the MP shocks series, we complement the dataset with quarterly indicators of US economic activity. Specifically, we include the GDP growth rate, the Consumer Price Index (CPI) growth rate, the Excess Bond Premium (EBP), and the 1-Year Treasury rate change, as reported by the Federal Reserve of St. Louis.

3.2 Markups Measurement

Firm-level markups reflect the ability of firms to price above marginal costs and are the main object of our analysis. Measuring markups is often difficult, as neither prices nor marginal costs tend to be directly observable in the data. To overcome this, we follow De Loecker and Warzynski (2012) and De Loecker, Eeckhout and Unger (2020), who propose a method to measure firm-level markups based on the production function approach pioneered by Hall (1988). In a nutshell, this estimation strategy is grounded on firms' cost-minimization behavior and allows the estimation of firm-level markups without specifying an explicit demand system. To illustrate that, consider a firm i employing the production technology given by:

$$Q_{i,t} = Q(\boldsymbol{X}_{i,t}, K_{i,t}, \omega_{i,t}), \tag{3}$$

where $\boldsymbol{X}_{i,t} \equiv \{X^{\nu}_{i,t}\}^{V}_{\nu=1}$ is a vector of variable inputs, $K_{i,t}$ is the predetermined input, and $\omega_{i,t}$ is firm productivity. The cost-minimization problem of this general framework is analogous to the one in the NK model, insifar as firms take prices and demand as given and choose inputs to minimize production costs. In particular, this optimization can be expressed as follows:

$$\min_{\{\boldsymbol{X}_{i,t}, K_{i,t}\}} \boldsymbol{P}'_{i,t}\boldsymbol{X}_{i,t} + R_t K_{i,t} + \lambda_{i,t}(Q_{i,t} - Q(\cdot)), \tag{4}$$

where $\boldsymbol{P}_{i,t} \equiv \{P^{\nu}_{i,t}\}^{V}_{\nu=1}$ is the vector of variable inputs prices, R_t is the price of the predetermined input, and $\lambda_{i,t}$ is the Lagrangian multiplier associated to the cost-minimization. The first order condition (FOC) for a generic variable input $X^{\nu} \in \boldsymbol{X}$ is then given by:

$$\frac{\partial \mathcal{L}(\cdot)}{\partial X^{\nu}_{i,t}} = P^{\nu}_{i,t} - \lambda_{i,t} \frac{\partial Q(\cdot)}{\partial X^{\nu}_{i,t}} = 0. \tag{5}$$

Note that the Lagrangian multiplier $\lambda_{i,t}$ can be interpreted as the marginal cost of producing at a given level of output. Moving one step forward, Equation (5) can be rearranged as:

$$\frac{\partial Q(\cdot)}{\partial X^{\nu}_{i,t}} \frac{X^{\nu}_{i,t}}{Q_{i,t}} = \frac{1}{\lambda_{i,t}} \frac{P^{\nu}_{i,t} X^{\nu}_{i,t}}{Q_{i,t}}. \tag{6}$$

Finally, defining the markup as price over marginal cost, $\mu_{i,t} \equiv \frac{P_{i,t}}{\lambda_{i,t}}$, it is possible to rearrange the FOC expressed in Equation (5) for a generic variable input $X^{\nu} \in \boldsymbol{X}$ so that it yields:

$$\mu_{i,t} = \mathcal{E}^{\nu}_{s,t} \frac{P_{i,t} Q_{i,t}}{P^{\nu}_{i,t} X^{\nu}_{i,t}}, \tag{7}$$

where $\mathcal{E}^{\nu}_{s,t} \equiv \frac{\partial Q(\cdot) X^{\nu}_{i,t}}{\partial X^{\nu}_{i,t} Q_{i,t}}$ is the elasticity of output with respect to the variable input X^{ν}. Hence, to measure firm-level markups, we need sales and variable inputs expenditure, which are available in our data, and the elasticity of output with respect to variable inputs, which requires the estimation of the firm-level production function. To increase comparability with previous estimates, we use as a benchmark the sector-level elasticities estimated yearly by De Loecker, Eeckhout and Unger (2020), and report in Appendix A the summary statistics for our baseline markup measure. However, in Appendix B.1.7, we double-check the results of our main analysis by estimating the input-output elasticity at the quarter and sector level under different production function specifications, including a Translog and a Cobb-Douglas.

While this estimator has already been successfully applied to study the high-frequency properties of markups – for instance in Meier and Reinelt (2022) –, several concerns remain. First, the presence of measurement error can introduce noise into our firm-level markup estimates. Second, as pointed out by Bond, Hashemi, Kaplan and Zoch (2021), when only sales data – rather than quantities – is available, using the revenue elasticity in place of the output elasticity in Equation (7) can compromise the accurate estimation of markups. To address these concerns, our empirical specification includes a comprehensive set of fixed effects, such as firm and sector-time fixed effects. These controls help mitigate both permanent and time-varying measurement errors and absorb much of the empirical variation attributable to the output elasticity relative to the variable input, thereby relying primarily on the variation in sales over variable input expenditure for the identification in our empirical strategy.

Regarding specifically the second concern, we also emphasize that our focus is on the *change* in firm-level markups in response to aggregate shocks. De Ridder, Grassi and Morzenti (2024) demonstrate that markup changes, whether computed using revenue elasticity or output-based elasticity, are highly correlated. This suggests that measurement concerns related to production function estimation are unlikely to significantly affect inferences about changes in markups. Nonetheless, to ensure the robustness of our results against potential measurement errors or limitations of the markup estimator, we also perform our main analysis using an alternative markup measure, the Lerner index,[4] as employed by Baqaee and Farhi (2020).[5]

4 Firm-Level Analysis

This section investigates whether there is micro-level heterogeneity in the response of firms' markups to MP shocks, specifically analyzing how firm-level variables correlate with changes in markups at different horizons following MP shocks. We begin by outlining our empirical strategy, and then follow by presenting and discussing our key firm-level findings.

4.1 Empirical Strategy

Relative response across firms. To estimate the relative markup responses across different groups of firms, we use a panel version of the local projections (LP) method proposed by Jordà (2005). Specifically, we estimate the following set of equations by ordinary least squares (OLS):

$$\Delta_h \log \mu_{i,t+h} = \sum_{x\in\mathcal{X}} \left(\alpha_{x,h} + \beta_{x,h}\Delta Y_{t-1} + \sum_{k=-\kappa}^{h} \gamma_{x,h}^{k} \varepsilon_{t+k}^{m} \right) \times \mathbb{1}_{i\in\mathcal{I}^x} + \sum_{\ell=1}^{L} \boldsymbol{\delta_h'} \boldsymbol{X_{i,t-\ell}} + \varphi_{i,h} + \varphi_{s,t,h} + \boldsymbol{\vartheta_h t} + u_{i,t+h} \tag{8}$$

[4]For instance, the Lerner index allows us not to rely on Compustat data on the cost of good sold (`COGSQ`) for the estimation of firm-level markups, and hence on the assumption that this variable captures frictionless factors.
[5]This second methodology is typically referred to as the "accounting-profit approach" and utilizes information on sales and operating income before depreciation, as discussed in Appendix B.1.8.

with horizons given by $h = 0, 1, \ldots, H$. The dependent variable on the left-hand side is the cumulative change in markups for firm i at horizon h, defined by the following expression:

$$\Delta_h \log \mu_{i,t+h} \equiv \log \mu_{i,t+h} - \log \mu_{i,t-1}. \tag{9}$$

On the right-hand side of Equation (8), our key explanatory regressor is the interaction between the MP shock ε_t^m and $\mathbb{1}_{i \in \mathcal{I}^x}$, which is an indicator equal to 1 if firm i was above the *median* value of a given set of variables in the previous year, and 0 otherwise. Specifically, we consider firm-level characteristics that serve as proxies for leading theories of performance heterogeneity, as discussed in Section 2. These firm-level characteristics are given by variables in $\mathcal{X} = \{$*age*, *assets*, *sales share*, *leverage*, *liquidity*, *markup*$\}$. The coefficients of interest are $\gamma_{x,h}^0$ for $x \in \mathcal{X}$, capturing the relative response (on impact) of firms that are (i) older, (ii) larger – both in absolute terms and relative to their sectors – (iii) more leveraged, or that have higher (iv) liquidity or (v) markups, conditional on variations in the monetary policy rate.

Our specification employs a semi-parametric approach, where we estimate the coefficients $\gamma_{x,h}^k$ using dummies instead of linear interactions between the variables in $\mathcal{X}$ and the MP shock series, as done in Cloyne, Ferreira, Froemel and Surico (2023). Robustness checks indicate that the results remain similar when using a parametric specification with linear interaction terms. We also include past and future MP shocks (i.e., $\kappa = 4$) to control for potential serial correlation in the identified MP shocks and to mitigate the attenuation bias that can arise in LP models applied to panel datasets with a short time dimension (Teulings and Zubanov, 2014).

We incorporate firm fixed effects (FE), $\varphi_{i,h}$, to account for unobserved, time-invariant heterogeneity. Therefore, our identification relies on within-firm variation in $\mathcal{X}$ over time. A potential concern is that certain characteristics, like firm age, only increase over time, unlike leverage, liquidity, size, or markups, which can instead vary over time. This could imply that heterogeneity is identified based on differences in firm responses across different periods within our sample. To address this, we introduce sector-time FEs, $\varphi_{s,t,h}$, which account for variations in firm-level responses over time. These FEs strengthen our identification strategy by controlling for time-varying shocks at the sector and aggregate levels, as well as sectoral differences in the degree of nominal pricing frictions. Additionally, a robustness check shows that removing both set of FEs, thereby fully exploiting permanent cross-firm variation and time variation, yields similar results. We also include linear and quadratic trends, $\boldsymbol{\vartheta_h t}$, to account for the growth in markups documented by De Loecker, Eeckhout and Unger (2020).

Following Ottonello and Winberry (2020), we interact $\mathbb{1}_{i\in\mathcal{I}^x}$ with previous-quarter GDP growth ΔY_{t-1} to account for varying sensitivities of firm markups to the business cycle. We include this control in line with best practices, but robustness checks confirm that its inclusion does not affect our results. Additionally, the vector $\boldsymbol{X_{i,t}}$ comprises: (i) firm-level controls such as sales growth and overhead costs relative to sales, (ii) macro-level controls including GDP and CPI growth, changes in the 1-year Treasury rate, and the EBP,[6] and (iii) their lags (i.e., $L = 4$). The vector also includes fiscal-quarter dummies to address seasonality in accounting practices, as noted by Ottonello and Winberry (2020). Finally, standard errors $u_{i,t}$ are clustered at both the firm and quarter levels to control for any serial correlation in the error term.[7]

Level response across firms. The *relative* response estimated through Equation (8) enables us to identify meaningful heterogeneity in how firms' markups respond to MP shocks. However, it does not provide insights into the specific direction of the cyclicality of markups across firms of varying age, size, leverage, liquidity, and markup categories. This is because Equation (8) is saturated with sector-time FEs, which span out common time-series variation across firms. Then, to estimate the *level* response of different groups of firms to MP shocks, we therefore omit sector-time fixed effects and run instead the following regression:

$$\begin{aligned}\Delta_h \log \mu_{i,t+h} = \alpha_h + \beta_h \Delta Y_{t-1} + \sum_{k=-\kappa}^{h} \gamma_h^k \varepsilon_{t+k}^m \\ + \left(\alpha_{x,h} + \beta_{x,h} \Delta Y_{t-1} + \sum_{k=-\kappa}^{h} \gamma_{x,h}^k \varepsilon_{t+k}^m \right) \times \mathbb{1}_{i\in\mathcal{I}^x} \\ + \sum_{\ell=1}^{L} \boldsymbol{\delta_h' X_{i,t-\ell}} + \varphi_{i,h} + \boldsymbol{\vartheta_h t} + u_{i,t+h}\end{aligned} \tag{10}$$

for horizons $h = 0, 1, \ldots, H$ and $x \in \mathcal{X}$. The dependent variable remains the cumulative

[6] We include these aggregate controls, despite sector-time fixed effects absorbing them, to maintain comparability with the specification in Equation (10). Omitting these controls does not change our results.

[7] Clustering at the firm level accommodates flexible error term dependence across time within each firm, while clustering by time is necessary to account for firm-level shocks correlated within a quarter, beyond the co-movement due to industry-level shocks captured by sector-quarter dummies. Without quarter-level clustering, confidence intervals would be significantly narrower.

change in markups for any firm i at horizon h, as previously defined in Equation (9).

This second specification exploits time-variation to estimate the level effect of MP shocks on markups, as captured by γ_h^0, as well as the relative additional effect of belonging to group $x \in \mathcal{X}$, which is captured by $\gamma_{x,h}^0$. We nonetheless preserve firm FE $\varphi_{i,h}$ in the estimation to account for time-invariant firm heterogeneity. Moreover, note that we measure firm-level variables and MP shocks as already discussed for Equation (8), and allow for the same list of firm-level and macro-level controls, including their lags. Finally, we keep clustering the standard errors $u_{i,t}$ at the firm and quarter level to account for correlation in the error term.

4.2 Results

4.2.1 Main Results

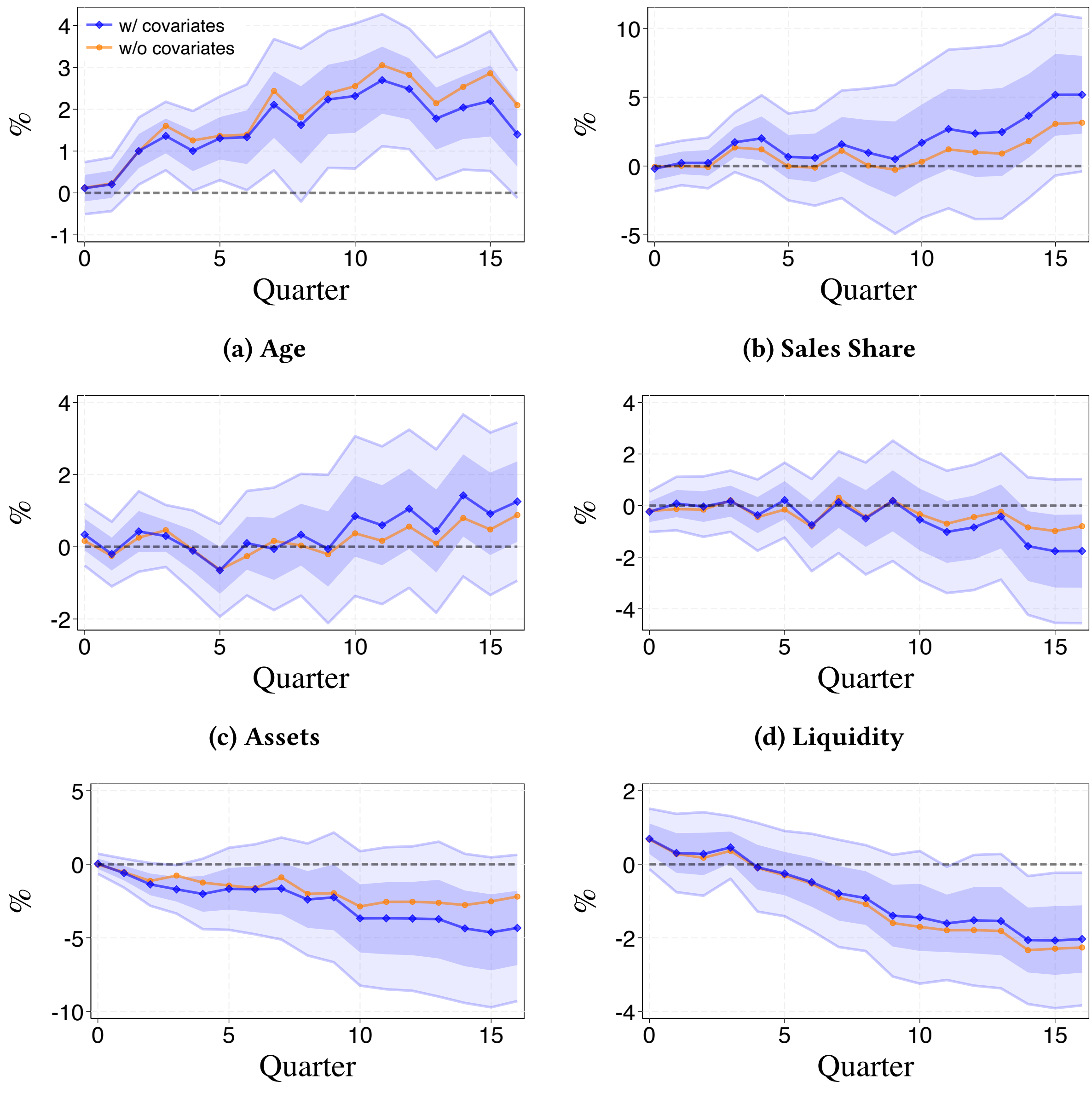
Relative response across firms. To start, Figure 1 shows our first result, namely the IRF obtained from the estimation of coefficients $\gamma_{x,h}^0$ in Equation (8) for $x \in \mathcal{X} = \{$*age*, *assets*, *sales shares*, *leverage*, *liquidity*, *markups*$\}$, along with confidence intervals (CI) around the point estimates. Note that Figure 1 is normalized to a 25 basis points (b.p.) contractionary MP shock, namely to a 25 b.p. increase in the monetary policy rate. Moreover, coefficients $\gamma_{x,h}^0$ for $x \in \mathcal{X}$ are shown for the estimation of Equation (8) with and without covariates – the solid blue and orange lines respectively – and for horizons $h = 1, ..., 16$.[8]

On the one hand, Figure 1 shows that the response of markups to MP shocks is substantially and significantly different when firms are relatively older. The magnitude of the $\widehat{\gamma_{age,h}^0}$ coefficient indicates that being above the median age before a contractionary MP shock of 25 b.p. implies up to a 2.69% statistically significant and positive difference in the subsequent response of markups. Moreover, the sign of $\widehat{\gamma_{sshare,h}^0}$ suggests that being above the median sales share in a 4-digit sector before a contractionary MP shock may imply a positive difference in the response of markups, which is almost statistically significant at the 90% level only in q3.

On the other hand, Figure 1 clarifies that neither the absolute size – measured as total assets – nor the liquidity position of firms in our sample are strong predictors of any heterogeneity in the response of firms' markups to MP shocks. Instead, when firms are highly

[8]Note that the estimation of Equation (8) with covariates allows for all elements of vector $\mathcal{X}$ simultaneously. On the contrary, the estimation of Equation (8) without covariates allows for one element of vector $\mathcal{X}$ at a time.

Figure 1: Heterogeneous Markups Cyclicality By Firm-Level Characteristics



(a) Age **(b) Sales Share**

(c) Assets **(d) Liquidity**

(e) Leverage **(f) Markups**

Note: Figure 1 shows the relative markup response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. Coefficients $\widehat{\gamma^0_{x,h}}$ are shown for the estimation of Equation (8) with and without covariates – the blue and orange lines – and for horizons $h = 1, \ldots, 16$. Note that Figure 1 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$ when estimating Equation (8) with all covariates. Standard errors clustered at the firm and quarter level.

leveraged, their markups show a smaller response to a 25 b.p. MP tightening compared to low leveraged ones, as captured by the estimated coefficient $\widehat{\gamma^0_{lev,h}}$. In particular, being above the median leverage before a contractionary MP shock implies up to a 1.70% negative difference in the subsequent response of markups, but the coefficient is rarely significant at the 90% level.

Finally, the last panel of Figure 1 shows that the conditional markup response of firms with high markups may be bigger than the one of low-markup firms. However, this difference turns negative after six quarters, and it is anyway statistically significant only after q13. As an additional remark, we also stress that the magnitude of the coefficients does not change significantly depending on whether the estimation of Equation (8) simultaneously includes or not the whole set of covariates described in Section 4.1 (represented by the orange line).

In Appendix B.2, we present additional evidence on the relative response of firm-level sales and cost of goods sold — the two main components of our firm-level markup estimator — across the same firm-level characteristics considered so far, and still conditional on a 25 b.p. contractionary MP shock. Overall, we find that firm age is the strongest predictor of heterogeneity in response to MP shocks for both firm-level sales and the cost of goods sold.

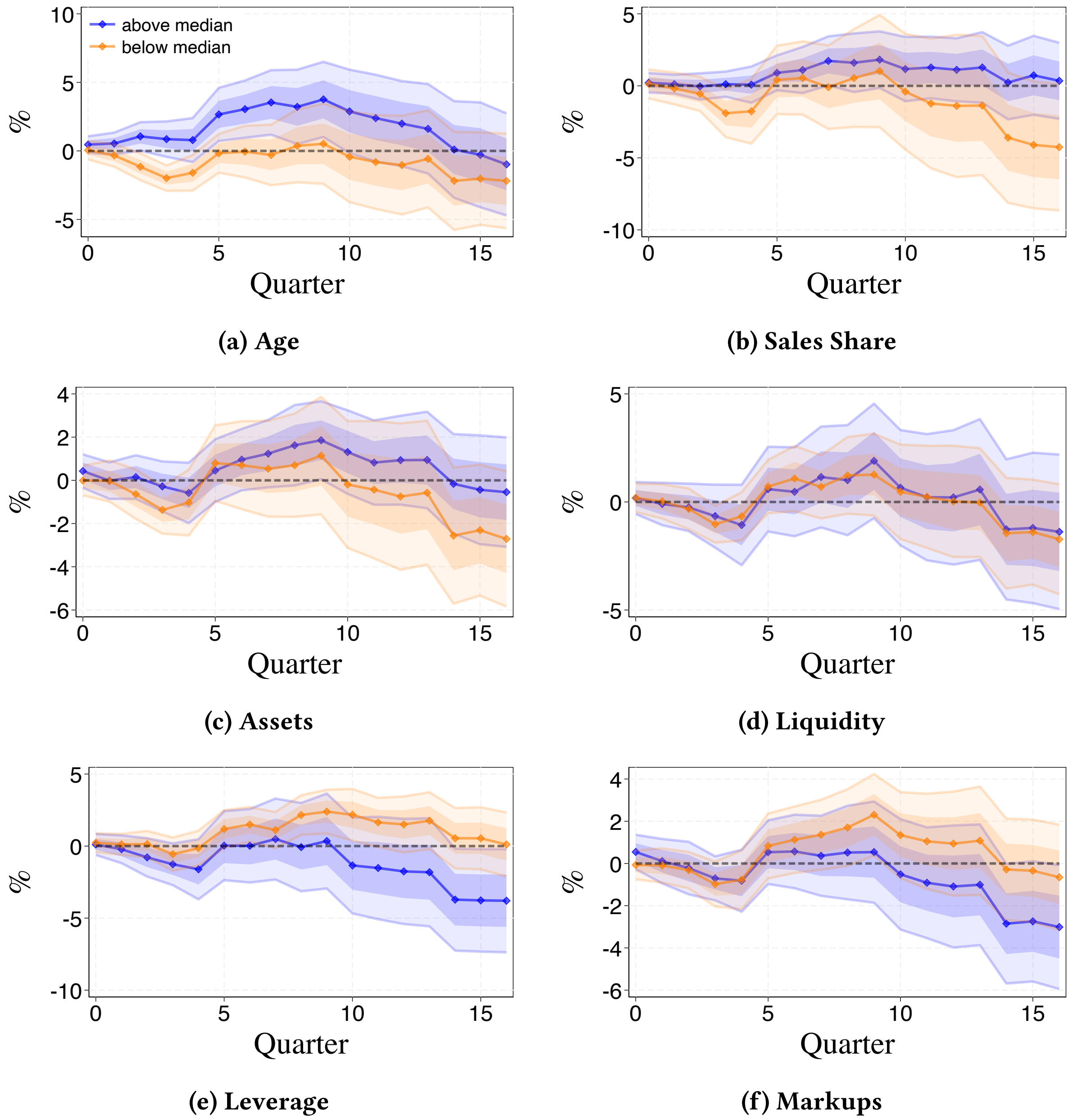
Level response across firms. To go beyond estimating the relative IRF differences across groups of firms and instead establish the specific direction of firms' markups cyclicality, Equation (10) drops the sector-time FE and exploits time series variation to estimate the level response of markups across firms that are above and below the median age, size – both in absolute terms and compared to their sectors –, leverage, liquidity, and markup. Note that the IRF is tracked by γ_h^0 for firms below the median in category $x \in \mathcal{X}$ and by $\gamma_h^0 + \gamma_{x,h}^0$ for those above. Figure 2 above reports the IRFs conditional on a 25 b.p. contractionary MP shock.

First, old firms increase their markups by up to 3.76% after a 25 b.p. contractionary MP shock, while young firms reduce them by 2%. Hence, markup responses vary across the age distribution, with old firms having countercyclical markups and young firms having procyclical ones. Similarly, Panel (b) in Figure 2 seems to suggest that firms with larger sales share show countercyclical markups, while firms with low sales share have (mildly) procyclical ones, although their respective coefficients are less precisely estimated and small in magnitude.

Second, Figure 2 clarifies that firms above and below the median in either the assets (size) or liquidity distribution have a similarly small and rather insignificant procyclical markup response to MP shocks. This further explains why we found no significant difference between these groups of firms when estimating their relative response coefficients in Equation (8).

Third, the last two panels of Figure 2 highlight that firms below the median leverage

Figure 2: Heterogeneous Markups Cyclicality Across Firms



(a) Age

(b) Sales Share

(c) Assets

(d) Liquidity

(e) Leverage

(f) Markups

Note: Figure 2 shows the markup response of (a) old and young, (b) high and low sales share, (c) big and small, (d) high and low liquidity, (e) high and low leverage, (f) high and low markup firms conditional on a MP shock. We plot the response of firms above (blue) and below (orange) the median in each category. Coefficients $\widehat{\gamma^0_{x,h}}$ are shown for the estimation of Equation (8) with covariates and for horizons $h = 1, ..., 16$. Note that Figure 2 is normalized to a 25 basis points contractionary MP shock. The dark and light-shaded areas report the 68% and the 90% confidence intervals. Standard errors clustered at the firm and quarter level.

show a more countercyclical markup response to MP shocks, with markups of firms with low leverage significantly increasing between q5 and q10 after a 25 b.p. monetary tightening. Moreover, we verify the intuition discussed for Figure 1 regarding the differential markup response of firms by their position in the markup distribution. In particular, firms below the median markup before a contractionary MP shock adjust downwards their markups by more in the short run and increase them by more in the medium run. Firms above the median markup have a qualitatively similar response, but the estimated coefficients are not significant.

In conclusion, our empirical findings yield several insights. We uncover substantial heterogeneity in the response of firms' markups to MP shocks when distinguishing between old and young firms. Moreover, when controlling for firm age, other firm characteristics such as sales share, leverage, size, and the level of their markup or liquidity vary in their salience with respect to predicting heterogeneous markup cyclicality conditional on MP shocks.

4.2.2 Robustness Analysis

In Appendix B.1, we provide a detailed discussion of the robustness exercises we conducted to assess the strength of our findings on the heterogeneous response of markups across firms of varying age, size, leverage, liquidity, and markup levels in response to MP shocks.

Here, we provide a brief explanation of each robustness exercise conducted to assess the strength of our results regarding the heterogeneous response of markups across firms with different characteristics. While our focus is on the robustness of Equation (8) as our benchmark, all these robustness checks apply to Equation (10), with results available upon request.

(i) We exclude the Zero Lower Bound (ZLB) period from our analysis, given that identified MP shocks exhibit very little variation during this time, which could potentially affect our findings. (ii) We examine the impact of excluding future MP shocks from the covariates to determine how these forward-looking components influence our results. (iii) We replicate alternative specifications commonly used in the empirical literature by interacting MP shocks linearly with firm-level characteristics, in contrast to our semi-parametric approach using dummies. (iv) We employ an alternative MP shock series, as estimated by Gürkaynak, Sack and Swanson (2005), to evaluate the extent to which our results depend on the chosen shock series. (v) We utilize firms' founding ages from Jay Ritter's database to assess the potential impact of measurement error in firm age, which in our baseline is measured from the date of incorporation. (vi) We redefine firm-level dummies within $\mathcal{I}^x$ by sector and quarter, rather than across the entire sample, to account for sectoral and temporal heterogeneity. (vii) We use alternative production function elasticities that vary at the sector-time and firm-time levels when computing firm markups, as well as (viii) an alternative markup measure, to evaluate the role of potential measurement errors in our baseline measure of firm-level markups. (ix) We assess the influence of control variables on our findings by excluding all controls and (x) also test the results without any fixed effects to fully exploit the variation present in the data.

Overall, our results remain qualitatively and quantitatively similar regardless of the variations introduced through our exercises. This suggests that the dominant role of age in explaining firm-level heterogeneity in response to MP shocks is a robust feature of the data.

4.2.3 Results Interpretation and Discussion

This paper does not aim to carefully microfound the reason why age drives the heterogeneous response of firm-level markups to MP shocks. It rather analyses which firm characteristics can predict a differential markup (conditional) cyclicality, and how changes in the distribution of firms along these characteristics may in turn shape the response of the aggregate markup to MP shocks. Nonetheless, we discuss two plausible interpretations regarding the significant role of age in driving heterogeneous markups responses across firms, which have been explored in the literature reviewed in Section 2. On the one hand, age may reflect varying degrees of financial frictions, with young firms being more prone to financial constraints as opposed to old ones. On the other, age could be linked to customer accumulation dynamics, insofar as old firms, with an established position in their respective market, may face a more inelastic demand and hence different incentives to change markups conditional on MP shocks.

A paper that relates to ours and focuses on the importance of firm age in driving the heterogeneous response of firm investment to MP shocks is Cloyne, Ferreira, Froemel and Surico (2023). They argue that firms are more likely to face financial constraints early in their life cycle when they typically lack stable cash flows and long credit history — a point emphasized in the firm dynamics literature (e.g., Haltiwanger, Jarmin and Miranda, 2013 and Davis and Haltiwanger, 2024). A similar argument is made by Dinlersoz, Kalemli-Ozcan, Hyatt and Penciakova (2018), who show that understanding the relationship between firm characteristics and financial frictions requires considering a firm's age. Additionally, young firms tend to secure a significantly larger share of their borrowing through collateral (Lian and Ma, 2021).

Overall, our result on young firms exhibiting procyclical markups and old firms countercyclical markups in response to contractionary MP shocks could align with theoretical and quantitative studies on the role of financial frictions in shaping differences in firms' marginal cost curves (e.g., Ottonello and Winberry, 2020; Cloyne et al., 2023; González et al., 2024; Jeenas, 2024). Specifically, they are consistent with the view that, after a contractionary MP shock, the marginal cost of financially unconstrained firms declines, leading to countercyclical markups, while constrained firms face rising marginal costs, resulting in procyclical markups, as reflected in our empirical findings. Note that these differences could also be amplified if constrained firms reduce prices more than unconstrained ones, for instance, to generate additional cash flows, as documented by Kim (2021) using detailed US scanner-level price data.

However, a key insight from Figure 1 is that firm age could predict heterogeneity in the response of markups to MP shocks beyond its potential correlation with firms' borrowing constraints, as we control for both leverage, size and liquidity. Indeed, theoretical and empirical works show that firms build their demand as they age, which implies that age may not just be informative of overcoming frictions to the installment of physical capital. On the contrary, firm age could reflect different pricing (and markup) behaviours because of its relation with customer accumulation and the elasticity of the demand faced by firms (Foster, Haltiwanger and Syverson, 2008).[9] This could be a reason why firm age predicts lower pass-through from costs to prices and, in turn, why markups of old firms are less responsive to aggregate shocks.

In conclusion, our evidence suggests that both demand forces and financial frictions may explain the role of age in driving the differential response of firm markups to MP shocks, and more detailed granular data would be necessary in order to carefully disentangle both channels. Here, we stress again that the specific reason behind the strong significance of firm age for the conditional cyclicality of markups does not change the scope of our question, as our goal is rather to highlight the contribution of the documented heterogeneity in firms' markups cyclicality to the cyclicality of the aggregate US markup conditional on MP shocks.

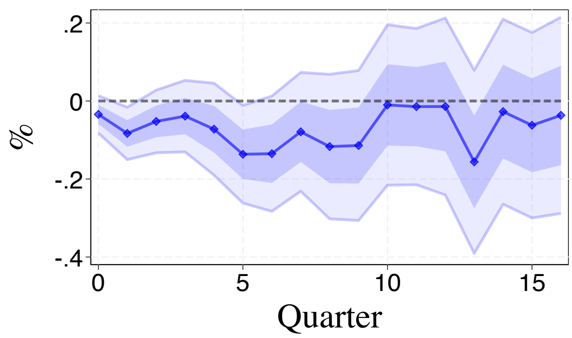
[9] Other studies on similar issues are Roldan-Blanco and Gilbukh (2021), Chiavari (2020), Foster, Haltiwanger and Syverson (2016), Hottman, Redding and Weinstein (2016), Eslava, Haltiwanger and Urdaneta (2024), and Argente, Fitzgerald, Moreira and Priolo (2021). Moreover, old firms tedn to have larger sales shares and are well-established in their (unobservable) market, which suggests they face a more inelastic demand. This observation aligns with the literature emphasizing that relatively large firms – behaving in an oligopolistic fashion – may be less willing to pass cost shocks onto prices (Gopinath and Itskhoki (2010), Klenow and Willis (2016), Mongey (2021), Wang and Werning (2022), Baqaee, Farhi and Sangani (2024), and Alvarez, Lippi and Souganidis (2022)).

4.2.4 Reallocation Across Firm-Age Groups

Having established firm age as the primary predictor of heterogeneous firm-level responses of markups to MP shocks, we go back to the conceptual framework outlined in Section (2), and examine how economic activity reallocates between young and old firms in response to MP shocks. As suggested by Equation (2), a demand shock induces both a firm-level response – which we have estimated in the previous paragraphs – and a reallocation effect, which instead summarizes the impact of demand shocks on the redistribution of costs across firms.

To measure this, we again use the specification in Equation (8), but this time with firm-level variable-cost shares as dependent variable on the left-hand side instead of firm-level (log) markups. Note that, since the sum of the variable cost responses to a MP shock between the two firm-age groups must be zero by construction, identifying the relative cost share response for older firms is equivalent to identifying the level of the cost share response of older firms.

Figure 3: Heterogeneous Cost Shares Cyclicality



Note: Figure 3 shows the variable cost shares response of old firms conditional on a MP shock. Coefficients $\widehat{\gamma^0_{x,h}}$ are shown for the estimation of Equation (8) with covariates and for horizons $h = 1, ..., 16$. Note that Figure 3 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$. Standard errors clustered at the firm and quarter level.

Figure 3 illustrates the response of variable-cost shares for older firms following a contractionary MP shock. We find that the variable cost share of older firms declines by approximately 0.1% after a 25 b.p. hike in the interest rate, also indicating that the variable cost share for younger firms increases by a similar magnitude. This finding suggests a reallocation of variable costs from older to younger firms, mirroring the theoretical result in Baqaee, Farhi and Sangani (2024), which describes a reallocation from large to small firms after a MP tightening. Moreover, Appendix C.2 demonstrates that the estimated reallocation effect looks qualitatively and quantitatively similar when using sales shares instead of variable-cost shares to capture the shift in economic activity conditional on MP shocks. In the next section, we will finally combine the micro-level evidence documented here – regarding the heterogeneity in firm-level markups responses to interest rate movements – to derive the subsequent implications for the cyclicality of the aggregate markup conditional on MP shocks.

5 Aggregate Implications

This section examines the aggregate implications of firm-level markup cyclicality conditional on MP shocks. First, we demonstrate how we aggregate our previously discussed firm-level responses and estimate a countercyclical markup in response to MP shocks across the entire sample. Next, we explore how shifts in the distribution of firms towards older ones could have shaped the cyclicality of the aggregate markup conditional on MP shocks in different periods of our sample, helping reconcile seemingly conflicting findings in the literature.

5.1 Aggregate Markup Cyclicality

Here, we demonstrate how the firm-level empirical analysis conducted in Section 4 allows us to explore the cyclicality of the aggregate markup. Given that one dimension of heterogeneity — e.g. firm age — proved particularly significant in predicting the response of markups to MP shocks, we redefine the aggregate markup from Section 2 without loss of generality as follows:

$$\mathcal{M} = \omega_O \mu_O + (1 - \omega_O)\mu_Y. \tag{11}$$

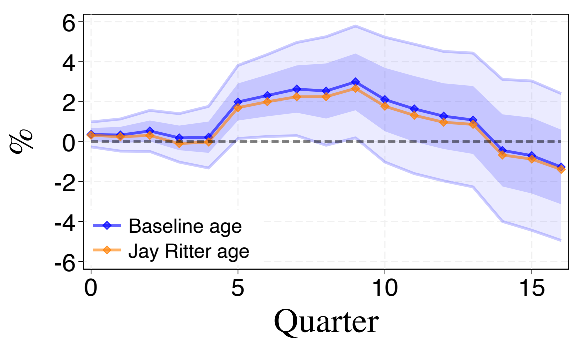
where ω_O is the share of variable costs for old firms (i.e., above the median age), and μ_O and μ_Y denote the variable cost-weighted markups of old and young firms, respectively. Taking a log-linear approximation of Equation (11) conditional on a MP shock allows to connect the firm-level IRFs documented in Section 4 with the IRF of the aggregate markup as follows:

$$\frac{\partial \Delta_h \log \mathcal{M}_h}{\partial \varepsilon^m} = \underbrace{\frac{\omega_O \mu_O}{\mathcal{M}} \frac{\partial \Delta_h \log \mu_{O,h}}{\partial \varepsilon^m} + \frac{(1-\omega_O)\mu_Y}{\mathcal{M}} \frac{\partial \Delta_h \log \mu_{Y,h}}{\partial \varepsilon^m}}_{\text{Direct effect}} + \underbrace{\frac{\mu_O - \mu_Y}{\mathcal{M}} \frac{\partial \Delta_h \omega_{O,h}}{\partial \varepsilon^m}}_{\text{Indirect effect}}, \tag{12}$$

where ω_O, μ_O, μ_Y, and $\mathcal{M}$ are defined as above and directly measurable in our data, while $\frac{\partial \Delta_h \log \mu_{O,h}}{\partial \varepsilon^m}$, $\frac{\partial \Delta_h \log \mu_{Y,h}}{\partial \varepsilon^m}$, and $\frac{\partial \Delta_h \omega_{O,h}}{\partial \varepsilon^m}$ represent the IRFs in response to a MP shock for old firms' markups, young firms' markups, and the old firms' variable cost share, respectively, as estimated in Section 4. Note that Equation (12) effectively represents the two-group equivalent of the more general Equation (2), simply splitting the sample of firms into old and young ones.

Using our baseline measure of age, we find that the share of variable costs for firms above the median age, ω_O, is 0.752; the variable cost-weighted markup for older firms, μ_O, is 1.299; the variable cost-weighted markup for younger firms, μ_Y, is 1.207; and the aggregate markup, $\mathcal{M}$, is 1.276. When using the Jay Ritter measure of age, these values are 0.650, 1.328, 1.244, and 1.299, respectively. By combining these numbers with our IRF estimates from Section 4 and Equation (12), we now compute the IRF of the aggregate markup in response to a MP shock.

Figure 4: Aggregate Markup Cyclicality



Note: Figure 4 presents the aggregate markup response conditional on a MP shock, derived using Equation (12). The results in Figure 4 are normalized to a 25 basis point contractionary MP shock. The dark blue line with squares represents the results obtained using our baseline measure of age, while the orange line with squares reflects the results using the Jay Ritter age measure. The dark and light blue shaded areas indicate the 68% and 90% confidence intervals around the point estimates, respectively.

Figure 4 presents the estimated IRF of the aggregate markup in response to a MP shock, using our baseline age distribution measure (blue line) and the alternative Jay Ritter measure (orange line). The IRF shows a countercyclical pattern, as a MP shock represents a contractionary demand shock, with the response of the aggregate markup peaking around 3% between q5 and q10. Overall, this pattern is consistent with standard NK models featuring price rigidity and aligns with the unconditional empirical findings of Bils, Klenow and Malin (2018).

We also highlight that the IRF estimated in Figure 4 is primarily driven by the direct effect, meaning that the heterogeneous impact across different firm-age groups is what quantitatively matters for the aggregate markup cyclicality in response to MP shocks. The indirect effect, while determining the reallocation of costs across firms (as discussed in Section 4.2.4), remains quantitatively weak. This is due to the fact that the markups of young and old firms, irrespective of the age measure used, are quite similar — i.e., $(\mu_O - \mu_Y)/\mathcal{M}$ is approximately zero. Appendix C.1 presents the IRF decomposition by direct and indirect effects, respectively.

Finally, although the most robust weighting method for the aggregate markup involves variable costs (e.g., Grassi, 2017; Edmond, Midrigan and Xu, 2023), Appendix C.2 demonstrates that using sales weights to compute the aggregate markup, as proposed by De Loecker et al. (2020), yields very similar results. This holds true for (i) estimates of the reallocation of economic activity across firm-age groups conditional on contractionary MP shocks, (ii) the aggregate markup's IRF conditional on the same shocks, and (iii) the estimate of the relative strength of direct and indirect effects for the conditional cyclicality of the aggregate markup.

Next, given the importance of the firm distribution for the response of the aggregate markup, highlighted by the explicit aggregation in Equation (12), we investigate how recent changes in the age distribution of firms may have influenced the aggregate markup's cyclicality in response to MP shocks over time, and its relation to existing findings in the literature.

5.2 Changing Markups Cyclicality: Firm Aging

5.2.1 Micro-to-Macro Approach

Importantly, Equation (12) shows that the conditional cyclicality of the aggregate markup does not only depend on the conditional cyclicality of firm-level markups and the conditional reallocation of variable costs across firms, but also on the relative size of each firm-age group.

On this note, a growing body of literature (e.g., Hathaway and Litan, 2014; Decker, Haltiwanger, Jarmin and Miranda, 2016) has documented a significant decline in the rate of new firm formation and in the exit rate of older firms, leading to a progressive aging of the US firm population. Indeed, Hopenhayn, Neira and Singhania (2022) has shown that this aging trend across US firms has important implications for aggregate factor shares. However, Equation (12) suggests that the impact of firm aging may extend beyond long-term trends and have first-order implications for business cycles, particularly through the cyclicality of markups.

Table 1: Macroeconomic Implications of Firm Aging

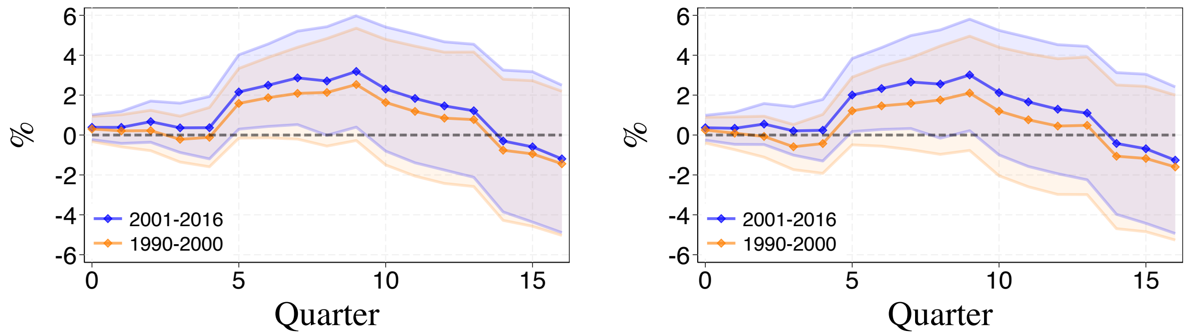
	1990-2000	**2001-2016**
Using *Years Since Incorporation*		
Avg. firm age	6	13
Share of old firms	20%	56%
Cost share by old firms (ω_O)	61%	81%
Markup old firms (μ_O)	1.296	1.299
Markup young firms (μ_Y)	1.222	1.197
Agg. markup ($\mathcal{M}$)	1.267	1.280
Using *Foundation Year* by Jay Ritter		
Avg. firm age	20	27
Share of old firms	26%	55%
Cost share by old firms (ω_O)	47%	76%
Markup old firms (μ_O)	1.326	1.330
Markup young firms (μ_Y)	1.229	1.255
Agg. markup ($\mathcal{M}$)	1.274	1.312

Note: Cost shares are computed using the variable input measure employed in the empirical analysis, i.e., cost of goods sold. The markup for young and old firms represents the average variable cost-weighted markup for firms above and below the median age, respectively. The aggregate markup is calculated as the sum of the markups for old and young firms, each weighted by their respective variable cost shares.

Table 1 confirms the findings of this literature, reporting that the average firm age in our sample has significantly increased over time, from the 1990-2000 period to the 2001-2016 period.[10] Since our baseline measure of firm age may contain measurement errors, as explained in Section 3.1, we also present results based on the exact age measure from Jay Ritter for the available subsample of Compustat firms. As Table 1 shows, the share of old firms has indeed doubled over the past few decades, and the share of production costs accounted for by these older firms has increased from 48-58% to 86-90%. Moreover, Table 1 presents statistics for

[10]We use the year 2000 as the cutoff because it is roughly in the middle of the pre-ZLB period.

Figure 5: Firm Aging and Aggregate Markup Cyclicality – Micro-to-Macro Approach



(a) Compustat Age **(b) Jay Ritter Age**

Note: Figure 5 shows the markup response to a contractionary MP shock before and after 2000 for changes in the distribution of firms age computed (a) using Compustat age, and (b) and Jay Ritter age. Figure 5 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$ when estimating Equation (8) with all covariates. Standard errors clustered at the firm and quarter level.

$(\omega_O, \mu_O, \mu_Y, \mathcal{M})$ across the two sub-periods. Overall, we find that the evolution of markups closely aligns with the findings of De Loecker, Eeckhout and Unger (2020), and we stress that we weight markups by variable costs as outlined in Grassi (2017) and Edmond, Midrigan and Xu (2023), and as required by the NK model, as explained by Baqaee, Farhi and Sangani (2024).

Thus, to compute the implications of firm aging on the conditional cyclicality of the aggregate markup, we follow the procedure outlined in Section 5.1 using the data from Table 1. Figure 5 presents our results, employing either our baseline measure of firm age or Jay Ritter's definition of age (shown in the right and left panels, respectively). Regardless of the age definition used, Figure 5 suggests that the aggregate markup was acyclical or mildly procyclical before 2000q3, though this is not statistically significant at the 90% confidence interval. However, after 2000, the aggregate markup became countercyclical across both firm age definitions, as indicated by the positive and statistically significant estimates from q5 to q10.

Moreover, in Appendix C.3 we show that firm-level markup cyclicality across different age groups has not changed over time, as firms' differential responses before and after 2000 are estimated to be close to zero. This means that the shift in the response of the aggregate markup to MP shocks over time should be driven by changes in the distribution of firms over time, and not by changes in the responsiveness of a particular group of firms over time.

Overall, this finding suggests that markups have become increasingly countercyclical in response to MP shocks due to a shift in the distribution of firms toward old ones, which exhibit more countercyclical markups under such shocks. From the perspective of the NK model, this is a significant aggregate implication, indicating a channel through which amplification forces are intensifying in the US economy. Next, we validate this result by comparing it with results obtained from a direct aggregate approach to assessing the aggregate markup cyclicality.

5.2.2 Macro Validation of the Micro-to-Macro Approach

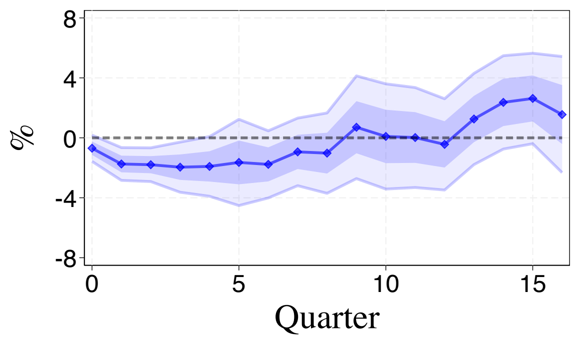
To measure the cyclicality of the aggregate markup using an aggregate economics approach and validate the results of our micro-to-macro methodology, we proceed to estimate the equivalent of Equation (10) at the aggregate level, employing the following aggregate LP model:

$$
\begin{aligned}
\Delta_h \log \mathcal{M}_{t+h} = \alpha_h + \beta_h \Delta Y_{t-1} + \sum_{k=-\kappa}^{h} \gamma_h^k \varepsilon_{t+k}^m \\
+ \left(\alpha_{t>2000q4,h} + \beta_{t>2000q4,h} \Delta Y_{t-1} + \sum_{k=-\kappa}^{h} \gamma_{t>2000q4,h}^k \varepsilon_{t+k}^m \right) \times \mathbb{1}_{t>2000q4} \\
+ \sum_{\ell=1}^{L} \boldsymbol{\delta}_h' \boldsymbol{X}_{t-\ell} + \boldsymbol{\vartheta}_h \boldsymbol{t} + u_{t+h},
\end{aligned}
\tag{13}
$$

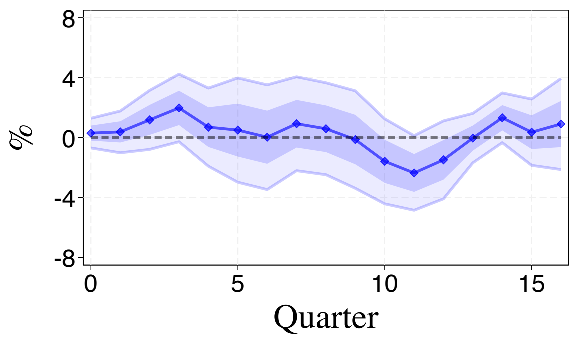
with horizons given by $h = 0, 1, \ldots, H$. The dependent variable measures the cumulative change in the aggregate markup, computed as the cost-weighted average of firms' markups.

Key regressors are the MP shocks and their interaction with $\mathbb{1}_{t>2000q4}$, a dummy variable for periods after the fourth quarter of 2000. As in Section 4, we include past (i.e., $\kappa = 4$) and future shocks to account for any serial correlation in the identified MP shocks and for attenuation biases. We also interact $\mathbb{1}_{t>2000q4}$ with the previous quarter's GDP growth ΔY_{t-1} to control for varying sensitivities of the aggregate markup to the business cycle. Additionally, the vector $\boldsymbol{X}_t$ includes: (i) macro-level controls such as GDP and CPI growth, changes in the 1-year Treasury rate, the EBP, and their lags (i.e. $L = 4$); and (ii) dummies for the Great Financial Crisis and the ZLB. A linear and quadratic trend $\boldsymbol{\vartheta}_h \boldsymbol{t}$ accounts for aggregate markup growth, and standard errors u_t are estimated using the Newey-West estimator with 3 lags, to control for autocorrelation.[11] For robustness, we also interact the MP shocks with a linear trend instead of the post-2000q4 dummy to capture changes in the response of $\mathcal{M}$ over time.

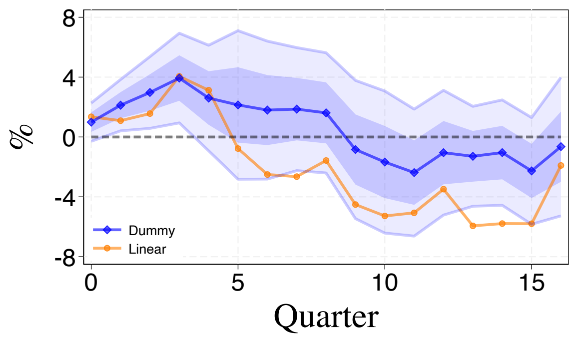
Figure 6: Aggregate Markup Cyclicality Macro Approach



(a) Before 2000



(b) After 2000



(c) Difference

Note: Figure 6 presents the aggregate markup response to contractionary monetary policy shocks before (Figure 6a) and after 2000 (Figure 6b) and the estimated difference (using a dummy or a linear specification) between the two impulse responses (Figure 6c). Impulse responses are normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$. Standard errors are calculated following Newey-West with 3 lags.

Figure 6 reports the aggregate markup response to contractionary MP shocks before (Figure 6a) and after 2000 (Figure 6b), along with the estimated difference between the two IRFs (Figure 6c) using either a dummy variable or a linear specification, as outlined in Equation (13). Findings align with those in Figure 5: the aggregate markup shifted from mildly procyclical before 2000 to mildly countercyclical afterward, and the difference is statistically significant.

To better compare these results with those obtained using the micro-to-macro approach in Section 5.2.1, we compute the cumulative impulse response function (CIR), as defined by Alvarez, Le Bihan and Lippi (2016). The CIR is a widely used statistic that measures the area

[11]We chose 3 lags following the rule of thumb $T^{1/4} = ((2016 - 1990) \times 4)^{1/4} \approx 3$.

under the IRF, summarizing both the impact and persistence of the economy's response. When focusing on the micro-to-macro approach, the CIR for the difference between the aggregate markup responses in the two subperiods, as shown in Panel (a) of Figure 5, is 8.41. A similar result of 11.62 is obtained when computing the CIR in Panel (b) of Figure 5. When instead considering the latter macro approach of this subsection, the CIR for Panel (c) in Figure 6 is 7.19. Overall, the two approaches align reasonably well qualitatively and quantitatively.[12]

Appendix C.4 shows that our results are robust to using alternative measures of the aggregate markup. Specifically, we test their robustness by using the inverse of the labor share, instead of the theoretically consistent aggregation of state-of-the-art firm-level markup estimates. While the inverse labor share is a popular proxy due to its simplicity, Bils, Klenow and Malin (2018) argue that it may not accurately capture aggregate markups, as it could fail to reflect the correct measure of marginal costs. Nonetheless, we observe differences in the IRFs that are qualitatively similar to our baseline in Figure 6c, though less precisely estimated.

5.2.3 Changing Markups Cyclicality in Relation to the Literature

The results we presented in Sections 5.2.1 and 5.2.2 suggest that the choice of the sub-period of analysis is not inconsequential when assessing the conditional cyclicality of the aggregate markup, and may help reconcile some otherwise conflicting evidence found in the literature. This is because the pervasive changes in the age structure of US firms began in the 1980s, suggesting that the increasingly significant role of these firms — with highly countercyclical markups in response to MP shocks — has likely played a larger role in more recent years.

Since the objective of this subsection is not to provide an exhaustive review of the literature but to highlight the importance of sample selection when assessing the cyclicality of firm-level and aggregate markups, we briefly compare our results with two papers that have gained particular attention: Bils, Klenow and Malin (2018), which finds countercyclical aggregate markups, and Nekarda and Ramey (2020), which finds procyclical aggregate markups.

Bils, Klenow and Malin (2018) is a recent paper that argues in favor of countercyclical markups. Although their methodology differs from ours and primarily relies on aggregated

[12] This remains valid, though to a lesser extent, if we use the absolute deviation at the peak as the metric of comparison. In that case, the absolute deviation at the peak for the micro-to-macro approach would be approximately 1 p.p., while for the macro approach, the difference would be around 4 p.p.

data, their findings are well aligned with those we report for the full sample in Section 5.1. In contrast, Nekarda and Ramey (2020), using a different methodology also based on aggregate data, finds that markups are mildly procyclical, or almost acyclical, in response to MP shocks.

Among the many differences between these papers, one stands out in light of our findings in Sections 5.2.1 and 5.2.2. Bils, Klenow and Malin (2018) focus on a sample starting in the late 1980s, whereas Nekarda and Ramey (2020) use a sample dating back to the 1950s.[13] As previously discussed and documented by extensive literature, these distinct periods had very different underlying distributions of firms. In particular, the former period was increasingly dominated by older firms, which we have shown to have highly countercyclical markups, while the latter was dominated by younger firms with highly procyclical markups.

Overall, we believe that differences in sample periods may have played a role in the divergent results of these two papers, potentially aiding in reconciling the existing evidence. Moreover, it underscores the importance of our micro-to-macro approach, which focuses on estimating the primitive response of firm-level markups to MP shocks and explicitly highlights the critical role of aggregation in shaping macroeconomic outcomes.

6 Concluding Remarks

This paper revisits the question on the cyclicality of the aggregate markup using a novel micro-to-macro approach that emphasizes the role of heterogeneous firm-level cyclicality, reallocation across firms, and aggregation. Moreover, rather than focusing on unconditional cyclicality — which is difficult to link to theory due to its dependence on the shock driving the cycle — we concentrate on a major type of shock in the literature: demand shocks.

We exploit quarterly firm-level data from Compustat from 1990q1 to 2016q4 to measure high-frequency firm-level markups using state-of-the-art IO techniques. Then, we estimate firms' heterogeneous impulse response functions to identified monetary shocks, a prominent type of demand shocks in this literature. Our findings reveal that young firms exhibit procyclical markups in response to such shocks, while old firms show countercyclical markups, which may reflect differences their financial constraints. Additionally, we observe a reallocation –

[13]The same is true for Cantore, Ferroni and León-Ledesma (2021) who find countercyclical labor share, i.e., procyclical aggregate markup, to MP in a sample dating back to the 1960s.

albeit small – of economic activity from old to young firms following monetary shocks.

Aggregating these firm-level findings, we find that the aggregate markup is countercyclical with respect to monetary shocks. Furthermore, we stress the importance of the choice of the aggregation method, and show that substantial changes in the conditional cyclicality of the aggregate markup over time may reflect changes in the underlying distribution of firms due to firm aging. Specifically, the aggregate markup shifts from being mildly procyclical or acyclical in the early part of the sample to strongly countercyclical in the latter, when old firms have been making up for a larger share of the overall economic activity. We confirm our results by estimating impulse response functions of the aggregate markup directly, and discuss how this may help to reconcile part of the conflicting views on its cyclicality proposed by the literature.

References

Afrouzi, Hassan and Luigi Caloi, "Endogenous firm competition and the cyclicality of markups," *The Review of Economics and Statistics*, 2022, pp. 1–45.

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

Alvarez, Fernando E, Francesco Lippi, and Takis Souganidis, "Price setting with strategic complementarities as a mean field game," Technical Report, National Bureau of Economic Research 2022.

Alvarez, Fernando, Hervé Le Bihan, and Francesco Lippi, "The real effects of monetary shocks in sticky price models: a sufficient statistic approach," *American Economic Review*, 2016, *106* (10), 2817–2851.

Anderson, G and A Cesa-Bianchi, ""Crossing the Credit Channel: Credit Spreads and Firm Heterogeneity," 2023.

Argente, David and Chen Yeh, "Product life cycle, learning, and nominal shocks," *The Review of Economic Studies*, 2022, *89* (6), 2992–3054.

__ , Doireann Fitzgerald, Sara Moreira, and Anthony Priolo, "How do entrants build market share? The role of demand frictions," Technical Report, Technical report, Mimeo, Penn State University 2021.

Baqaee, David R, Emmanuel Farhi, and Kunal Sangani, "The supply-side effects of monetary policy," *Journal of Political Economy*, 2024, *132* (4), 1065–1112.

Baqaee, David Rezza and Emmanuel Farhi, "Productivity and misallocation in general equilibrium," *The Quarterly Journal of Economics*, 2020, *135* (1), 105–163.

Bils, Mark, "The cyclical behavior of marginal cost and price," *The American Economic Review*, 1987, pp. 838–855.

_ **and James A Kahn**, "What inventory behavior tells us about business cycles," *American Economic Review*, 2000, *90* (3), 458–481.

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

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch, "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data," *Journal of Monetary Economics*, 2021, *121*, 1–14.

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

Cantore, Cristiano, Filippo Ferroni, and Miguel León-Ledesma, "The missing link: monetary policy and the labor share," *Journal of the European Economic Association*, 2021, *19* (3), 1592–1620.

Chiavari, Andrea, "Customer Accumulation, Returns to Scale, and Secular Trends," *Unpublished manuscript, Oxford University*, 2020.

Cloyne, James, Clodomiro Ferreira, Maren Froemel, and Paolo Surico, "Monetary policy, corporate finance, and investment," *Journal of the European Economic Association*, 2023, *21* (6), 2586–2634.

Crouzet, Nicolas and Neil R Mehrotra, "Small and large firms over the business cycle," *American Economic Review*, 2020, *110* (11), 3549–3601.

Darmouni, Olivier, Oliver Giesecke, and Alexander Rodnyansky, "The bond lending channel of monetary policy," *Columbia Business School Research Paper Forthcoming*, 2022.

Davis, Steven J and John Haltiwanger, "Dynamism diminished: The role of housing markets and credit conditions," *American Economic Journal: Macroeconomics*, 2024, *16* (2), 29–61.

_ **,** _ **, Ron Jarmin, Javier Miranda, Christopher Foote, and Eva Nagypal**, "Volatility and dispersion in business growth rates: Publicly traded versus privately held firms [with comments and discussion]," *NBER macroeconomics annual*, 2006, *21*, 107–179.

Debortoli, Davide and Jordi Galí, "Idiosyncratic income risk and aggregate fluctuations," Technical Report, National Bureau of Economic Research 2022.

__ and __ , "Heterogeneity and aggregate fluctuations: insights from TANK models," Technical Report, National Bureau of Economic Research 2024.

Decker, Ryan A, John Haltiwanger, Ron S Jarmin, and Javier Miranda, "Declining business dynamism: What we know and the way forward," *American Economic Review*, 2016, *106* (5), 203–207.

Deng, Minjie and Min Fang, "Debt maturity heterogeneity and investment responses to monetary policy," *European Economic Review*, 2022, *144*, 104095.

Dinlersoz, Emin, 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.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "How costly are markups?," *Journal of Political Economy*, 2023, *131* (7), 1619–1675.

Eslava, Marcela, John Haltiwanger, and Nicolas Urdaneta, "The size and life-cycle growth of plants: The role of productivity, demand, and wedges," *Review of Economic Studies*, 2024, *91* (1), 259–300.

Fabiani, Andrea, Luigi Falasconi, and Janko Heineken, "Monetary Policy and Corporate Debt Maturity," 2020.

Foster, Lucia, John Haltiwanger, and Chad Syverson, "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?," *American Economic Review*, 2008, *98* (1), 394–425.

__ , __ , and __ , "The slow growth of new plants: Learning about demand?," *Economica*, 2016, *83* (329), 91–129.

Gabaix, Xavier, "The granular origins of aggregate fluctuations," *Econometrica*, 2011, *79* (3), 733–772.

Galeotti, Marzio and Fabio Schiantarelli, "The cyclicality of markups in a model with adjustment costs: econometric evidence for US industry," *Oxford Bulletin of Economics and*

Statistics, 1998, *60* (2), 121–142.

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

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

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

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

González, Beatriz, Galo Nuño, Dominik Thaler, and Silvia Albrizio, "Firm heterogeneity, capital misallocation and optimal monetary policy," 2024.

Gopinath, Gita and Oleg Itskhoki, "Frequency of price adjustment and pass-through," *The Quarterly Journal of Economics*, 2010, *125* (2), 675–727.

Grassi, Basile, "IO in IO: Competition and volatility in input-output networks," *Unpublished Manuscript, Bocconi University*, 2017.

Gürkaynak, Refet, Hatice Gökçe Karasoy-Can, and Sang Seok Lee, "Stock market's assessment of monetary policy transmission: The cash flow effect," *The Journal of Finance*, 2022, *77* (4), 2375–2421.

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

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

_ , "What the cyclical response of advertising reveals about markups and other macroeconomic wedges," *NBER working paper*, 2014, *18370.*

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

Hathaway, Ian and Robert E Litan, "Declining business dynamism in the United States: A look at states and metros," *Brookings institution*, 2014, *2.*

Hong, Sungki, "Customer capital, markup cyclicality, and amplification," *FRB St. Louis Work-*

ing Paper, 2017, (2017-33).

Hopenhayn, Hugo, Julian Neira, and Rish Singhania, "From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share," *Econometrica*, 2022, *90* (4), 1879–1914.

Hottman, Colin J, Stephen J Redding, and David E Weinstein, "Quantifying the sources of firm heterogeneity," *The Quarterly Journal of Economics*, 2016, *131* (3), 1291–1364.

Ippolito, Filippo, Ali K Ozdagli, and Ander Perez-Orive, "The transmission of monetary policy through bank lending: The floating rate channel," *Journal of Monetary Economics*, 2018, *95*, 49–71.

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

Jeenas, Priit, "Firm Balance Sheet Liquidity, Monetary Policy Shocks, and Investment Dynamics," 2024.

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

Jungherr, Joachim, Matthias Meier, Timo Reinelt, and Immo Schott, "Corporate debt maturity matters for monetary policy," 2024.

Kim, Ryan, "The effect of the credit crunch on output price dynamics: The corporate inventory and liquidity management channel," *The Quarterly Journal of Economics*, 2021, *136* (1), 563–619.

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

Lian, Chen and Yueran Ma, "Anatomy of corporate borrowing constraints," *The Quarterly Journal of Economics*, 2021, *136* (1), 229–291.

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

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

Meier, Matthias and Timo Reinelt, "Monetary policy, markup dispersion, and aggregate tfp," *Review of Economics and Statistics*, 2022, pp. 1–45.

Meinen, Philipp and Ana Cristina Soares, "Markups and Financial Shocks," *The Economic Journal*, 2022, *132* (647), 2471–2499.

Mongey, Simon, "Market structure and monetary non-neutrality," Technical Report, National Bureau of Economic Research 2021.

Nakamura, Emi and Jón Steinsson, "Price setting in forward-looking customer markets," *Journal of Monetary Economics*, 2011, *58* (3), 220–233.

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

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

Ravn, Morten, Stephanie Schmitt-Grohé, and Martin Uribe, "Deep habits," *The Review of Economic Studies*, 2006, *73* (1), 195–218.

Ridder, Maarten De, Basile Grassi, and Giovanni Morzenti, "The Hitchhiker's Guide to Markup Estimation: Assessing Estimates from Financial Data," 2024.

Roldan-Blanco, Pau and Sonia Gilbukh, "Firm dynamics and pricing under customer capital accumulation," *Journal of Monetary Economics*, 2021, *118*, 99–119.

Rotemberg, Julio J and Michael Woodford, "The cyclical behavior of prices and costs," *Handbook of macroeconomics*, 1999, *1*, 1051–1135.

Santos, Carlos D, Luís F Costa, and Paulo B Brito, "Demand, supply and markup fluctuations," *The Economic Journal*, 2022, *132* (644), 1620–1645.

Teulings, Coen N and Nikolay Zubanov, "Is economic recovery a myth? Robust estimation of impulse responses," *Journal of Applied Econometrics*, 2014, *29* (3), 497–514.

Wang, Olivier and Iván Werning, "Dynamic oligopoly and price stickiness," *American Economic Review*, 2022, *112* (8), 2815–2849.

Heterogeneous Markups Cyclicality and Monetary Policy

Andrea Chiavari, Marta Morazzoni, and Danila Smirnov

Online Appendix

Contents

A Data Appendix

This section provides details on the variables used in the main analysis. First, following standard practices in the literature and ensuring that firms in our Compustat sample face the interest rate set by the FED as their benchmark, we restrict our attention to firms that are incorporated in the US. Second, we exclude firms in utilities (SIC codes between 4900 and 4999) because they have heavily regulated prices. Finally, we exclude financial firms (SIC codes between 6000 and 6999) because their balance sheets are extremely different from those of other firms.

Furthermore, we drop all the observations in our sample with missing industry classification, as well as those observations with negative or missing sales (SALEQ) and cost of goods sold (COGSQ). Whenever applicable, we deflate variables using a GDP deflator from the NIPA tables. Table A.1 reports summary statistics for the variables used in our regression analyses.

Table A.1: Summary Statistics

	Sales	Cogs	Assets	Leverage	Liquidity	Age (Compustat)	Age (Jay Ritter)	Markups
Mean	447.69	303.17	4919.69	0.45	0.16	9.46	23.61	1.78
P25	6.06	3.31	37.83	0.03	0.02	4	9	1.03
P50	31.00	17.18	229.50	0.18	0.07	8	16	1.30
P75	164.58	100.60	1118.33	0.39	0.22	14	29	1.86
N	715,874	715,874	685,784	641,316	683,696	715,874	146,112	715,874

Note: Summary statistics of cleaned quarterly Compustat between 1990q1 and 2016q4. Sales, Cogs, and Assets are measured in millions of real 2012 US$, while Leverage and Liquidity are ratios and Age is in years since IPO (Compustat) and since foundation (Jay Ritter).

In particular, we exploit information on sales (SALEQ) to measure firm-level production, while we use the cost of goods sold (COGSQ) to determine the variable inputs used in production, and gross capital (PPEGTQ) to measure tangible capital. In line with the literature, we use selling, general, and administrative expenses (XSGAQ) as a measure of overhead costs, and the variable ATQ as a measure of total assets. Finally, we use cash and short-term investments (CHEQ) and short and long-term liabilities (DLCQ and DLTTQ) to compute liquidity and leverage (i.e. CHEQ/ATQ and DLCQ/ATQ + DLTTQ/ATQ respectively). As explained in Section 3, our benchmark measure for markups is the ratio between sales and cost of goods sold, multiplied by the input-output elasticities computed by De Loecker, Eeckhout and Unger (2020).

B Firm-Level Analysis Robustness and Additional Results

B.1 Robustness

B.1.1 Excluding the ZLB

In what follows, we estimate Equation (8) between 1990q1 and 2008q4, thereby excluding the Zero Lower Bound (ZLB) period. We do this for several reasons: during the ZLB period, (i) the measure of monetary policy (MP) shocks from Jarociński and Karadi (2020) exhibits nearly no variation, and (ii) central banks favored new unconventional monetary policy practices over standard ones. Both reasons suggest that the ZLB period could be atypical and we want to ensure that this is not driving our main insights. Figure B.1 shows the result. Overall, we notice that excluding the ZLB from the period of analysis does not affect our qualitative conclusions, but increases the magnitude and significance of the estimated coefficient $\widehat{\gamma^0_{age,h}}$.

Figure B.1: Heterogeneous Markups Cyclicality - No Zero Lower Bound Period

(a) Age **(b) Sales Share** **(c) Assets**

(d) Liquidity **(e) Leverage** **(f) Markups**

Note: Figure B.1 shows the relative markup response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. Coefficients $\widehat{\gamma^0_{x,h}}$ are shown for the estimation of Equation (8) with all covariates and for horizons $h = 1, ..., 16$, limiting the sample between 1990q1 and 2008 q4 (before the zero lower bound period). Figure B.1 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$. Standard errors clustered at the firm and quarter level.

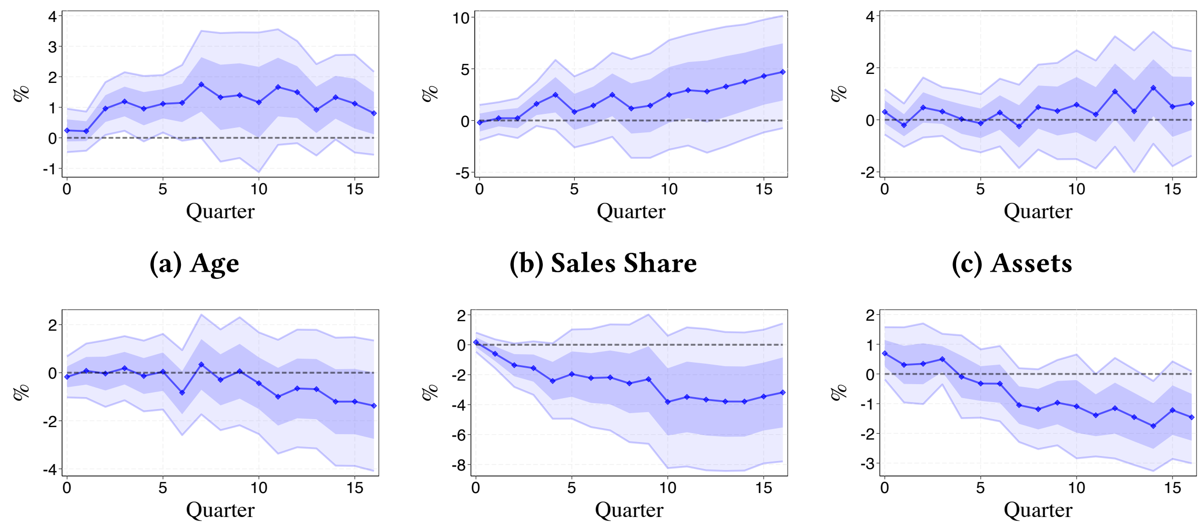
B.1.2 Excluding Future Shocks

In the following robustness analysis, we estimate an alternative specification of Equation (8) in which we do not include future monetary policy shocks as controls, and that is given by:

$$\Delta_h \log \mu_{i,t+h} = \sum\nolimits_{x\in\mathcal{X}} \left(\alpha_{x,h} + \beta_{x,h}\Delta Y_{t-1} + \sum\nolimits_{k=0}^{\kappa} \gamma^{k}_{x,h}\varepsilon^{m}_{t-k} \right) \times \mathbb{1}_{i\in\mathcal{I}^x}$$
$$+ \sum\nolimits_{\ell=1}^{L} \boldsymbol{\delta}'_{\boldsymbol{h}}\boldsymbol{X}_{\boldsymbol{i,t-\ell}} + \varphi_{i,h} + \varphi_{s,t,h} + \boldsymbol{\vartheta_h t} + u_{i,t+h}, \tag{14}$$

with horizons $h = 0, 1, \ldots, H$. Figure B.2 shows the result. Overall, excluding future MP shocks from Equation (8) does not qualitatively affect our results, but it increases the downward bias present in LP models, consistent with the insights in Teulings and Zubanov (2014).

Figure B.2: Heterogeneous Markups Cyclicality - Excluding Future Shocks



(a) Age **(b) Sales Share** **(c) Assets**

(d) Liquidity **(e) Leverage** **(f) Markups**

Note: Figure B.2 shows the relative markup response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. Coefficients $\widehat{\gamma^0_{x,h}}$ are shown for the estimation of Equation (8) with all covariates except for future MP shocks (note that the baseline estimation contains instead 4 leads), and for horizons $h = 1, \ldots, 16$. Figure B.2 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$. Standard errors clustered at the firm and quarter level.

B.1.3 Linear Parametric Interaction

In the following exercise, we estimate an alternative specification of Equation (8) in which we use a linear parametric approach, similarly in spirit to the analysis in Ottonello and Winberry (2020). Instead of characterizing firms with dummies, we linearly interact the MP shock series with measures of firm age, assets, sales share, leverage, liquidity and markup, and estimate:

$$\begin{aligned}\Delta_h \log \mu_{i,t+h} = \sum\nolimits_{x \in \mathcal{X}} \left(\alpha_{x,h} + \beta_{x,h} \Delta Y_{t-1} + \sum\nolimits_{k=-\kappa}^{h} \gamma_{x,h}^{k} \varepsilon_{t+k}^{m} \right) \times x_{t-1} \\ + \sum\nolimits_{\ell=1}^{L} \boldsymbol{\delta}_{\boldsymbol{h}}' \boldsymbol{X}_{\boldsymbol{i,t-\ell}} + \varphi_{i,h} + \varphi_{s,t,h} + \boldsymbol{\vartheta}_{\boldsymbol{h}} \boldsymbol{t} + u_{i,t+h},\end{aligned} \tag{15}$$

for horizons $h = 0, 1, \ldots, H$. Figure B.3 shows our results. Overall, we notice that the coefficients estimated through Equation (15) still reflect the main insights of our benchmark specification. Specifically, Figure B.3 confirms that older firms show more countercyclical markups in response to MP shocks, while there is no evident heterogeneity in the response of markups by firms' size, sales share, leverage, liquidity, or the level of their markup.

B.1.4 Monetary Policy Shocks from Gürkaynak, Sack and Swanson (2005)

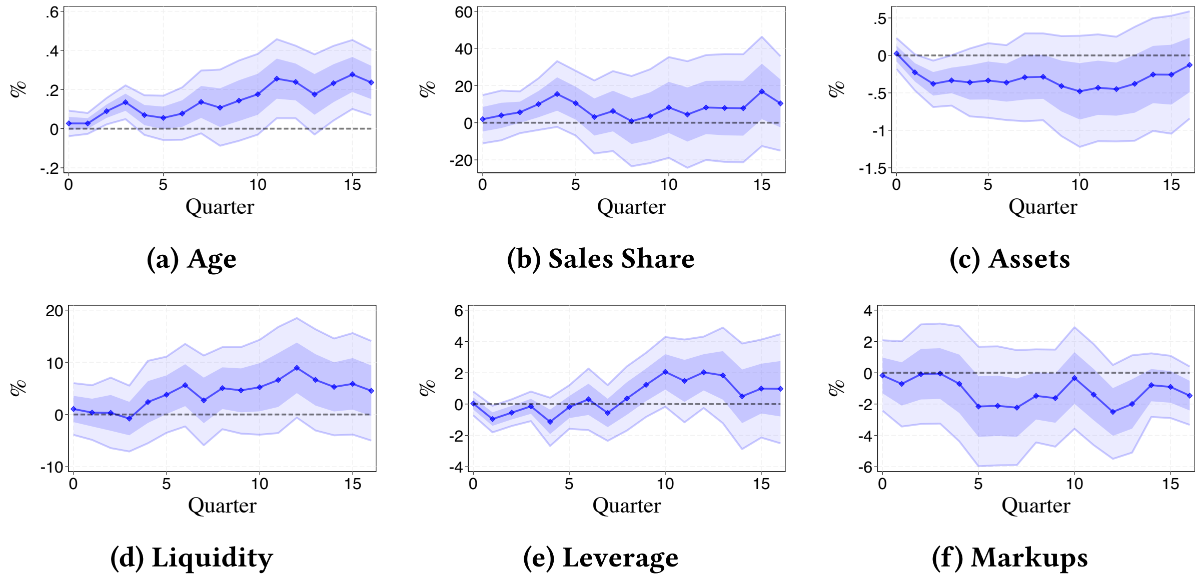
Here, we estimate Equation (8) using an alternative monetary policy shocks measure from Gürkaynak, Sack and Swanson (2005). Their series is similar compared to the one in Jarociński and Karadi (2020) but it does not take into account the information channel of monetary policy.

Figure B.4 shows the results. Overall, we see that using the alternative monetary policy shocks proposed by Gürkaynak, Sack and Swanson (2005) produces a set of IRF that is both qualitatively and quantitatively close to the one in our benchmark.

B.1.5 Using Founding Age by Jay Ritter

In the robustness analysis that follows, we estimate Equation (8) using the "true" founding age of firms instead of corporate age, which is made available by Jay Ritter for a subsample of Compustat (about 20% of the observations approximately, hence a significant empirical limitation if we were to use only founding age in our baseline estimation). Results are shown in Figure B.5. Overall, we see that using this alternative measure of firm age does not sig-

Figure B.3: Heterogeneous Markups Cyclicality - Linear Parametric Specification



(a) Age
(b) Sales Share
(c) Assets
(d) Liquidity
(e) Leverage
(f) Markups

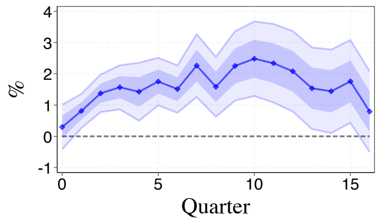
Note: Figure B.3 shows the relative markup response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. Coefficients $\widehat{\gamma^{0}_{x,h}}$ are shown for the estimation of Equation (15) using a linear parametric approach, with all covariates and for horizons $h = 1, ..., 16$. Figure B.3 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^{0}_{x,h}}$. Standard errors clustered at the firm and quarter level.

nificantly alter our main conclusions. In particular, older firms keep their markups relatively higher compared to young ones after the arrival of an MP shock. Neither leverage, liquidity or absolute size (measured in assets) are strong predictors of heterogeneity in markup responses to a monetary tightening, as is also the case for sales share and the markup level.

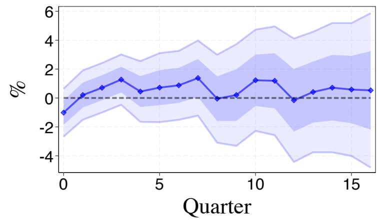
B.1.6 Grouping Firms by Sector and Quarter

In the following exercise, we estimate Equation (8) using a different definition for $\mathbb{1}_{i\in\mathcal{I}^x}$. In the main analysis, we define firms' categories by being above or below the median of $x \in \mathcal{X} = \{\textit{age, sales share, leverage, liquidity, assets}\}$, considering the entire sample and according to firms' characteristics in the previous year. Here, we instead define firms' categories by being above or below the median of those same variables but within a given sector and a given quarter of the previous year. Figure B.6 shows the result. Overall, using a different definition of $\mathbb{1}_{i\in\mathcal{I}^x}$ does not significantly change our conclusions, suggesting that our results do not depend on the particular definition of the dummies capturing firm-level heterogeneity.

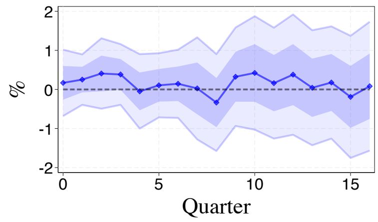
Figure B.4: Heterogeneous Markups Cyclicality - Gürkaynak, Sack & Swanson (2005)



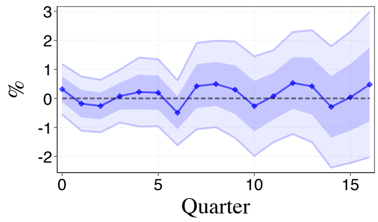
(a) Age



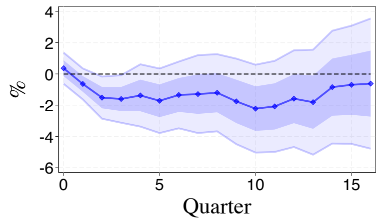
(b) Sales Share



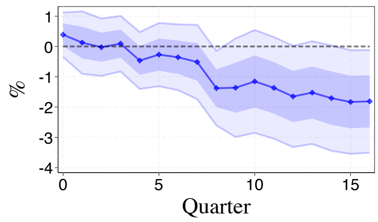
(c) Assets



(d) Liquidity



(e) Leverage



(f) Markups

Note: Figure B.4 shows the relative markup response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. We use the MP shocks series exogenously identified in Gürkaynak, Sack and Swanson (2005). Coefficients $\widehat{\gamma^0_{x,h}}$ are shown for the estimation of Equation (8) with covariates and for horizons $h = 1, ..., 16$. Figure B.4 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$. Standard errors clustered at the firm and quarter level.

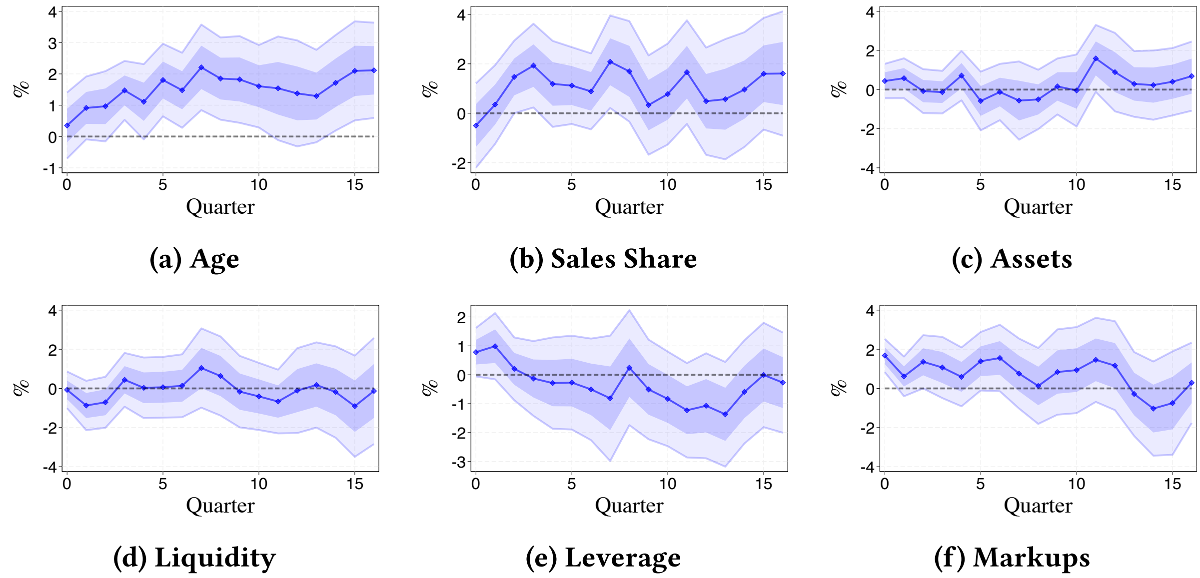
B.1.7 Using Alternative Elasticities

In the following robustness exercise, we estimate Equation (8) using firm-level markups calculated with different production function elasticities. In particular, instead of using the elasticities provided by De Loecker, Eeckhout and Unger (2020), which are common for all the quarters within a year, we calculate our own input-output elasticities allowing them to vary at a quarter level. In particular, we estimate (i) a Cobb-Douglas production function which varies at quarter and sector levels, and (ii) a Translog production function which also varies at quarter and sector levels. Note that this second specification also allows us to have production function elasticities that are heterogeneous across firms within sectors and quarters.[14]

Results are shown in Figure B.7 and B.8. Both specifications show similar patterns regarding the magnitude and significance of the estimated coefficients, both among themselves and compared to our benchmark. This suggests that our conclusions on the different cyclicality

[14]The elasticity coming from the Translog production function is a function of firm-level capital stock (PPEGT in Compustat). For data quality, we interpolate between adjacent points our estimates whenever this item is missing. We check that this does not affect results, although it improves precision.

Figure B.5: Heterogeneous Markups Cyclicality - Using Founding Age



(a) Age (b) Sales Share (c) Assets

(d) Liquidity (e) Leverage (f) Markups

Note: Figure B.5 shows the relative markup response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. We use the founding age of firms in Compustat as collected in the database by Jay Ritter. Coefficients $\widehat{\gamma^{0}_{x,h}}$ are shown for the estimation of Equation (8) with covariates and for horizons $h = 1, ..., 16$. Figure B.5 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^{0}_{x,h}}$. Standard errors clustered at the firm and quarter level.

of markups by firm-specific characteristics are not driven mainly by heterogeneity in firm-level elasticities but by heterogeneity in the firm-level ratio of sales to variable costs. In fact, different firm-level patterns in production function elasticities do not affect our results.

B.1.8 Using Alternative Markups Measure

In this exercise, we estimate again Equation (8) using firm-level markups calculated with a different methodology compared to that presented in Equation (7). In particular, we follow the "accounting-profit approach" adopted by Baqaee and Farhi (2020), where markups are:

$$\mu = \frac{1}{1-\ell} \quad \text{and} \quad \ell = \frac{\texttt{OIBDPQ} - \texttt{DPQ}}{\texttt{SALEQ}}, \tag{16}$$

and $\texttt{OIBDPQ} - \texttt{DPQ}$ denotes operating income before depreciation, net of depreciation.

Results are shown in Figure B.9. Overall, we see that using this alternative measure of market power, which should be considered second-best compared to our main measure, does

Figure B.6: Heterogeneous Markups Cyclicality - Sector/Quarter Dummies

(a) Age | (b) Sales Share | (c) Assets

(d) Liquidity | (e) Leverage | (f) Markups

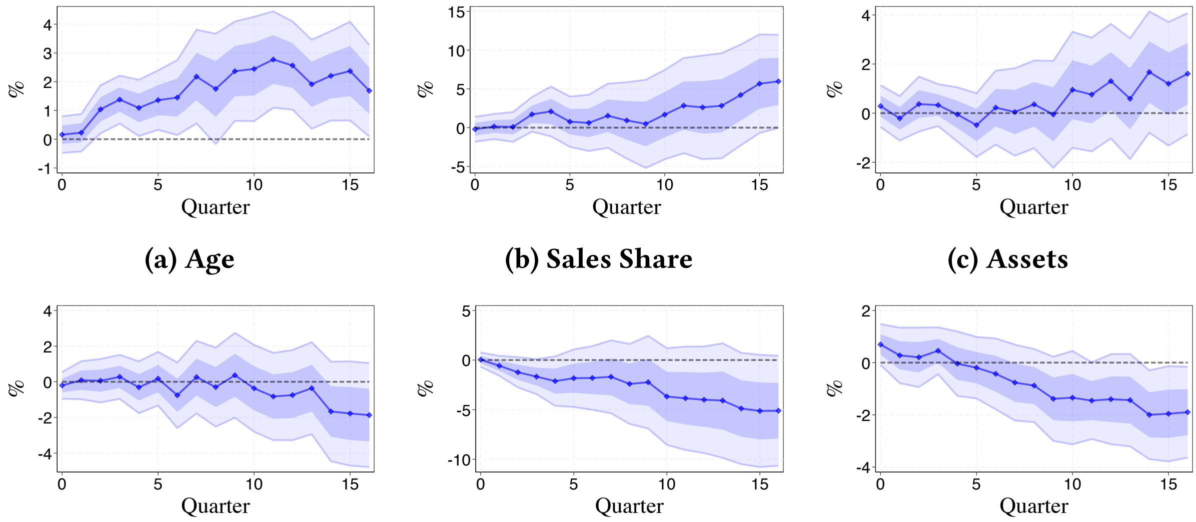
Note: Figure B.6 shows the relative markup response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. We construct dummies equal to one if the variable is above the median within a sector and quarter. Coefficients $\widehat{\gamma_{x,h}^{0}}$ are shown for the estimation of Equation (8) with covariates and for horizons $h = 1, ..., 16$. Figure B.6 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma_{x,h}^{0}}$. Standard errors clustered at the firm and quarter level.

not affect the qualitative implications of our results, although it dampens slightly the magnitude of most coefficients compared to our benchmark estimates from the main analysis.

B.1.9 Excluding All Controls

In this exercise, we estimate again Equation (8) but excluding most of the controls from the analysis, to show that results are not driven by their inclusion. Specifically, we remove the interaction of the MP hock series with past GDP, as well as aggregate and firm-level controls. Recall that macro-level controls include GDP and CPI growth, changes in the 1-year Treasury rate, and the EBP (and 4 lags for each variable). Firm-level controls comprise sales growth and overhead costs relative to sales (and 4 lags for each variable). Results are presented in Figure B.10 and show no significant qualitative change to our preferred specification.

Figure B.7: Heterogeneous Markups Cyclicality - Cobb-Douglas Elasticity



(a) Age **(b) Sales Share** **(c) Assets**

(d) Liquidity **(e) Leverage** **(f) Markups**

Note: Figure B.7 shows the relative markup response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. When measuring firm-level markups, we calculate the input-output elasticity $\theta^{\nu}_{s,t}$ in Equation (7) by estimating a Cobb-Douglas production function that varies at quarter and sector-level. Coefficients $\widehat{\gamma^0_{x,h}}$ are shown for the estimation of Equation (8) with covariates and for horizons $h = 1, ..., 16$. Figure B.7 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$. Standard errors clustered at the firm and quarter level.

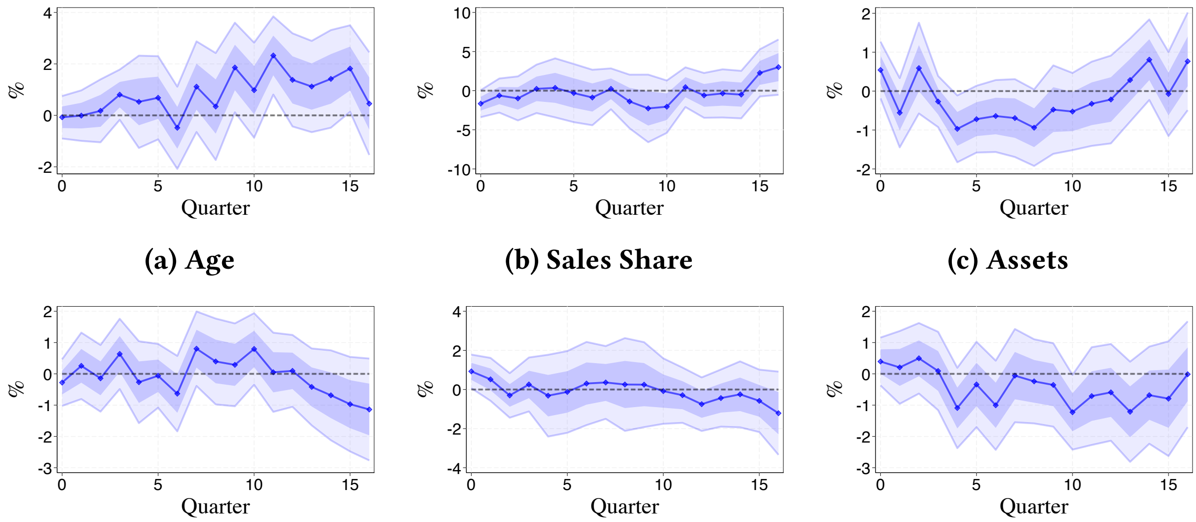
B.1.10 Excluding All Fixed Effects

In this exercise, we estimate again Equation (8) but excluding both firm and sector-time fixed effects, to show that results are not driven by their inclusion. Results are presented in Figure B.11 and show no significant qualitative change to our preferred specification.

B.2 Relative Response of Sales and Cogs

Figure B.12 and Figure B.13 present additional results on the relative response of sales and cost of goods sold — the two main components of our firm-level markup estimator — across various firm characteristics, conditional on a MP shock. Overall, we find that firm age is the strongest predictor of heterogeneity in response to MP shocks for both sales and cost of goods sold. Moreover, while both variables show a negative response, the cost of good sold declines more strongly and significantly for old firms, compared to firm sales, which is consistent with the relative increase of markups for old firms conditional on a contractionary MP shock.

Figure B.8: Heterogeneous Markups Cyclicality - Translog Elasticity



(a) Age **(b) Sales Share** **(c) Assets**

(d) Liquidity **(e) Leverage** **(f) Markups**

Note: Figure B.8 shows the relative markup response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. When measuring firm-level markups, we calculate the input-output elasticity $\theta^{\nu}_{s,t}$ in Equation (7) by estimating a Translog production function that varies at quarter and sector level. Coefficients $\widehat{\gamma^0_{x,h}}$ are shown for the estimation of Equation (8) with covariates and for horizons $h = 1, ..., 16$. Figure B.8 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$. Standard errors clustered at the firm and quarter level.

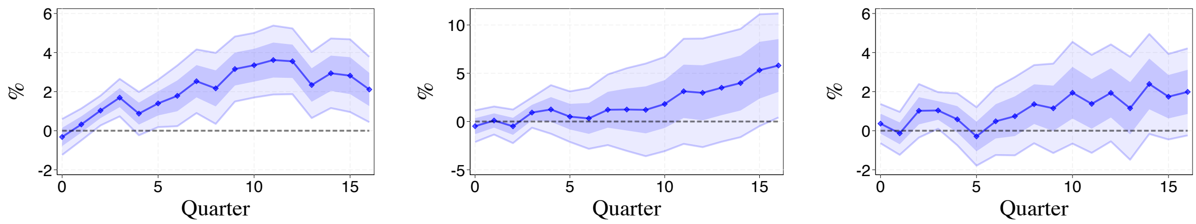
Figure B.9: Heterogeneous Markups Cyclicality - Alternative Markup Measure

(a) Age **(b) Sales Share** **(c) Assets**

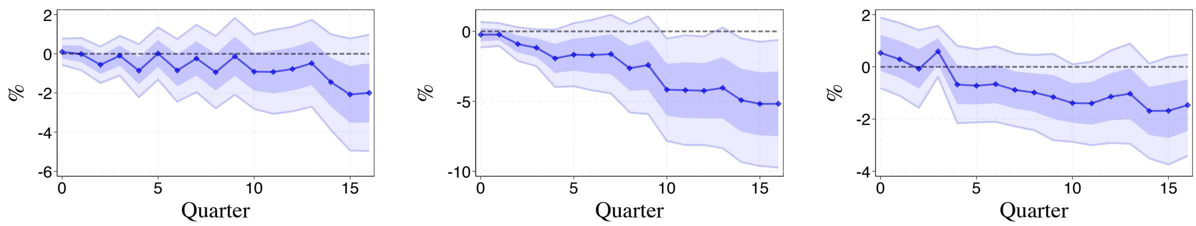
(d) Liquidity **(e) Leverage** **(f) Markups**

Note: Figure B.9 shows the relative markup response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. We use the Lerner index as an alternative measure to firms' markups, as defined in Equation (16). Coefficients $\widehat{\gamma_{x,h}^{0}}$ are shown for the estimation of Equation (8) with covariates and for horizons $h = 1, ..., 16$. Figure B.9 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma_{x,h}^{0}}$. Standard errors clustered at the firm and quarter level.

Figure B.10: Heterogeneous Markups Cyclicality - No Controls



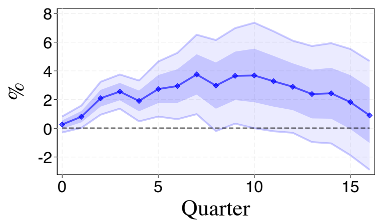
(a) Age **(b) Sales Share** **(c) Assets**



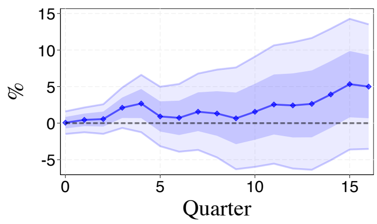
(d) Liquidity **(e) Leverage** **(f) Markups**

Note: Figure B.10 shows the relative markup response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. Coefficients $\widehat{\gamma^0_{x,h}}$ are shown for the estimation of Equation (8) with all covariates and for horizons $h = 1, ..., 16$, but without any control. Figure B.10 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$. Standard errors clustered at the firm and quarter level.

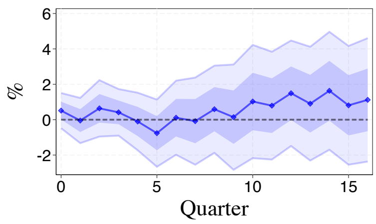
Figure B.11: Heterogeneous Markups Cyclicality - No Fixed Effects



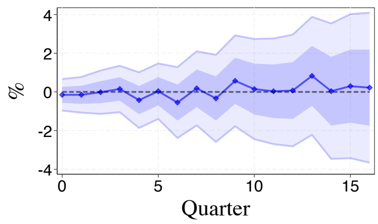
(a) Age



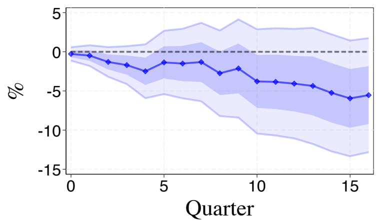
(b) Sales Share



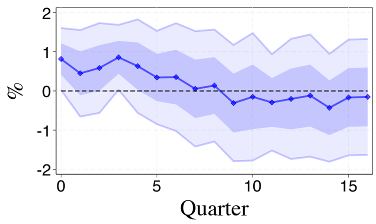
(c) Assets



(d) Liquidity



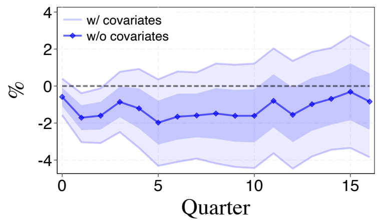
(e) Leverage



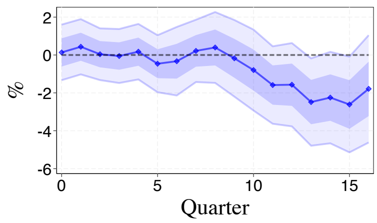
(f) Markups

Note: Figure B.11 shows the relative markup response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. Coefficients $\widehat{\gamma^0_{x,h}}$ are shown for the estimation of Equation (8) with all covariates and for horizons $h = 1, ..., 16$, but without any FE. Figure B.10 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$. Standard errors clustered at the firm and quarter level.

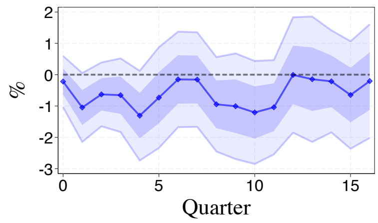
Figure B.12: Heterogeneous Sales Cyclicality By Firm-Level Characteristics



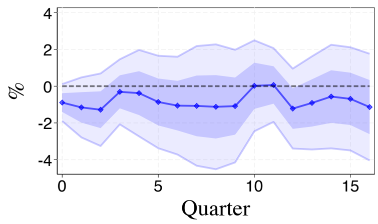
(a) Age



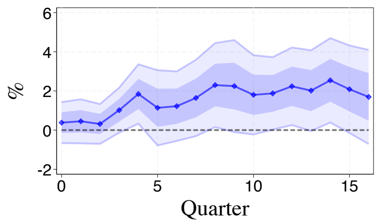
(b) Sales Share



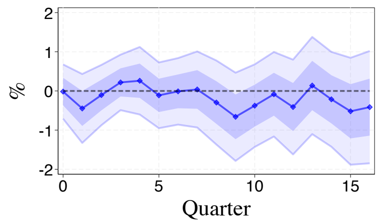
(c) Assets



(d) Liquidity



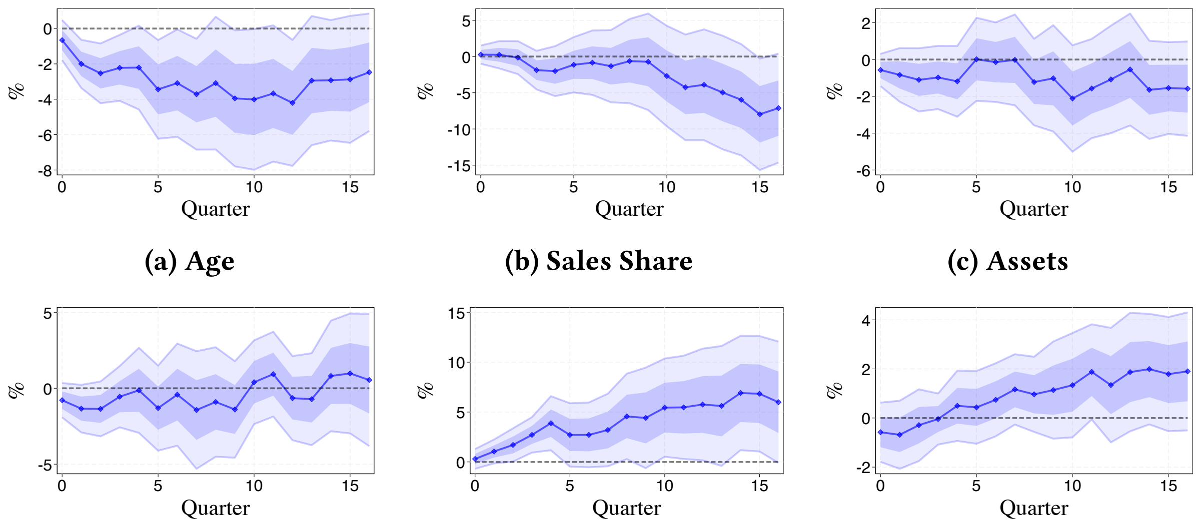
(e) Leverage



(f) Markups

Note: Figure B.12 shows the relative sales response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. Coefficients $\widehat{\gamma_{x,h}^{0}}$ are shown for the estimation of Equation (8) with covariates and for horizons $h = 1, ..., 16$. Note that Figure B.12 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma_{x,h}^{0}}$ when estimating Equation (8) with all covariates. Standard errors clustered at the firm and quarter level.

Figure B.13: Heterogeneous Variable Costs Cyclicality By Firm-Level Characteristics



(a) Age **(b) Sales Share** **(c) Assets**

(d) Liquidity **(e) Leverage** **(f) Markups**

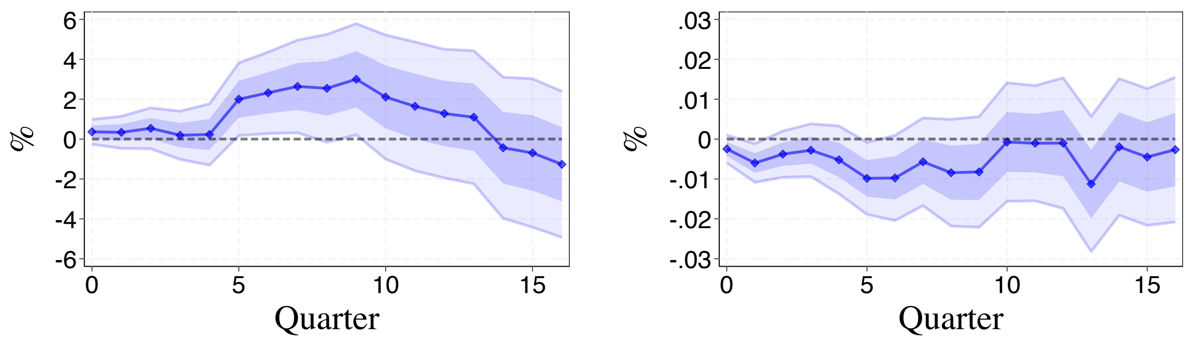
Note: Figure B.13 shows the relative variable costs response of (a) old vs young, (b) high vs low sales share, (c) big vs small, (d) high vs low liquidity, (e) high vs low leverage, (f) high vs low markup firms conditional on a MP shock. Coefficients $\widehat{\gamma^0_{x,h}}$ are shown for the estimation of Equation (8) with covariates and for horizons $h = 1, ..., 16$. Note that Figure B.13 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$ when estimating Equation (8) with all covariates. Standard errors clustered at the firm and quarter level.

C Aggregate Implications: Additional Exercises

C.1 Direct and Indirect Effects

In what follows, we decompose the IRF estimated in Figure 4 to highlight the contribution of both the direct ad indirect effect of firm-level dynamics to the conditional cyclicality of the aggregate markup, as specified in Equation (12). Interestingly, Figure C.14 clarifies that the heterogeneous impact on markups across different firm-age groups is what quantitatively matters for the aggregate markup cyclicality in response to MP shocks. The indirect effect determines the reallocation of costs across firms discussed in Section 4.2.4, but remains quantitatively weak. This is due to the fact that the markups of young and old firms, irrespective of the age measure used, are quite similar — i.e., $(\mu_O - \mu_Y)/\mathcal{M}$ is approximately zero.

Figure C.14: Direct and Indirect Effects



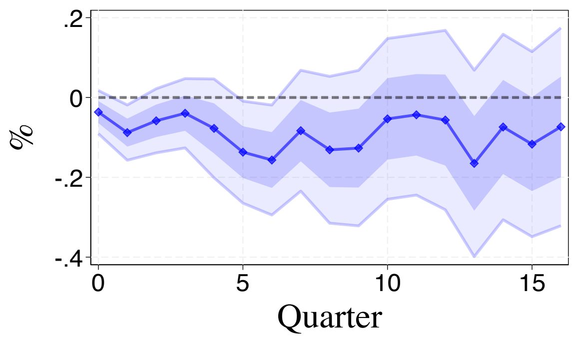
(a) Direct Effect **(b) Indirect Effect**

Note: Figure C.14a and Figure C.14b present the direct and indirect effects as defined in Equation (12). The IRFs are normalized to a 25 basis point contractionary MP shock. The dark blue line with squares represents the results obtained using our baseline measure of age. The dark and light blue shaded areas indicate the 68% and 90% confidence intervals around the point estimates, respectively.

C.2 Sales-Weighted Markup Aggregation

If we compute the aggregate markup as the sales – instead of the variable cost – weighted average of firm-level markups, we can examine how sales reallocate between young and old firms in response to MP shocks, as suggested by the conceptual framework outlined in Section (2). For this, we use Equation (8), but with firm-level sales shares as dependent variable on the left-hand side instead of firm-level (log) markups. Note that since the sum of sales responses to a MP shock between the two firm-age groups must be zero by construction, identifying the relative sales share response for old firms is equivalent to identifying the level of the sales-share response of old firms. Figure C.15 illustrates the response of sales shares for older firms following a contractionary MP shock of 25 b.p. and show a similar pattern to the one estimated for variable cost-shares in Figure 3. Specifically, there is reallocation of sales (and variable costs) from older to younger rms, mirroring the theoretical result in Baqaee, Farhi and Sangani (2024), which describes a reallocation from large to small firms after a MP tightening.

Figure C.15: Heterogeneous Sale Shares Cyclicality

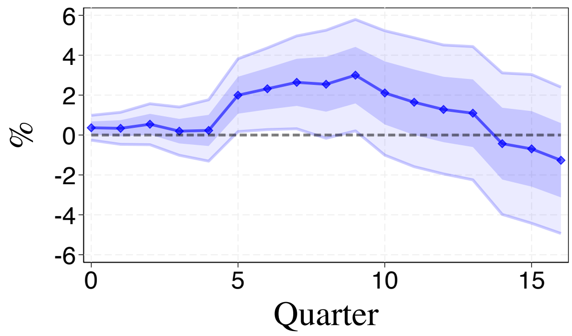


Note: Figure C.15 shows the variable sale shares response of old firms conditional on a MP shock. Coefficients $\widehat{\gamma^0_{x,h}}$ are shown for the estimation of Equation (8) with covariates and for horizons $h = 1, ..., 16$. Note that Figure 3 is normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around $\widehat{\gamma^0_{x,h}}$. Standard errors clustered at the firm and quarter level.

Using our baseline measure of age, we find that the share of variable sales for firms above the median age, ω_O, is 0.761; the variable sales-weighted markups for older firms, μ_O, is 1.624; the variable sales-weighted markups for younger firms, μ_Y, is 1.573; and the aggregate markup using sales instead of costs-weights, $\mathcal{M}$, is 1.612. Notice that here the aggregate markups is higher than the one calculated with cost shares, as explained in De Loecker et al. (2020).

By combining these numbers with our IRF estimates from Section 4 and Equation (12), we now compute the IRF of the aggregate markup in response to a MP shock, as well as its decomposition by direct and indirect effect. Figure C.16 shows that – conditional on a 25 b.p. MP tightening – the estimated sales-weighted markup cyclicality qualitatively and quantitatively reflect the cost-weighted markup cyclicality presented in our baseline specification (Figure 4). Similarly, Figure C.17 confirms that the "direct effect", summarized by the heterogeneous markup responses across different firm-age groups following a MP shock, is

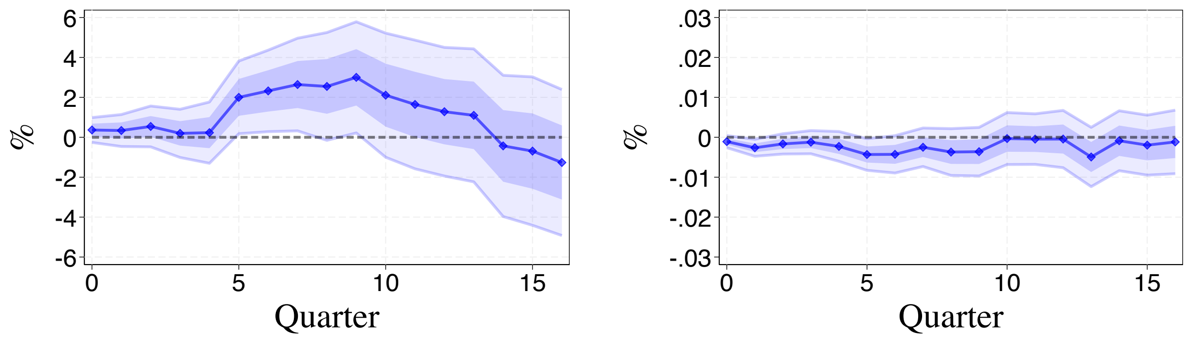
Figure C.16: Aggregate Sales-Weighted Markup Cyclicality



Note: Figure C.16 presents the aggregate markup response conditional on a MP shock, derived using Equation (12). The results in Figure C.16 are normalized to a 25 basis point contractionary MP shock. The dark blue line with squares represents the results obtained using our baseline measure of age. The dark and light blue shaded areas indicate the 68% and 90% confidence intervals around the point estimates, respectively.

what drives the conditional countercyclicality of the aggregate markup. This is regardless of whether the aggregate markup is weighted using sales shares or cost shares, as in Figure C.14.

Figure C.17: Direct and Indirect Effects of Sales-Weighted Aggregate Markup



(a) Direct Effect **(b) Indirect Effect**

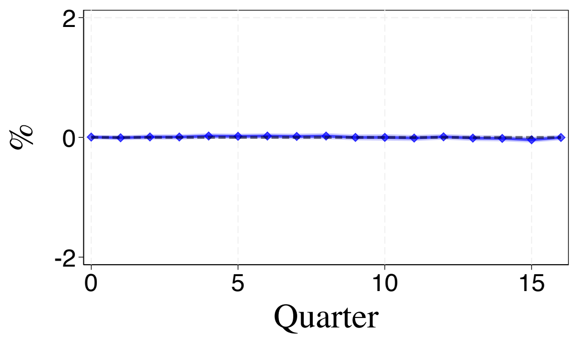
Note: Figure C.17a and Figure C.17b present the direct and indirect effects as defined in Equation (12). The IRFs are normalized to a 25 basis point contractionary MP shock. The dark blue line with squares represents the results obtained using our baseline measure of age. The dark and light blue shaded areas indicate the 68% and 90% confidence intervals around the point estimates, respectively.

C.3 Changing Firm-Level Responses Over Time

In what follows, we estimate an alternative version of Equation (8) to show that firm-level markup cyclicality across different age groups has not changed over time, as firms' differential responses before and after 2000 are estimated to be close to zero. Figure C.18 illustrates the results of this estimation and suggests therefore that the shift in the response of the aggregate markup to MP shocks over time should be driven by changes in the distribution of firms over time, and not by changes in the responsiveness of a particular group of firms over time.

$$\begin{aligned}\Delta_h \log \mu_{i,t+h} = & \sum_{x\in\mathcal{X}} \left(\alpha_{x,h} + \beta_{x,h}\Delta Y_{t-1} + \sum_{k=-\kappa}^{h} \gamma^k_{x,h}\varepsilon^m_{t+k}\right) \times \mathbb{1}_{i\in\mathcal{I}^x} \\ & + \sum_{x\in\mathcal{X}} \left(\alpha_{t\geq 2000q4,x,h} + \beta_{t\geq 2000q4,x,h}\Delta Y_{t-1} + \sum_{k=-\kappa}^{h} \gamma^k_{t\geq 2000q4,x,h}\varepsilon^m_{t+k}\right) \\ & \times \mathbb{1}_{i\in\mathcal{I}^x} \times \mathbb{1}_{t\geq 2000q4} + \sum_{\ell=1}^{L} \boldsymbol{\delta}'_{\boldsymbol{h}}\boldsymbol{X}_{\boldsymbol{i,t-\ell}} + \varphi_{i,h} + \varphi_{s,t,h} + \boldsymbol{\vartheta}_{\boldsymbol{h}}\boldsymbol{t} + u_{i,t+h}\end{aligned} \tag{17}$$

Figure C.18: Changing Firm-Level Responses Over Time

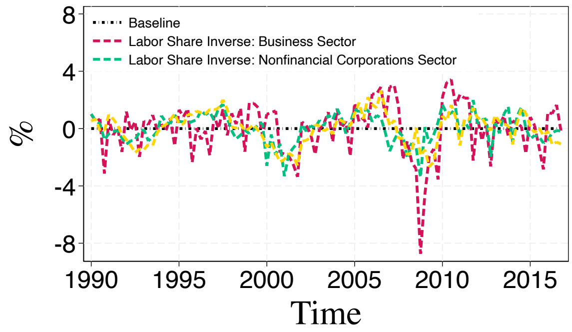


Note: Figure C.18 presents the estimate change before and after 2000 in the differential response of old relative to young firms. The results in Figure C.18 are normalized to a 25 basis point contractionary MP shock. The dark blue line with squares represents the point estimates. The dark and light blue shaded areas indicate the 68% and 90% confidence intervals around the point estimates, respectively.

C.4 Macro Validation of the Micro-to-Macro Approach: Robustness

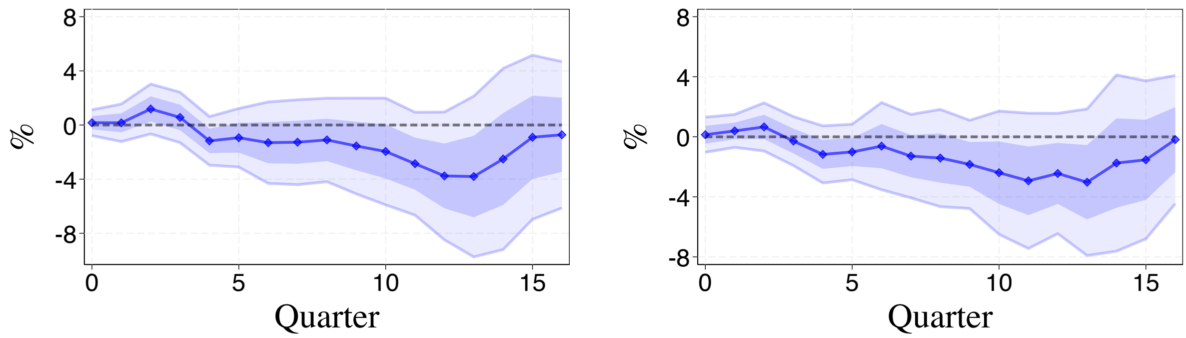
In what follows, we show that our results on the cyclicality of the aggregate markup conditional on MP shocks are robust to using an alternative measures of the aggregate markup. Specifically, we test the robustness of our macro approach by using the inverse of the labor share instead of the theoretically consistent aggregation of state-of-the-art firm-level markup estimates. Note that Figure C.19 compares our baseline aggregate markup measure with the inverse of the labor share across the business and the non-financial corporation sectors.

Figure C.19: HP-Filtered Aggregate Markup Measures



Note: Figure C.19 shows alternative HP filtered measures of the aggregate markup. Each measure is in percentage deviation form its trend.

Figure C.20: Aggregate Markup Cyclicality Using the Inverse of the Labor Share



(a) Nonfinancial Corporations Sector **(b) Business Sector**

Note: Figure C.20 presents the estimated difference (using the nonfinancial corporations sector and business sector labor share) between the two impulse responses before and after 2000. Impulse responses are normalized to a 25 basis points contractionary MP shock. The dark and light blue areas report the 68% and the 90% confidence intervals around the point estimate. Standard errors are calculated following Newey-West with 3 lags.

While the inverse of the labor share is a popular proxy of the aggregate markup due to its simplicity, Bils, Klenow and Malin (2018) argue that it may not accurately capture aggregate markups, as it could fail to reflect the correct measure of marginal costs. Nonetheless, Figure C.20 reports the difference in the aggregate markup response to contractionary MP shocks before and after 2000, as outlined in Equation (13) (using the dummy approach). Findings align with our baseline evidence presented in Figure 6: the aggregate markup shifted from mildly procyclical before 2000 to mildly countercyclical afterward, which is reflected in the estimated difference between the two IRFs, regardless of the specific labor share definition.