Monetary Policy, Intangibles, and the Cyclicality of Markups*

Yumeng Gu†

November 29, 2022

Job Market Paper
Click here for the most recent version

Abstract

I study the role of intangible inputs and firm heterogeneity in determining how price-cost markups respond to interest rate changes. Empirically, I find that firm-level markups are conditionally pro-cyclical, i.e., they go up following an interest rate cut, and that firms that intensively use intangible inputs - such as software and marketing - display more pro-cyclical markups. I use a heterogeneous firm New-Keynesian model with sticky prices to show that through the adoption of intangible inputs, firms exhibit pro-cyclical markups following a monetary easing. Meanwhile, the model preserves conditionally pro-cyclical real wages and inflation. I test key predictions of the model in data and find empirical support for this mechanism. In the cross-section, as in the data, larger firms with higher market share adopt intangible inputs aggressively and become larger and more profitable at the expense of smaller firms following a monetary easing. My findings suggest that a monetary easing could induce inflation driven by both rising costs and rising profits margins.

Keywords: Monetary Policy, Heterogeneous Firms, Intangible Inputs, Market Power

JEL Code: D21; D22; D43; E32; E52

*I own special thanks to my advisor James Cloyne for his continuous guidance and invaluable support. I am grateful to Sanjay Singh and Alan Taylor as well as Nicolas Caramp, Andrès Carvajal and Òscar Jordà for discussion and feedback throughout this project. I thank Gonzalo Basante Pereira, Hanguo Huang, Xian Jiang, Emile Marin, Haopeng Shen, Katheryn Russ and seminar participants at UC Davis for helpful comments and discussion.

†Department of Economics, University of California, Davis (ymgu@ucdavis.edu). All errors are my own.

1 Introduction

What causes inflation when aggregate demand increases? In booms, inflation rises as output increases and unemployment falls. The standard argument in macro models focuses on rising production costs, and that increasing marginal costs drive up inflation. Another strand of literature documents that along with prices, firm profits go up in booms, hinting at the possibility that inflation could be driven by rising profit margins.[1]

In popular macroeconomic models, counter-cyclical profit margins are instrumental for the rise in employment in booms.[2] One prominent measure of firms' profit margins is price-cost markups.[3] In a standard sticky-price New-Keynesian (NK) model, conditionally counter-cyclical markups (i.e., markups fall conditional on a monetary easing) play a key role in the transmission of monetary policy shocks to the real economy through their effect on firm profits.[4] In the presence of sticky prices, a monetary easing increases aggregate demand. Since inflation is driven by the expected path of future real marginal costs, higher demand leads to lower markups and higher labor shares of income, and as the inverse of markups, rising real marginal costs drive up inflation.

Despite the importance of this theoretical transmission mechanism, the related empirical evidence is mixed at best.[5] In fact, the literature finds largely pro-cyclical aggregate markups conditional on monetary shocks (Nekarda and Ramey, 2020). Cantore, Ferroni and León-Ledesma (2021) document complementary evidence that the labor share of income responds counter-cyclically to monetary policy shocks. The disconnect between theory and data casts doubt over its potency as a monetary transmission mechanism and raises the following questions: (1) What causes the pro-cyclical response of markups following a monetary policy shock? (2) What are the implications for the transmission of monetary policy to inflation?

In this paper, I propose that firms' adoption of intangible inputs could explain the conditionally pro-cyclical response of markups, and induce profit-driven inflation following a monetary easing - a type of inflation that is distinct from the usual cost-driven inflation. Specifically, I show that the properties of intangible inputs - production inputs that are not physically embodied, such as software, information technology, advertising, and marketing - allow firms to (1) reduce marginal cost through an endogenous increase in economies of scale, and (2) increase market power through shifting consumer demand. Both a reduction in marginal cost and a rise in market power lead to pro-cyclical markups following a monetary easing. More-

[1] Christiano, Eichenbaum and Evans (2005) offer VAR evidence that profits increase in response to a monetary easing. Hall (2012) find highly pro-cyclical advertising, and note that the finding is hard to be reconciled with counter-cyclical profit margins that constitute a prevalent feature in macroeconomic modelling.

[2] A lowered profit margin implies a higher real wage in booms, and an array of macro models ranging from Keynesian to search-and-marching maps an increase in wages to higher employment.

[3] Markup measures the marginal profit margin of a firm, whereas gross profit measures the average profit margin.

[4] Broer, Harbo Hansen, Krusell and Öberg (2020) note that counter-cyclical markups imply higher output in booms through a negative wealth effect induced by falling profits such that households increase labor supply.

[5] A large empirical literature studies the *unconditional* cyclical variations of markups and related labor market variables, such as the labor share of income (Bils, 1987; Rotemberg and Woodford, 1999; Galí, Gertler and López-Salido, 2007). Recently, Nekarda and Ramey (2020) emphasize studying the *conditional* cyclicality of markups to a known type of business-cycle shocks.

over, standard NK models face an uphill battle to simultaneously deliver pro-cyclical real wages, markups, and inflation following interest rate changes.[6] I show that through firms' adoption of intangible inputs, it is possible to reproduce these properties in a sticky-price NK framework.[7]

To identify what might be driving the pro-cyclical markups, I first look at the cyclical behavior of firm-level markups to investigate the interaction between interest rate changes and intangible inputs. To this end, I empirically examine the cyclical behaviors of markups with firm-level panel data from the US, using local projections instrumental variable techniques (Jordà and Taylor, 2016). I document that firm-level markups are conditionally pro-cyclical to a monetary policy shock, corroborating earlier finding in Nekarda and Ramey (2020) based on industry-level data. In the cross-section, I find new evidence that firms that operate an intangible-intensive technology exhibit more pro-cyclical markups relative to those that rely less on intangible inputs. Intangible intensity measures the share of intangible inputs relative to total inputs. The results are robust to an array of sensitivity analyses, including the use of alternative markup measures and identification strategies for monetary policy shocks, and conditioning on alternative firm characteristics, in particular financial frictions.[8] These empirical findings are puzzling in light of standard NK theory. Moreover, the cross-sectional heterogeneity I document cannot be easily rationalized in a framework that does not take firms' intangible input decisions into account.

Motivated by these empirical findings, I analyze the cyclical behavior of markups in a heterogeneous firm model with multiple *sources* of markup variation, in which market power arises from (1) cost advantage from high productivity, and (2) a firm's ability to influence and shift consumer demand.[9] In the model, firms differ in their levels of productivity that is exogenously given and fixed over time. Meanwhile, through the use of intangible inputs, firms are able to vary both types of market power by (1) endogenously changing their measured productivity through a cost-shifter, and (2) by directly influencing their demand through a demand-shifter. Specifically, I develop a heterogeneous firm New-Keynesian model with sticky prices that has three key features: (i) *strategic complementary* in the form of oligopolistic competition (static Cournot quantity game), (ii) *heterogeneity in firm-level productivity* in a given sector, and (iii) *two types of intangible*

[6] Christiano et al. (2005), among others, document conditionally pro-cyclical real wages to interest rate changes. The challenges for the NK framework are (1) to reproduce higher real wages and higher markups, i.e., how can firms pay more to each worker, while enjoying higher profit margins? And (2) to have higher markups and higher inflation, i.e., why would firms increase price, when they face lessened cost-push pressure?

[7] A standard sticky-price NK model predicts pro-cyclical real wages and inflation, but counter-cyclical price markups following a monetary easing. Incorporating working capital induces conditionally pro-cyclical markups and preserves pro-cyclical real wages, but it also induces a *price puzzle* as declining marginal costs put downward pressure on inflation. Alternatively, if only wage rigidities are present, Galí (2015) shows that a standard NK model would imply a counter-cyclical real wage to monetary shocks.

[8] A common device used in the NK literature to reproduce pro-cyclical price markups is the assumption of working capital, which operates through a credit channel, as firms have to borrow in advance to finance a fraction or all of their labor costs. With working capital, an interest rate cut reduces the cost of borrowing, and if the benefit out-weights the increase in real wage, firms face a lower marginal cost and charge a pro-cyclical markups. I do not find evidence that financially constrained firms are showing more pro-cyclical markups as implied by the working capital assumption.

[9] The approach is motivated by Hottman, Redding and Weinstein (2016), who separate the contribution of cost and "appeal" (quality and taste) and document that firms' idiosyncratic demand is a more important factor in determining firm size than cost.

inputs - a scalable production technology and sales/advertising/marketing - that exhibit complementarity.[10]

In the model, firms compete strategically in a sector both inter-temporally through price competition, and intra-temporally through optimal adoption of intangible inputs. The discrete number of firms in an oligopoly induces variations in markups, but it is through the adoption of intangible inputs that the model generates a pro-cyclical aggregate markup response following a monetary easing.[11] In the cross-section, the model predicts that following a monetary easing, highly productive firms (the "larger" firms with higher market share and charge higher markups in steady state) aggressively adopt both intangible inputs. As such they are able to grow even larger and charge higher markups, and this drives the aggregate response.

The intuition for the mechanism is as follows. The first ingredient is heterogeneity in firms' productivity. At the core of the mechanism is that the asymmetrical marginal benefits of intangible inputs' adoption magnify the initial exogenous productivity differences. Specifically, their adoption creates a mutually reinforcing feedback loop for larger firms in a sector. They are incentivized to increase advertising and marketing efforts because the potential profit from gaining additional demand is proportional to their markup and is thus greater. They then attract more demand and thus produce more, incentivizing further adoption of the scalable production technology to further reduce cost and increase markup. In contrast, smaller businesses are facing the opposite effects and are disincentivized to spend on intangibles. Thus, increased economies of scale, when coupled with firms' increased ability to pass-through cost to consumers, would reinforce each other and put upward pressure on both markups and prices for larger firms.

The second ingredient is an endogenous increase in economies of scale induced by the use of a scalable production technology. As a "cost-shifter", it naturally induces pro-cyclical markups by shifting costs away from a variable cost component and its adoption is observationally equivalent to an increase in measured productivity. Notable examples are software and information and communication technology (ICT), technologies that are *scalable* because they are costly to deploy, yet they incur minimal marginal cost when productions are scaled up. Notably, pro-cyclical markups induced by the scalable technology alone do not necessarily lead to higher prices, as they reflect "lowered marginal cost".

Lastly, firms need market power in equilibrium to implement scalable technologies that come with an upfront cost. Importantly, the cyclical behaviors of markups across firms are mediated by the sources of market power. In particular, firms could directly influence consumer demand such that they increase market power beyond what is implied by their cost advantage. Firms deploy the second major type of intangible

[10] Corrado, Hulten and Sichel (2005) broadly classify business intangibles into three types: computerized information, economic competencies, and innovative property, and correspondingly, software, brand value, and R&D are prime examples for each type. I discuss in the paper why I focus on the first two types and why R&D is unlikely to be driving my results.

[11] In a standard CES-oligopoly model, a monetary easing leads to a reduction in desired markups because of strategic complementarity in pricing, complementing the counter-cyclical markups induced by pricing rigidities. Likewise, if a Kimball (1995) aggregator is used, a monetary easing leads to a reduction in desired markups, unless a negative superelasticity is adopted, in which case, more productive firms would have counterfactual, lower markups.

inputs - sales, advertising, and marketing, as a "demand-shifter" - to shift demand towards their products, such that they effectively face a less elastic demand curve. The two complementary intangible inputs exhibit *synergies* with one another, thus creating the positive feedback loop for the larger firms.[12] Furthermore, pro-cyclical markups reflect both "lowered marginal cost" and "increased market power", such that whether they imply higher prices depends on the relative strength of these two forces.

I test the proposed mechanism by bringing the model's predictions to data. I find evidence that larger firms operating an intangible-intensive technology display more pro-cyclical markups following a monetary easing, consistent with the prediction of the model. I also show that, broadly consistent with the model's mechanism, there exist co-movements between expenditures on intangible inputs and sales at the firm-level following a monetary policy shock. Finally, to rule out possible confounding factors, I use a triple-sorting of the data to condition on alternative firm characteristics including Research and Development (R&D) intensity, financial frictions, and exporter status. I show that market share and intangible intensity remain strong predictors for the cross-sectional heterogeneity, and that these confounding factors are unlikely to be driving my empirical findings.

The mechanism has several implications. A monetary expansion could induce a rise of firm market power that translates into inflationary pressure as pro-cyclical profit margins (as measured by markups) put upward pressure on prices. The notion that markups contribute to inflation is at odds with the conventional argument for cost-driven inflation; i.e., an expansionary monetary policy leads to a strong labor market with shrinking profit margins and rising wages that move inflation along the Phillips curve.[13] My finding points to the opposite. Following a monetary easing, profit margins widen, resulting in profit-driven inflation. In particular, larger firms experience greater profit-driven pressure from an increase in market power that dominates the lowered cost-driven pressure from a reduction in marginal cost, and end up charging a higher relative price. Inflation driven by increasing market power differs from inflation driven by rising labor costs in its implication for the distributional consequence of monetary policy. Conventionally, monetary easing leads to a redistribution of income from firm profits to labor. My results instead imply that a monetary easing could lead to a redistribution of income from workers to firm profits.

Related Literature. The paper is related to a vast empirical literature studying the cyclical behavior of markups and related labor market variables (e.g., labor share of income and real wage). Most of the early contributions focus on the the dynamics of wage markups over the business cycle and generally find markups to be unconditionally counter-cyclical (Bils, 1987; Rotemberg and Woodford, 1999; Galí et al., 2007;

[12] Haskel and Westlake (2018) argue that intangible assets exhibit fundamentally different properties to their tangible counterparts: (1) intangible investment tends to be a *sunk* cost, and (2) generates *spillovers*; (3) intangible assets tend to be *scalable*, and (4) have *synergies* with one another. I focus on the latter two properties that are arguably more relevant to intangible inputs.

[13] Despite the emphasis on costs, leading DSGE models in Smets and Wouters (2007) and Justiniano, Primiceri and Tambalotti (2010) allow for exogenous "cost-push" shocks to the natural markups and consider them important in explaining inflation dynamics.

Bils, Klenow and Malin, 2018; Burstein, Carvalho and Grassi, 2020). More recent studies instead focus on the *conditional* cyclicality of markups.[14] Among others, Anderson, Rebelo and Wong (2018) and Stroebel and Vavra (2019) document that markups are conditionally pro-cyclical to demand shocks. Likewise, Nekarda and Ramey (2020) show that markups are pro-cyclical to monetary policy shocks.[15] Cantore et al. (2021) instead focus on the observable labor shares of income and document that they are conditionally counter-cyclical to monetary policy shocks in a group of advanced economies. My contribution to this empirical literature is that first I provide additional empirical evidence based on firm-level data that corroborates the previous finding of conditionally pro-cyclical markups based on aggregate data, and second, utilizing the rich micro data, I identify the characteristics of firms that are driving the pro-cyclical responses.

My work is also related to an extensive theoretical literature studying models with variable elasticity of demand. The literature is founded upon the robust empirical finding that firms consistently charge heterogeneous markups and pass-through heterogeneous fractions of cost shocks (e.g., exchange rate shocks) into their prices.[16] Various alternative demand systems (relative to the baseline constant-elasticity-of-substitution (CES) demand structure) have been proposed to reproduce this robust feature of the data, most notably Kimball preferences (Kimball, 1995; Klenow and Willis, 2016; Gopinath and Itskhoki, 2010). Alternative approach replaces the baseline market strucrure of monopolistic competition with that of oligopolistic competition in the vein of Dornbusch (1987), Krugman (1986), and Atkeson and Burstein (2008). Amiti, Itskhoki and Konings (2019) combine both non-CES demand and firms' strategic interaction to match evidence of exchange rate shock pass-through across firm sizes.[17]

A closely related strand of literature incorporates nominal rigidities into a variable markup model. Sbordone (2007), Benigno and Faia (2010), and Guilloux-Nefussi (2020) propose a two-country NK model to analyze the implication of globalization on domestic output-inflation trade-offs. More recently, Mongey (2021) and Wang and Werning (2020) study the implications of oligopoly for monetary non-neutrality in a closed economy with dynamic strategic complementarities. In general, these papers imply that real rigidites from strategic interaction complement nominal rigidities in pricing, leading to more counter-cyclical markups and stronger monetary non-neutrality. A separate strand of literature, relying on consumer market frictions, is able to induce pro-cyclical markups. Examples include labor market search (Kaplan and Menzio, 2016), directed search in good markets (Qiu and Ríos-Rull, 2022), and endogenous assortment (Anderson

[14]The approach is arguably more elucidative. Business cycles are likely driven by a variety of shocks; the *unconditional* cyclicality of markups is rather an outcome driven by potentially counteracting forces. Therefore, unconditional counter-cyclicality results do not in principle reject pro-cyclical conditional markups on an identified shocks.

[15]One criticism is that markups are not directly observable and thus the empirical results could be systematically biased contingent on empirical implementation (Bond, Hashemi, Kaplan and Zoch, 2021).

[16]The workhouse models in macro and international economics assume CES demand structure and monopolistic competitive market structure. They collectively imply a constant markup and complete pass-through of cost shocks in equilibrium.

[17]Variable markup channel provides one source of firm-level real rigidities. Alternative sources of real rigidities occur at the aggregate-level (e.g., round-about production structure (Basu, 1995), real wage rigidities (Blanchard and Galí, 2007)). Models with rational inattention and sticky information can also induce incomplete pass-through of cost shocks to price.

et al., 2018). My focus on firms' role in actively influencing consumer demand differs from previous papers that center on consumer-side frictions and in which firms passively facing a less elastic demand curve in booms.[18] My contribution is two-fold: first I propose a novel mechanism through intangible inputs adoption to reproduce pro-cyclical markups in a sticky-price NK framework, and second, the mechanism proposed helps rationalizing my novel cross-sectional empirical finding.

Finally, the paper is related to an emerging literature studying the rise of business intangible and its macroeconomic implications. A large literature documents the salient rise of intangible capital, both in the US and around the globe (Corrado, Hulten and Sichel, 2009; Haskel and Westlake, 2018; Bhandari and McGrattan, 2021).[19] I focus on intangible inputs and its impacts on the cost structure of firms, as opposed to intangible capital and its implication for the capital structure studied extensively in the literature. Notable examples include Crouzet and Eberly (2019), who examine intangible capital's role in explaining weak investment in physical capital. Focusing on monetary policy, Döttling and Ratnovski (2021) contend that the rise of intangible capital makes monetary policy less effective in stimulating investment. Others emphasize the role of one type of intangible capital - customer capital, and its implications for firm dynamics (Gourio and Rudanko, 2014), firm price setting (Paciello, Pozzi and Trachter, 2019; Roldan-Blanco and Gilbukh, 2021), industry concentration (Morlacco and Zeke, 2020; Afrouzi, Drenik and Kim, 2020), and economic growth (Cavenaile, Celik, Roldan-Blanco and Tian, 2021).

Perhaps the more closely related paper is De Ridder (2019), who studies the implication of intangible inputs on long-run trends in corporate market power and industry concentration. Another closely related paper is Altomonte, Favoino, Morlacco and Sonno (2021), who shows that the less financially constrained firms invest more in intangible inputs and charge higher markups. These papers do not focus on monetary policy, or the link between markup dynamics and inflation. In contrast, I explore the role of intangible inputs in explaining firms' markup dynamics over the business cycles, and in particular, the connection between pro-cyclical markups and inflation.

Roadmap. The paper proceeds as follows: Section 2 provides an overview of the data and the empirical specification, then presents the motivational evidence from both aggregate- and firm-level data. Section 3 develops a simple theory model, illustrates the mechanism proposed, and discusses some of the key implications. Section 4 then proceeds to test the proposed mechanism by checking if the data support some of the key testable predictions of the model. Section 5 discusses several alternative mechanisms and argues that none of them, on its own, could plausibly explain my findings. Section 6 offers some concluding remarks.

[18] Hyun, Kim and Lee (2021) use a flexible translog production function to induce pro-cyclical markups. I keep the CES structure.

[19] The literature largely focuses on trends. A firm's heterogeneous ability to scale up gives the firm a competitive edge over its rivals, leading to the emergence of the "superstar" firms (Autor, Dorn, Katz, Patterson and Van Reenen, 2020), which in turn has implications for industry concentration and business dynamism (Akcigit and Kerr, 2018; Akcigit and Ates, 2019).

2 Motivational Evidence: Pro-cyclical Markups

This section introduces the data and outlines the empirical specification before turning to empirical evidence. I focus on the two main firm-level variables of interest, namely, Sales, General, and Administrative expenses (SG&A) and markup. I then present the motivational evidence from firm-level data. I first show that markups are conditionally pro-cyclical to a monetary policy shock, and in the cross-section, firms with high intangible intensity display more pro-cyclical markups than those that rely less on intangible-intensive technologies.

2.1 Data and Empirical Framework

Data sources and variables definition. In order to construct firm-level markups, I use quarterly balance-sheet data for U.S. publicly listed firms from Compustat for 1980 - 2009. The sampling period starts before the main monetary policy shock series (1990Q1 - 2009Q4) because I estimate some series of markups in backward-looking rolling windows.[20] I obtain quarterly aggregate time-series data from national accounts to use as macroeconomic controls that include the interest rate, real output, inflation, unemployment, and credit spread. To obtain aggregate-level evidence as a benchmark, I further include various measures of aggregate markups estimated from sectoral-level data in Nekarda and Ramey (2020).[21]

For firm-level data from Compustat, I define industry primarily at 3-digit NAICS level, such that my definition of market share and specification of sectoral production function used in the estimation of markups are consistent throughout.[22] I further use firm-level data from annual Compustat as a supplement to the quarterly data for only the former contains relevant information such as firm-level foreign income. Finally, I apply a standard cleaning procedure to the balance-sheet data to construct a baseline sample. Details on data sources and variable construction can be found in Appendix B.

Measure of Intangible intensity. The first firm-level variable central to my analysis is intangible intensity, which measures the ratio of a firm's expenditure on intangible inputs over total operating expense. I follow Peters and Taylor (2017) and define my measure of intangible input as SG&A net of R&D expendi-

[20] Production function may vary both over time and across sectors. Using a backward-looking rolling windows allows for estimation of markups in a sector-time-specific production function. For instance, markups for 1990Q1 can be estimated using data from the past ten years (1980Q1 - 1990Q1), and so forth.

[21] These markup series are estimated based on labor input margins with a wide range of adjustments made to take into account complications introduced by the presence of the overhead labor and the non-unitary elasticity of substitution between labor and capital in a flexible CES production function.

[22] A number of my exercises call for a definition of sector, for instance, in order to define market share, and to specify a sectoral production function in order to estimate markups. The former calls for a narrowly defined sector, while the latter requires a sufficiently large sample size per sector to consistently estimate markups. To strike a balance, I choose to focus on 3-digit sectors for the main analysis.

ture.[23] SG&A appears on the income statement as a selling expense that includes expenditures on various business intangibles (for instance, R&D, sales, advertising, and marketing, and expenses on IT staff). I remove R&D to alleviate the concern that at least a fraction of SG&A represents an investment for future profitability, for the benefits of R&D typically lag behind the expenditures.[24] In Appendix B, I provide a detailed discussion of the matter and lay out my reasoning for my definition of intangible intensity, and in particular, why I perceive an increase in the ratio of SG&A over total operating costs as indicative of an increase in the use of intangible inputs in general.

Accordingly, my preferred measures of intangible intensity is the ratio between SG&A net of R&D and total operating costs (OPEX, Compustat item *xopr*); i.e., firms with a higher SG&A share of OPEX are relying more on intangible inputs in their production process.[25] Notably, the definition is distinct from the intangible capital intensity defined in Peters and Taylor (2017) as a ratio of intangible capital to total capital stock, and it is for this reason that I do not take externally purchased intangible capital (measured by Compustat item *intan*) into account in my measure of intangible intensity. The focus is on intangible input intensity in the cost structure of firms.

Markup estimation. Markup is the second key variable in my analysis. In order to estimate firm-level markups, I closely follow De Loecker and Warzynski (2012) to estimate the output elasticity of a variable input using COGS (Costs of Goods Sold) as a proxy for the variable cost in a production function. Firm-level markups are then recovered as the ratio between the output elasticity and the input's cost share of total revenue. The goal of the exercise is to deploy a conventional empirical methodology to estimate markups in a consistent way. [26] Details of markup estimation and a summary statistics for estimated series of markups are relegated to Appendix C.

[23] That is, I measure SG&A as Compustat item *xsga* minus *xrd* minus *rdip*, with adjustments made according to the procedure proposed in Peters and Taylor (2017).

[24] Interpreting SG&A under the current accounting practices is not straightforward, as it is difficult to classify intangible expenditure as an investment in capital stock that supports future profitability versus an input that contributes to current profits. Peters and Taylor (2017) treats 30% of SG&A net of R&D as intangible capital expenditures, while the remaining 70% as intangible inputs. I choose not to take the 70% fraction of SG&A in my main analysis, for the uniform fraction would not change the distribution of firm-level intangible intensity. There exists, however, evidence suggesting that there are industrial heterogeneity in fractions of SG&A as intangible inputs (Ewens, Peters and Wang, 2019).

[25] Firm-level SG&A (Compustat item *xsgaq*) is highly volatile and features significant lumpiness. I take a backward-looking (current and previous three quarters) moving averages to smooth the measure of SG&A net of R&D. Later on, when intangible intensity is used as a conditioning firm characteristic variable, whether or not a moving average is used does not change the results. However, when SG&A is used as a dependent variable in Section 4, taking a moving average helps to remove the initial jump of the impulse response function that I attribute to the lumpiness of the micro data.

[26] There is an active debate over this methodology and its application on firm-level revenue data that goes beyond the scope of this paper. In particular, there are concerns over identification and estimation of markups over the usage of the ratio estimator on firm-level revenue data that are typically devoid of separate prices and output quantities data. At the core of the criticism is the use of firm revenue data to proxy for gross output, a real quantity (Bond et al., 2021). In practice, revenue is purged out in the first stage of the procedure, and markups estimated by the De Loecker and Warzynski (2012) methodology do not in fact measure the revenue elasticity. Using detailed product-level data from France that contains information on prices, Burstein et al. (2020) compare the two set of markups estimated using revenue data versus quantity data and find that they are highly correlated with a correlation of 0.83.

Recent literature notes that the increase in SG&A as a share of total operating cost is precisely the type of technological advancement that may jeopardize the validity of the traditional COGS-based markups, resulting in an overestimation bias (Traina, 2018; Covarrubias, Gutiérrez and Philippon, 2020). Alternatively, I estimate OPEX-based markups using the adjusted operating costs (COGS + SG&A - R&D), following Traina (2018). One criticism of the approach is that incorporating SG&A into the variable cost introduces an additional source of measurement error as some fraction of SG&A should be classified under the fixed cost component, as opposed to the variable cost, and any heterogeneity in fixed cost across firms leads to a systematic underestimation of markups. Given the additional trade-off when taking SG&A into account, I focus on the traditional COGS-based markups in the main text, and show that my main results are robust to alternative measures of markups in Appendix E.

Finally, the production function-based estimation approach calls for a pre-specified production function and a definition of sector. I estimate markups with both a Cobb-Douglas and a more flexible translog production function in 3-digit sectors.[27] As a robustness check, I estimate markups in a backward-looking five-year rolling window for a time and 2-digit-industry specific production functions.

Identification and empirical specification. Estimating the dynamic causal effects of monetary policy on markups relies on overcoming the familiar issue of reverse causality in empirical macroeconomics; namely, interest rate changes respond to the underlying macroeconomic conditions and affect them at the same time. For identification, I rely on high-frequency identified monetary shocks (GSS shocks) from Gürkaynak, Sack and Swanson (2005) with an extension by Gorodnichenko and Weber (2016), i.e., monetary policy shocks derived from high-frequency movements of Fed Funds future contracts within a sufficiently narrow window of time around FOMC announcements. The identification assumption holds as long as no event happens during these narrow windows would affect the action of the private sector as well as the decision of the central bank. Following convention in the macro literature, I sum the monthly shock series to quarterly frequency. I then employ a Local Projection Instrument Variable (LP-IV) specification following Jordà and Taylor (2016); Stock and Watson (2018), in which the interest rate changes are instrumented by the identified monetary policy shocks, to estimate the dynamic causal effects of monetary policy on aggregate markups. For the firm-level data, I use a panel LP-IV setup to analyze both the average effect of a monetary easing on markups and the potential heterogeneity based on firm characteristics.

[27] I focus on 3-digit sectors and not the more finely defined 4-digit sectors nor the broadly defined 2-digit sectors, because it balance the demand for sufficient number of firms per sectors to estimate markups consistently, and the need to compute market shares based on a narrowly defined sector later on.

2.2 Firm-level Evidence

I use firm-level panel data to estimate the dynamic causal effects of interest rate changes on markups in a panel LP-IV set up. I first estimate the average impulse response functions for markups following a monetary easing; I then identify the heterogeneous effects of a monetary easing on markups across firms according to their characteristics. In Appendix E.1, I show the dynamic responses of the aggregate markups and an assortment of macroeconomic variables to interest rate changes as a benchmark. In particular, I show that real GDP and inflation respond pro-cyclically to a monetary easing, and the empirical specification I adopt produces sensible impulse responses consistent with those reported in Gertler and Karadi (2015). Additionally, I show that the median real wage also responds pro-cyclically to a monetary easing, consistent with the VAR evidence in Christiano et al. (2005). Finally, various series of the aggregate markups from Nekarda and Ramey (2013) also respond pro-cyclically to an interest rate cut. The peak effect is typically reached one year after the shock and ranges between 2-4%.

Average effects. To estimate the average dynamic causal effects, I run the following second stage regressions in a panel:

$$\Delta \log y_{i,t+h} = \alpha_i^h + \beta^h \Delta \hat{R}_t + \Gamma_l^h L(p).Y_t + \gamma_l^h L(p).x_{i,t} + \epsilon_{i,t+h} \tag{1}$$

Time is denoted by t and firm by i. The dependent variable y is the variable of interest - log of firm-level markups in this Section, and later log of SG&A and sales in Section 4. On the right hand side, β^h is the parameter of interest that captures the cumulative average effect of interest rate changes on markups at each horizon. I include firm-level fixed effect α_i^h, and quarterly dummies, in addition to a standard set of macroeconomic controls in lags of Y_t.[28] Additional firm-level controls in $x_{i,t}$ are lags of log total asset as a proxy for size, and liquidity and leverage are proxies for financial frictions. Standard errors are clustered at firm-time level using the method of Driscoll and Kraay (1998) to correct for possible serial correlation in the forecast errors, as a standard practice in the local projection method. Finally, I focus on the sample of firms that have a minimal of 16 observations, in accordance with the forecast horizon of 16 quarters.

Figure 1 plots the impulse response functions for my baseline measure of firm-level markups (estimated in a 3-digit sector-specific Cobb-Douglas production function based on COGS). It shows that following a one percent cut in the interest rate, average markups increase significantly and reach a peak of around 1.5% one year and a half after the shock. The shape of the dynamic response is largely consistent with my aggregate benchmark. The magnitude of the peak effect is on the lower end of the spectrum compared to those across the aggregate markups. Given the differences in both the estimation methodology and the samples used, I interpret the finding as supportive of the pro-cyclical responses of markups to a monetary easing in a sample

[28] Specifically, I include four lags of the Fed Fund rate, the monetary policy shock, changes in log real GDP, credit spread, unemployment rate, and changes in log CPI and log commodity price index.

of public firms.

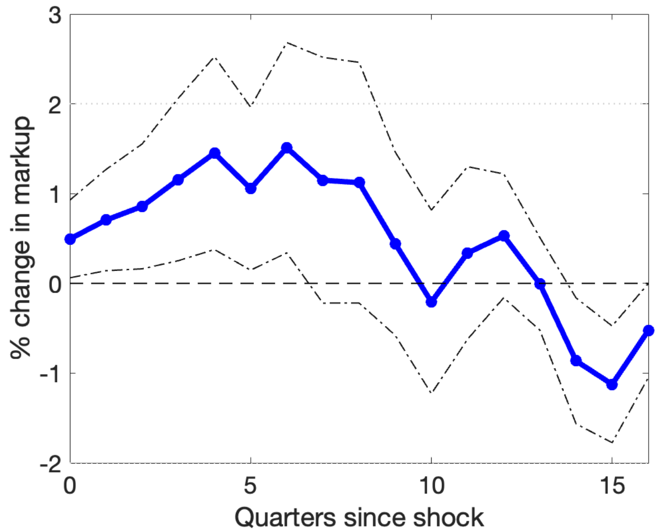


Figure 1: Average Effects of Monetary Policy Shocks on Markups

Note: The figure shows the impulse responses of the baseline measure of markups (estimated in a 3-digit sector-specific Cobb-Douglas production function based on COGS) in LP-IV regressions with high-frequency identified shocks (GSS shocks) from Gürkaynak et al. (2005) instrumenting for one percentage decrease in the Fed Funds rate. The sampling period is 1990Q1-2009Q4 (corresponding to the span of GSS shoocks), and the forecast horizon is 16 quarters. The regression does not allow for possible heterogeneous responses across firms, and hence the IRFs obtained represent the average effect of an expansionary monetary policy shock across all firms in the sample, and form the basis of comparison with the aggregate IRFs obtained from the time-series data. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

The natural followup question arises: is there any heterogeneity in the cross-section? And if so, what characterizes the firms that display the more conditionally pro-cyclical markups? The answers to these questions are *ex ante* ambiguous. In standard theory models with variable elasticity of demand, when facing an increase in the marginal cost, if firms are able to freely adjust their prices, those with higher market shares are more inclined towards acting strategically and absorbing a higher fraction of the cost shocks into their markups, and therefore would display a more counter-cyclical flexible-price markup.[29] The presence of pricing frictions further muddles the picture, especially if there is heterogeneity in price-setting frictions across firm sizes.[30]

Importantly, neither this type of strategic consideration nor nominal pricing rigidities could explain the

[29] Amiti et al. (2019) show that in a flexible-price setting with non-CES demand and oligopoly, both *heterogeneity* in pass-through elasticities and cost shocks are needed to generate relative price movement (and hence markup movement). Therefore, when all firms are facing the same marginal cost shock (as in the case of a monetary policy shock), and in the absence of any nominal pricing friction, there would be a complete pass-through of the cost shock into prices, which leaves markups unchanged.

[30] Empirically, the evidence is not clear cut. Large firms are found to changing prices more frequently, but by smaller amounts (Goldberg and Hellerstein, 2011). That is, whether or not large firms have more rigid price depends on which whether we examine the intensive margin of adjustment sizes versus the extensive margin of adjustment frequencies. In the theory section, I hold nominal rigidity fixed for all firms, and the relationship is therefore outside the scope of this paper.

pro-cyclical average effect. A pro-cyclical markup implies a counter-cyclical real marginal cost. Recent literature has linked scalable production technologies and network economies of scale to lower marginal cost and the rise of market concentration. I therefore examine the interaction between interest rate changes and intangible intensity for intangible inputs are associated with economies of scale, and thus may have implications for the dynamics of markups.

Cross-sectional heterogeneity. In order to estimate the heterogeneous effects of monetary policy on markups, I follow Cloyne, Ferreira, Froemel and Surico (2019) to estimate a flexible semi-parametric specification based on the distribution of firm characteristics $z_{i,t-1}$ from the previous quarter. A firm characteristic $z_{i,t-1}$ takes the form of an indicator function that equals to one if the firm falls into a particular group based on - for instance, higher intangible intensity versus lower intangible intensity. The flexible set up avoids any linearity assumption imposed in the interactions between firm characteristics and interest rate changes, and allows for highly versatile grouping of firms along multiple dimensions of firm characteristics.

I first document the empirical link between firm-level markups and intangible intensity. Table 1 shows that there exists a positive correlation between intangible intensity and markups at the firm level. The result is robust to controlling for additional firm-level variables, as well as firm-fixed effects and time-fixed effects; i.e., within a firm, higher markup is associated with higher intangible intensity. The finding is consistent with the recent literature that links firms' adoption of intangible input to levels and elasticities of markups (Altomonte et al., 2021), pointing to the possibility that firms with high intangible intensity might be driving the pro-cyclical responses of markups.

Table 1: Firm-level relationship: markups on intangible intensity

	(1)	(2)	(3)
Intangible intensity	0.418^{***}	0.393^{***}	0.366^{***}
	(0.021)	(0.018)	(0.018)
Observations	182597	182597	176271
Firm f.e.	Yes	Yes	Yes
Time f.e.	No	Yes	Yes
Firm-level controls	No	No	Yes

Note: Firm-level relationship between markups and intangible intensity. The dependent variable is the baseline measure of markups, and the independent variable is the baseline measure of intangible intensity (as the ratio ïŒ$SG\&A - R\&D)/OPEX$). Regression (1) contains firm-level fixed effect. Regression (2) contains both firm- and time-fixed effects. Regression (3) contains additional firm-level controls that include: size (log(atq)), leverage, and liquidity. Standard errors in parentheses are clustered at firm-time level. * $p < 0.1$,** $p < 0.05$ *** $p < 0.01$

I then estimate impulse response functions in the following second stage LP-IV panel regression in Equation 2. The baseline firm characteristic $z_{i,t-1}$ is intangible intensity, which measures the extent to which firms rely on intangible inputs in their cost structure.

$$\Delta \log y_{i,t+h} = \alpha_i^h + \sum_{g=1}^{G} \beta_g^h \cdot \mathbf{1}[z_{i,t-1} \in g] \cdot \Delta \hat{R}_t + \sum_{g=1}^{G} \zeta_g^h \cdot \mathbf{1}[z_{i,t-1} \in g] + \Gamma_l^h L(p).Y_t + \gamma_l^h L(p).x_{i,t} + \epsilon_{i,t+h}, \quad (2)$$

I likewise include firm-level fixed effect, quarterly dummies, and a rich set of firm-level and aggregate controls as before. Indicator functions for firm characteristics interact with both the interest rate changes in the second stage and the GSS shocks in the first stage. The group-specific coefficient β_g^h thus captures the potential heterogeneous responses of markups to interest rate changes according to group characteristics at each horizon h.

Figure 2 reports the baseline results in the cross-section; firms with high intangible intensity display more pro-cyclical markups than those with low intangible intensity. Specifically, I allow the effects of monetary policy to vary across the distributions of firm-level intangible intensity such that firms fall into three groups based on high, medium, and low intangible intensity.[31] I divide the distribution equally into three parts. Low intensity corresponds to firms with intangible intensity below the 33th percentile of its distribution in the previous quarter, whereas high intensity corresponds to firms with intangible intensity above the 66th percentile.

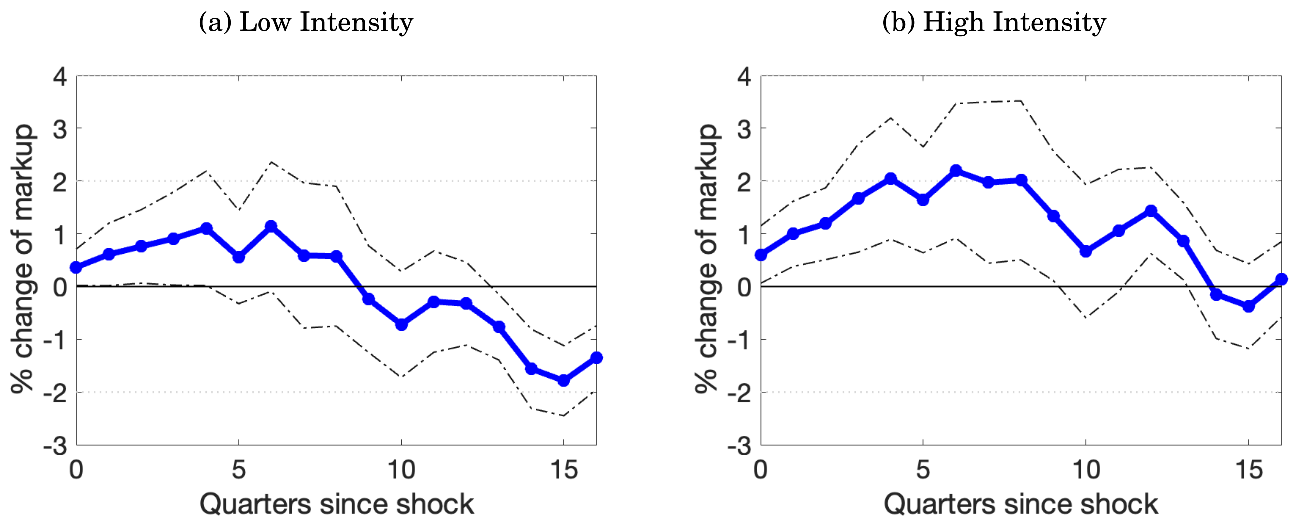


Figure 2: Heterogeneous Effects of Monetary Policy Shocks on Markups

Note: The figure shows the impulse responses of the baseline measure of markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The sampling period is 1990Q1-2009Q4, and the forecast horizon is 16 quarters. The interest rate is interacted with a binary indicator based on the distribution of firm-level intangible intensity, in which low (high) intensity corresponds to firms with intangible intensity below the 33th percentile (above the 66th percentile) of its distribution in the previous quarter. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

[31]Because of the lumpiness in SG&A in Compustat, I use a backward-looking (current and previous three quarters) moving averages to smooth SG&A before computing intangible intensity. When used as a conditioning variable, whether or not to to take the additional step of smoothing does not alter my results.

Firms that are characterized by high intangible intensity in cost structure are showing the most pro-cyclical markups, reaching a peak of 2.1% after one year and a half, whereas firms with low intangible intensity display the least pro-cyclical markups that reach a peak of 1% and turn counter-cyclical after two years. The medium group shows impulse responses that fall in between. For expository purpose, I show the impulse responses for the bottom and top tertiles in the main text and relegate the full set of impulse responses to Appendix E.2.2.

Taking together, these findings suggest that in a monetary policy-induced expansion, an average firm's marginal profitability increases. In other words, profits are conditionally pro-cyclical to a monetary easing. Furthermore, firms with high intangible intensity in their production process, become even more profitable.[32]

2.3 Robustness

I verify the robustness of the baseline findings to alternative measures of markups and monetary policy shocks. Specifically, I use a variety of markups that include: (1) OPEX-based markups, (2) markups estimated with a more flexible translog production function, (3) markups estimated in a backward-looking five-years rolling windows in 2-digit sector.[33] Using the narratively-identified Romer and Romer (2004) shocks, I test if the results are robust to an alternative identification strategy for monetary policy shocks. In Appendices E.2.1 and E.2.2, I show that both my average and cross-sectional heterogeneity results are robustness to these checks, respectively.

Furthermore, I use an alternative specification to test for heterogeneity by imposing a linear interaction between intangible intensity and interest rate changes. I also check if the heterogeneity remains robust after controlling for alternative firm characteristics, in particular, financial frictions, as the working capital assumption operates through the credit channel of monetary policy. Contrary to its implication that the financially constrained firms would exhibit more pro-cyclical markups, I find evidence that suggests the opposite.[34] My baseline cross-sectional heterogeneity results are robust to these sensitivity analyses.

To summarize, the key empirical finding is that markups are conditionally pro-cyclical to monetary policy surprises, and in the cross-section, the pro-cyclical responses are most pronounced among firms that utilize a greater share of intangible inputs in their operation. The baseline results are broadly robust to alternative

[32]It should be noted that if upfront costs are incurred in the production technology, a firm with positive upfront costs would have a lower average profit margin than its marginal profit margin as measured by markups. In regressions that I did not report here, I find that the heterogeneity based on intangible intensity holds for firm profitability measured by gross profit margin (a non-parametric way of computing firm profitability).

[33]To address the concern that OPEX-based markups are estimated with a variable cost measure that contains heterogeneous fixed cost components, I also use an alternative specification in which I treat SG&A net of R&D and COGS are separate inputs, and compute markups based on COGS alone. Both my average and cross-sectional results remain robust.

[34]Results based on imposing the linear interaction and conditioning on alternative firm-level characteristics can be found in Appendix E.2.2

measures of markups and monetary policy shocks.

3 Theory: a CES-oligopoly Model with Intangible Inputs Adoption

Motivated by the empirical findings, I use a heterogeneous firm New-Keynesian (NK) model to study the effect of intangible inputs on the dynamics of markups. The main departure from the literature is that I allow a firm to optimally adopt intangible inputs to alter its cost structure and shift demand towards its product. The focus on markup heterogeneity requires a departure from the typical CES demand system and monopolistic competition in the NK literature, for they collectively imply constant markups across firms, I instead allow a finite number of firms of heterogeneous productivity in a given sector to compete strategically in a static Cournot game in a setup that resembles the CES-oligopoly models studied in the real rigidities literature.

The other key ingredients of the model are: (1) a nested-CES demand structure augmented with a good-specific demand-shifter (Atkeson and Burstein, 2008; Ravn, Schmitt-Grohé and Uribe, 2006; Gilchrist, Schoenle, Sim and Zakrajšek, 2017),[35] (2) individual good firms face quadratic price-setting costs à la Rotemberg (1982), (3) labor unions face wage-setting frictions à la Calvo (1983).[36] The model (i) generates pro-cyclical aggregate markup, at the same time, preserves pro-cyclical real wage and inflation, and (ii) predicts that in the cross-section, larger firms (with higher market share) in a sector adopt intangible inputs aggressively, and are driving the overall pro-cyclical markups following a monetary easing.

3.1 Environment

Final Good Production. The final consumption good remains the Dixit-Stiglitz aggregation of imperfectly substitutable sectoral goods, but now there are two layers of production taking place in a nested structure, in which (a) the final good firm aggregates from a continuum of differentiated sectoral goods, and (b) a continuum of sectoral good firms aggregate from a finite number of differentiated intermediate goods produced by individual good firms within each sector.

Specifically, the economy is supplied by a continuum of sectors indexed by $j \in [0,1]$. Time is discrete. At time t, aggregating from sectors $j \in [0,1]$ to the economy gives the final consumption good Y_t according to:

$$Y_t = \left(\int_0^1 X_{jt}^{\frac{\eta-1}{\eta}} dj\right)^{\frac{\eta}{\eta-1}}, \tag{3}$$

[35] The demand-shifter takes a similar form as a good-specific habit stock in the deep habit models. Relative to these models, I model a firm's ability to shift demand, conditional on relative price, through its engagement in sales/advertising/marketing.

[36] The household side of the model is similar to the baseline NK model in Galí (2015), as such I rely on Calvo frictions for wage rigidities. The reason why I instead use Rotemberg frictions in pricing rigidities is that Calvo frictions introduce another source of heterogeneity to a heterogeneous firm model, rendering the model intractable as soon as the continuum of firms assumption is relaxed.

where X_{jt} is the demand-adjusted sectoral good, and η is the between-sector elasticity of substitution.

In each sector j, a finite number of N_j individual good firms indexed by i supply differentiated goods y_{ijt} at time t, and the sectoral good is aggregated from demand-adjusted individual goods according to:

$$X_{jt} = \left[\sum_{i=1}^{N_j} \left(\frac{y_{ijt}}{b_{ijt}^{\theta}} \right)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}}, \theta < 0, \rho > \eta > 1, \tag{4}$$

where y_{ijt} is the output of an individual good firm i in sector j, b_{ijt} is the demand-shifter associated with the individual good ij, and ρ denotes the within-sector elasticity of substitution. The good-specific demand-shifter is taken as given by the final good firm and is thus external.[37] I consider a finite within-sector elasticity of substitution $\rho < \infty$ such that within each sector, goods supplied by individual firms are imperfect substitutes. Moreover, the elasticity of substitution within-sector ρ is assumed to be greater than that of between-sector η, consistent with empirical evidence.

A perfectly competitive final good firm produce the final good by solving the dual problem of cost minimization, in which it chooses quantities of sectoral goods X_{jt} to achieve a given level of output. Likewise, perfectly competitive sectoral good firms choose the quantities of each individual good y_{ijt}, taking its demand-shifter and price as given to minimize production costs given levels of output. This implies the following demand function for individual good y_{ijt}:

$$y_{ijt} = \left(\frac{P_{ijt}}{\tilde{P}_{jt}} \right)^{-\rho} b_{ijt}^{\theta(1-\rho)} \left(\frac{\tilde{P}_{jt}}{\tilde{P}_t} \right)^{-\eta} Y_t, \tag{5}$$

where P_{ijt} is the price of individual good i in sector j at time t. $\tilde{P}_{jt} \equiv \left[\sum_{i=1}^{n_j} (P_{ijt} b_{ijt}^{\theta})^{1-\rho} \right]^{\frac{1}{1-\rho}}$ is the ideal sectoral price index, and $\tilde{P}_t \equiv \left(\int_0^1 \tilde{P}_{jt}^{1-\eta} dj \right)^{\frac{1}{1-\eta}}$ is the ideal aggregate pricing index. The good-specific demand-shifter b_{ijt} increases consumer demand for the individual good conditional on its relative price. For instance, through advertising and marketing, firms can improve the attractiveness or appeal of their products to consumers or increase the "perceived" quality of their products, such that for a given relative price, a larger b_{ijt} is associated with an increased demand for the individual good.

Individual Good Firm's problem. Strategic interaction occurs within a sector and among individual good firms heterogeneous in their productivity in the form of a static Cournot game. Importantly, the model has two *sources* of markup variation, in which market power arises from (1) cost advantage from high measured productivity, and (2) a firm's ability to influence and shift consumer demand. Measured productivity

[37] The demand-shifter functions similarly as a "habit stock" from the deep habit literature. Within a sector, the set-up is similar to a quasi-product deep habit models as in Gilchrist et al. (2017). I likewise follow the convention in the deep habit literature such that the demand-shifter operates at the individual good-level is assumed to be external (Ravn et al., 2006; Ravn, Schmitt-Grohė and Uuskula, 2010).

consists of two parts: an exogenously given productivity level z_{ij} and an endogenous cost advantage from the firm's adoption of the scalable technology. It follows that through the use of the first intangible input - a scalable production technology, firms could vary the first type of market power by increasing their measured productivity, and through the second intangible input - sales/advertising/marketing - firms could vary its second type of market power by directly influencing their demand.

Specifically, in a given sector, a finite number of individual good firms decide on the quantities of labor to hire for three purposes: (1) implementation of a scalable production technology to reduce marginal cost at the expense of a convex adjustment cost, (2) engagement in sales/advertising/marketing to influence consumer demand through the demand-shifter b, and (3) actual production of the individual good. Each period, they also set prices subject to Rotemberg frictions; i.e., whenever firms change prices, they pay quadratic adjustment costs in units of the final good.

Individual good firms produce according to the following production technology:

$$y_{ijt} = \frac{1}{1-f_{ijt}} z_{ij} l_{p,ijt}, \tag{6}$$

where $l_{p,ijt}$ is the production labor used, f_{ijt} is the amount of the scalable production technology adopted, and z_{ij} is the firm-level productivity that is assumed to be permanent and provides the only source of heterogeneity in the model. The nominal marginal cost of producing good i is therefore $MC_{ijt} = (1-f_{ijt})\frac{W_t}{z_{ij}}$, and it is declining in both a firm's productivity z and the amount of the scalable production technology adopted (i.e., measured productivity is increasing in f). The use of the scalable technology comes at the cost of a strictly convex expenditure $F(f_{ijt})\forall f_{ijt} \in [0,1)$ and it has the following properties: (1) $F' > 0$, $F'' > 0$, (2) $\lim_{f\to 1} F = \infty$. Following De Ridder (2019), I adopt the following function in the amount of labor $l_{f,ijt}$ hired to implement the scalable production technology as an upfront convex cost:

$$F(f_{ijt}) = (1-\phi_F)\left[\left(\frac{1}{1-f_{ijt}}\right)^{\psi_F} - 1\right]. \tag{7}$$

The function implies a zero adjustment cost, should a firm choose not to implement any scalable production technology, it also makes it practically impossible for firms to reduce its marginal cost to zero.

Likewise, I assume that firms could also vary its demand-shifter b_{ijt} by hiring sales, advertising, and marketing labor $l_{s,ijt}$, should they choose to incur a symmetric upfront cost in units of sales labor as that of the scalable production technology. At the core of the setup is the assumption that firms engage in marketing and advertising to shift demand and build market shares by influencing customers' willingness to pay.[38]

[38] A large literature on customer markets instead builds on an inter-temporal trade-offs inherent in the pricing decision, in which firms temporarily lower markups in order to attract customers, and later raise markups as their customers are locked in (Nakamura and Steinsson, 2011; Ravn et al., 2006; Gilchrist et al., 2017). Fitzgerald and Priolo (2018) find empirical evidence that supports the marketing and advertising view of market share building. Specifically, they find that markups remain the same upon a firm's

At time t in a given sector j, each individual good firm i takes its rivals' (denoted by $-i$) present and future prices $\{P_{-ij\tau}\}\forall\tau \geq t$ and relevant aggregate variables as given; a firm optimally chooses the amount of scalable production technology f_{ist} to adopt and commits to pay the associated convex cost $F(f_{ijt})$, as well as the optimal level of its demand-shifter b_{ijt} by hiring the required sales, advertising, and marketing labor and paying analogously the convex cost $F(b_{ijt})$. The intangible input adoption decisions are assumed to be static. It then sets its own price subject to Rotemberg adjustment costs and chooses quantity y_{ijt} to produce in a static Cournot quantity game in a dynamic output-pricing decision, internalizing its production decision's impact on the level of the sectoral output.

The motivation for modelling the choices of intangible inputs as static decisions is twofold. First, both types of intangible inputs have high depreciation rates, such that firms need to spend considerable amount of resources each year to maintain constant levels of both. Take software, a typical example of a highly scalable input, for example, Li and Hall (2020) estimate a depreciate rate in the order of 30% and 40% annually; similarly for advertising, Corrado et al. (2009) find that advertising capital depreciates annually at a rate as high as 60%. Second, as noted in De Ridder (2019), an increasing proportion of software is sold to businesses as a service (SaaS), such that firms incur periodic costs for the right to use it (as opposed the traditional one-off payment for purchasing a piece of equipment). Similarly for advertising and marketing, firms often maintain a permanent staff of sales and marketing personals on their payrolls instead of hiring them on the spot when need arises. Even though the assumption is made in part for tractability, I view the setup as a parsimonious way to model a firm's periodic spending on intangible inputs in support of its *current* profits and hence it aligns with my focus on the cost structure (as opposed to capital structure) of firms.[39]

The only dynamic decision firms are making is the output-pricing decision subject to Rotemberg adjustment costs, and I make a similar behavior assumption as in Etro and Rossi (2015) and Corhay, Kung and Schmid (2020) that when setting prices, firms, despite their granularity in a given sector, still behave atomistically, taking competitors' current and future prices as given.[40] This assumption that firms are price-maker with respect to own prices but price-takers with respect to rivals' prices helps me avoid analyzing an otherwise analytically intractable Markov Perfect equilibrium (MPE) and go beyond the case of a duopoly ($N = 2$) when there is firm-level heterogeneity involved. The goal is to abstract from complicated inter-

successful entry while its market share grows. To best preserve tractability and to mirror the setup for the scalable production technology (especially since in the quarterly data, SG&A net of R&D is a broad measure that captures both types of intangible expenditures), I model customer acquisition through firms' engagement in sales/advertising/marketing in the same vein as Arkolakis (2010), Drozd and Nosal (2012) and more recently Afrouzi et al. (2020).

[39]Treating intangible inputs as capital and solving two dynamic problems (intangible capital investment subject to partial depreciation and output-pricing decisions subject to nominal frictions) that show interdependence is technically challenging.

[40]That is, firms form correct expectation about their competitors' current and future prices, but incorrectly assume that their price-setting today would not affect their competitors' future price setting. An emerging literature relaxes this assumptions and analyzes a dynamic oligopoly game with pricing frictions in a Markov Perfect Equilibrium. Recent examples include Mongey (2021), who studies a dynamic duopoly with stochastic menu costs, and Wang and Werning (2020), who analyse a symmetric dynamic oligopoly with Calvo frictions.

temporal strategic consideration, and at the same time, preserve the key elements needed to address my initial research question.

Solving Individual Good Firm's problem. Firms solve the maximization problem in two stages: in the static first stage, firms take prices as given and optimally choose the amount of the scalable production technology and the level of advertising/marketing, they commit to paying the associated convex costs upfront, and subsequently post their prices and produce the goods demanded. In other words, the assumption is that firms are myopic - they do not internalize pricing decisions' impacts on intangible investment (i.e., taking prices as given in the first stage). This assumption enables me to solve fully the dynamic equilibrium. In Appendix D, I relax this assumption to analytically study a model in which firms internalize the pricing decision in order to show that this assumption is unlikely to be pivotal for the main implications of the model.[41] The first-stage maximization problem is as follows:

$$\max_{f_{ijt},b_{ijt}} \left(P_{ijt} - \underbrace{(1-f_{ijt})\frac{W_t}{z_{ij}}}_{\text{marginal cost}} \right) \underbrace{\left(\frac{P_{ijt}}{\tilde{P}_{jt}}\right)^{-\rho} b_{ijt}^{\theta(1-\rho)} \left(\frac{\tilde{P}_{jt}}{\tilde{P}_t}\right)^{-\eta} Y_t}_{\text{demand}} \\ - W_t \underbrace{(1-\phi_F)\left[\left(\frac{1}{1-f_{ijt}}\right)^{\psi_F} - 1\right]}_{l_{ijt,f}} - W_t \underbrace{(1-\phi_B)\left[\left(\frac{1}{1-b_{ijt}}\right)^{\psi_B} - 1\right]}_{l_{ijt,s}}, \tag{8}$$

where W_t is the nominal wage, $l_{f,ijt}$ is the labor used to implement the scalable production technology, while $l_{s,ijt}$ denotes the labor hired for sales, advertising, and marketing.

FOCs with respect to f_{ijt} and b_{ijt} imply that:

$$\psi_F(1-\phi_F)(1-f_{ijt})^{-(\psi_F+1)} = \frac{y_{ijt}}{z_{ij}} \tag{9}$$

$$\psi_B(1-\phi_B)(1-b_{ijt})^{-(\psi_B+1)} W_t = \theta(1-\rho)\left(1-\frac{1}{\mu_{ijt}}\right)\frac{P_{ijt}y_{ijt}}{b_{ijt}} \tag{10}$$

The two first order conditions link the the marginal cost of adopting intangible inputs on the left hand side to the marginal benefit of acquiring them on the right.

[41]Specifically, in the first theoretical extension, I show that it is possible to solve the problem in the forward order to internalize the impacts of prices on the optimal choices of intangible inputs. Mathematically, it adds a number of wedges to the optimal pricing equation. Numerically it proves to be a challenge to implement when solving for the full dynamic equilibrium. I instead log-linearize the competitive equilibrium, and show that under specific assumptions, the main takeaways of the baseline model would be preserved. Alternatively, I solve fully the dynamic equilibrium of a flexible-price model in a backward manner to show that the main insights of the model remain unchanged. This is because in a sticky-price model, it is technically challenging to solve the problem backwards, for the optimal pricing expression is fairly complicated, and in particular is a function of the past and future expected variations in prices. The challenge is mitigated in a flexible-price model. However, the goal is to produce conditionally pro-cyclical markups in a sticky-price model, so while removing pricing rigidities alleviates the technical challenge, it does not help as much in terms of answering my original research question.

In the second stage, given rivals' prices and intangible input choices, firms solve a dynamic output-pricing problem subject to Rotemberg costs to maximize the present discounted values of profits. Through a static Cournot game, they internalize their production decisions' impact on the sectoral output. They solve the following maximization problem subject to the inverse demand function they face:

$$\max_{\{P_{ij\tau}, y_{ij\tau}\}_{\tau\geq t}} \quad \mathbf{E}_t \sum_{\tau\geq t} \frac{Q_{t,\tau}}{P_\tau} \left\{ \left(P_{ij\tau} - (1-f_{ij\tau})\frac{W_\tau}{z_{ij}}\right) y_{ij\tau} - \underbrace{\frac{\phi_p}{2}\left(\frac{P_{ij\tau}}{P_{ij,\tau-1}} - 1\right)^2 P_{ij\tau} y_{ij\tau}}_{\text{Rotemberg costs}} \right\} \tag{11}$$

$$\text{s.t. } P_{ij\tau} = \left(\frac{y_{ij\tau} b_{ij\tau}^{\theta(\rho-1)}}{X_{j\tau}}\right)^{-\frac{1}{\rho}} \left(\frac{X_{j\tau}}{Y_\tau}\right)^{-\frac{1}{\eta}} \tilde{P}_\tau,$$

where $Q_{t,\tau} = \beta \frac{C_\tau^{-\sigma}}{C_t^{-\sigma}}$ is the stochastic discount factor, and ϕ_p denotes the Rotemberg coefficient for pricing friction. Production labor choices are embedded in the equation and is otherwise defined as $l_{p,ij\tau} = \frac{y_{ij\tau}(1-f_{ij\tau})}{z_{ij}}$.

The first order conditions imply that the optimal price is a markup over the firm's marginal cost:

$$P_{ijt} = \mu_{ijt}(1-f_{ijt})\frac{W_t}{z_{ij}}, \tag{12}$$

where

$$\mu_{ijt} = \frac{\Theta_{ijt}}{(\Theta_{ijt}-1)\left[1-\frac{\phi_p}{2}(\Pi_{ijt}-1)^2\right] + \phi_p \Pi_{ijt}(\Pi_{ijt}-1) - \Gamma_{ijt}}$$

$$\Theta_{ijt} = \left[\frac{1}{\rho} + (\frac{1}{\eta} - \frac{1}{\rho})s_{ijt}\right]^{-1}, s_{ijt} = \frac{P_{ijt}y_{ijt}}{\sum_{l=1}^{n_j} P_{ljt}y_{ljt}} = \frac{(P_{ijt}b_{ijt}^{\theta})^{1-\rho}}{\sum_{l=1}^{n_j}(P_{ljt}b_{ljt}^{\theta})^{1-\rho}} = \left(\frac{\tilde{P}_{ijt}}{\tilde{P}_{jt}}\right)^{1-\rho}$$

$$\Gamma_{ijt} = \phi_p \mathbf{E}_t \left[Q_{t,t+1}\Pi_{ij,t+1}^2(\Pi_{ij,t+1}-1)\frac{y_{ij,t+1}}{y_{ijt}}\right]$$

$$\Pi_{ijt} = \frac{P_{ijt}}{P_{ij,t-1}}$$

The output-pricing decision leads to optimal time-varying markups μ_{ijt} that depends on both the firm's market share s_{ijt} and the degree of nominal rigidities ϕ_p. The firm's market share is in turn determined by its relative ideal price $\tilde{P}_{ijt}$. Conditional on actual sticker prices P_{ijt}, the more a firm spends on advertising/marketing, the lower its ideal relative price would be, and the greater the market share it can thus gain. Since there exits an direct, inverse relationship between a firm's market share and its variable elasticity of demand Θ_{ijt}, this also implies that the firm is able to operate on the less elastic portion of its demand curve, and hence enjoys greater market power as evidenced by a higher markup.

Household Preferences. The economy is populated by a continuum of infinitely lived heterogeneous households that consume final good C_t and supply labor $N_t(h)$ to a labor union. Households supply differentiated labor varieties h, which give them some monopolistic power to set their own wages $W_t(h)$ subject to Calvo frictions. Each household has the following preferences over a final consumption good and leisure. They save by investing in a one-period risk-free government bond B_{t+1} that pays the nominal interest rate i_t. In addition to their wage incomes, households receive Φ_t in dividends from an equal share of all firms' profits each period. Similar to the setting in Erceg, Henderson and Levin (2000), I assume that there exist complete markets for state-contingent securities that insure households against idiosyncratic wage risks due to the Calvo frictions, and since the preference is separable in labor and consumption, heterogeneous households become *de facto* identical in their consumption-saving decision (and the index h is thus omitted for C_t and B_{t+1}). Household h seeks to maximize its lifetime utility subject to its budget constraint and a labor demand function:

$$\max_{C_t, N_t(h), W_t(h), B_{t+1}} \mathbf{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma} - 1}{1-\sigma} - \theta_L \frac{N_t(h)^{1+\psi_L}}{1+\psi_L} \right) \right],$$

s.t.

$$P_t C_t + B_{t+1} \leq W_t(h) N_t(h) + (1 + i_{t-1}) B_t + \Phi_t,$$

$$N_t(h) = \left(\frac{W_t(h)}{W_t} \right)^{-\eta_L} N_t,$$

where P_t is the price of the final consumption good, W_t is the prevailing nominal wage, and N_t is the aggregate labor demand. $\beta \in (0,1)$ is the subjective discount factor, and σ is the inverse elasticity of intertemporal substitution. Moreover, ψ_L is the inverse of the Frisch elasticity of labor supply, θ_L is a parameter that governs the dis-utility of labor, and finally η_L is the elasticity of substitution between labor varieties.

Wage setting. A representative, perfectly competitive labor union hires labor varieties supplied by household h and aggregate them into a homogeneous composite of labor good N_t using a Dixit-Stiglitz aggregator:

$$N_t = \left(\int_h N_t(h)^{\frac{\eta_L - 1}{\eta_L}} dh \right)^{\frac{\eta_L}{\eta_L - 1}}.$$

The union then supplies the labor composite to individual good firms. The perfectly competitive labor union takes both the nominal wage W_t prevailing in the economy and the differentiated labor wage $W_t(h)$ charged by type h household as given, and choose quantities of labor varieties $N_t(h)$ to solve a maximization problem, which in turn yields the labor demand function for the labor variety h as defined above in household's maximization problem and an aggregate wage index $W_t = \left(\int_h W_t(h)^{1-\eta_L} dh \right)^{\frac{1}{1-\eta_L}}$. Thus, the labor demand for variety h is effectively determined by its relative wage and the elasticity of substitution η_L across varieties.

In each period t, a fraction $1 - \gamma_w$ of households is allowed to re-optimize their wages, whereas the remaining γ_w fraction of households are permitted to partially index their wages in accordance with past inflation, with the degree of indexation dictated by the parameter $\chi_w \in [0,1]$. In other words, if a household cannot adjust wage for a duration of τ periods, she would index her normalized wage at period τ to equal to $\prod_{s=1}^{\tau} \frac{\Pi_{t+s-1}^{\chi_w}}{\Pi_{t+s}} w_t(h)$, where $\Pi_t = \frac{P_t}{P_{t-1}}$ is the gross rate of inflation between periods $t-1$ and t. Note that here I re-write the problem in terms of real wage.[42] This implies that the household sets wages by solving the following maximization problem subject to a labor demand function:

$$\max_{w_t(h)} \mathbf{E}_t \sum_{\tau=0}^{\infty} (\beta\gamma_w)^{\tau} \left\{ -\theta_L \frac{N_t(h)^{1+\psi_L}}{1+\psi_L} + C_{t+\tau}^{-\sigma} \prod_{s=1}^{\tau} \frac{\Pi_{t+s-1}^{\chi_w}}{\Pi_{t+s}} w_t(h) N_{t+\tau}(h) \right\}$$

subject to

$$N_{t+\tau}(h) = \left(\prod_{s=1}^{\tau} \frac{\Pi_{t+s-1}^{\chi_w}}{\Pi_{t+s}} \frac{w_t(h)}{w_{t+\tau}} \right)^{-\eta_L} N_{t+\tau}.$$

When solving the maximization problem, all households would set the same wage in equilibrium because of the complete markets for state-contingent securities and the separable utility assumptions, I can therefore suppress the index h to analyze a symmetric equilibrium. The household's problem is largely standard, and the relevant first order conditions can be found in Appendix A.

Government policy. Finally, the central bank sets the nominal interest rate on a short-term government bond following a simple Taylor rule to target deviation of both inflation and output from their respective steady state level:

$$1 + i_t = (1 + i_{ss}) \left(\frac{\Pi_t}{\Pi_{ss}} \right)^{\phi^{\pi}} \left(\frac{Y_t}{Y_{ss}} \right)^{\phi^{y}} \epsilon_t^m, \phi^{\pi} > 0, \phi^{y} > 0 \tag{13}$$

where ϵ_t^m is a random shock to the policy rule that follows an exogenous AR(1) process with persistence $\rho^m \in (0,1)$. The short-term government bonds are in net zero supply, and the government budget is balanced every period.

Market clearing. In equilibrium, labor market clears such that the sum of production, the scalable production technology, and sales/advertising labor choices equals total labor supplied by the labor union:

$$N_t = \int_0^1 \left(\sum_i^{N_j} l_{p,ijt} + \sum_i^{N_j} l_{f,ijt} + \sum_i^{N_j} l_{s,ijt} \right) dj \tag{14}$$

[42] This is because, contingent on the choice of monetary policy rule, inflation could be non-stationary, rendering nominal wage non-stationary as well. Real wage, however, would remain stationary.

The final output Y_t can either be consumed by households, or used by firms to pay for the price adjustment costs. Market clearing for final good implies the following resource constraint:

$$Y_t = C_t + \int_0^1 \left(\sum_i^{N_j} \frac{\phi_p}{2} \left(\frac{P_{ijt}}{P_{ij,t-1}} - 1 \right)^2 \frac{P_{ijt}}{P_t} y_{ijt} \right) dj. \tag{15}$$

3.2 Equilibrium

I formally define the competitive equilibrium of the economy in Appendix A. Specifically, I focus on the normalized competitive equilibrium with stationary prices by normalizing all nominal quantities by the aggregate price level P_t. For instance, real wage is defined as $w_t = \frac{W_t}{P_t}$, and relative price of individual good ij is defined as $p_{ijt} = \frac{P_{ijt}}{P_t}$, and so forth.

Steady State. In order to solve the steady state equilibrium of the model, I focus on the zero-inflation steady state with fully symmetric sectors; i.e., all sectors have the same number of firms ($N_j = N \forall j$) and a symmetric distribution of firm-level productivity draws. Henceforth, for notation convenience, I replace individual variety index ij with firm productivity draw z to denote all firm-level variables. The formal definition of the steady state equilibrium is provided in Appendix A. In the Appendix, I also describe how I compute the steady state equilibrium in what amounts to a two-dimensional fixed point problem in real wage rate w and aggregate demand Y.

3.3 Calibration

In order to illustrate the dynamics for the baseline model, I calibrate the model with parameters provided in Table 2. A model period corresponds to a quarter, and the corresponding discount factor is $\beta = 0.99$. Focusing on the case of symmetric sectors, I now set the number of firms per industry to be $N = 3$ such that it corresponds to representative firms of low, medium, and high productivity, respectively. I draw firm-level productivity from a Pareto distribution with a shape parameter θ^s and a lower bound for productivity draws z_{min}, following the estimates provided in Amiti et al. (2019).

Parameters governing the degree of nominal rigidities and various elasticities of substitution are standard and largely follow previous estimates in the literature. Specifically, Rotemberg pricing friction ϕ_p is chosen to match a price duration of three quarters, whereas Calvo wage friction γ_w is set to allow a probability of being unable to re-optimize wages to be 0.75 in one year.[43] I choose a wage indexation parameter to equal to $\chi_w = 0.85$. Choices of elasticity of substitution are such that the within-secter elasticity of sub-

[43] The price duration of three quarters corresponds to a Calvo pricing friction of $\alpha = 0.66$; i.e., the probability of being unable to adjust price each quarter is 0.66. I choose the Rotemberg pricing friction such that it leads to a Phillips curve slope of $\frac{\rho-1}{\phi_p}$ that would be equal to the Phillips curve slope arisen from the Calvo setting, i.e., the Rotemberg slope = $\frac{(1-\alpha)(1-\alpha\beta)}{\alpha}$.

stitution ($\rho = 10$) is greater than that of between-sector ($\eta = 1.1$). This follows conventional calibration in the literature of nested-CES demand and Cournot competition as in Atkeson and Burstein (2008), and is consistent with the empirical findings in Broda and Weinstein (2006).

The parameter θ for the strength of the demand-shifter is set to equal to -0.8, according to Ravn et al. (2006), and the elasticity/curvature in the cost structure of adopting scalable production technology ψ_F is set to 2, following De Ridder (2019). I set the efficiency in adopting the technology to $\phi_F = 0.9$, and is uniform across firms such that technical efficiency z would remain the sole source of heterogeneity. I adopt the same cost structure for that of sales and advertising labor, in order to preserve the symmetry and account for the lack of distinction between these two types of intangible inputs in data.

Table 2: Parameters

	Value	Source/Target
Elas. intertemporal subs.	$1/\sigma = 0.5$	Standard value
Discount factor	$\beta = 0.99$	Standard value
Demand-shifter parameter	$\theta = -0.8$	Ravn et al. (2006)
Efficiency in adopting scalable tech.	$\phi_F = 0.9$	
Scalable tech. cost elasticity	$\psi_F = 2$	De Ridder (2019)
Efficiency in advertising	$\phi_B = 0.9$	
Advertising cost elasticity	$\psi_B = 2$	
Between-sector elas. of subs.	$\eta = 1.1$	Atkeson and Burstein (2008)
Within-sector elas. of subs.	$\rho = 10$	Atkeson and Burstein (2008)
Pareto shape of prod. distr.	$\theta^s = 8$	Amiti et al. (2019)
Log of lower bound of prod.	$z_{min} = \log(1)$	Amiti et al. (2019)
Rotemberg pricing friction	$\phi_p = 28$	Price duration = 3 quarters
Labor elasticity of sub.	$\eta_L = 12$	Standard value
Inverse of Frisch elas.	$\psi_L = 0.5$	
Slope of working disutility	$\theta_L = 0.5208$	
Calvo wage friction	$\gamma_w = 0.75$	Wage duration = 4 quarters
Wage indexation parameter	$\chi_w = 0.85$	
Taylor rule coefficients	$\phi^\pi = 0.9, \phi^y = 0.5$	Standard value
Persistence of monetary shock	$\rho^m = 0.75$	Standard value

Notes: Model period corresponds to a quarter.

Finally, I calibrate the remaining parameters that govern the household's optimization problem, including elasticity of intertemporal substitution ($\sigma = 2$) , labor elasticity of substitution ($\eta_L = 12$), and inverse of Frisch elasticity of labor supply ($\psi_L = 0.5$), following conventional calibration in the NK literature. The slope of working dis-utility of θ_L is determined within the model. The Taylor rule coefficients to target deviations in inflation and output are set to $\phi^\pi = 0.9$ and $\phi^y = 0.5$, respectively.

3.4 Theoretical Results

In this section, I discuss the main theoretical results obtained when the modelled economy starting from a zero-inflation steady state is subject to a one-time expansionary monetary policy shock. I solve for the full dynamic equilibrium in a model with symmetric sectors and three representative firms of low, medium, and high productivity in each sector, using standard techniques. I first show the aggregate-level responses of relevant macro variables; specifically, the model reproduces pro-cyclical aggregate markups, real wage, and inflation conditional to a monetary easing. I then examine the impulse responses of key firm-level variables to show that the model induces a re-allocation of production resources to the larger, more productive firms in a sector and that they are driving the pro-cyclical overall markup responses.

Aggregate pro-cyclical markups. Figure 3 shows the model-implied impulse response functions (IRFs) of key aggregate variables - nominal interest rate, output, employment, real wage, price inflation, and aggregate markup - to an expansionary monetary policy shock for four iterations of the model. Following a -50 basis point monetary policy innovation to the Taylor rule, nominal interest rate falls, and in the presence of pricing frictions, the real interest rate falls as well. Inter-temporal substitution implies that aggregate demand increases. Firms are unable to adjust their nominal prices sufficiently to monetary shocks, instead they adjust production and expenditures on the scalable production technology and advertising/marketing. Firms' increased labor demand puts upward pressures on wages. The presence of wage-setting frictions mutes the response of workers' real wage to the monetary easing, but nevertheless the response remains positive.

In a baseline representative-firm NK model with monopolistic competition, sticky wages, and sticky prices (yellow dotted line), an expansionary monetary shock produces counter-cyclical markups as expected. This is because when prices are sticky, and in the absence of strategic interaction that causes cyclical movement in desired markups, counter-cyclical movement in the sticky-price markups alone dictates the dynamics of markups such that markups fall in response of a monetary easing. A lowered firm profit margin, together with an increase in real wage both contribute to an increase in employment through (1) a negative wealth effect, as households are receiving less of their share of profit incomes and choose to work more, and (2) a substitution effect, as the increase in real wage has a direct impact on the number of hours worked. The increase in firm marginal cost in turn leads to a pro-cyclical inflation as firms pass-through cost pressure to consumers by increasing prices over time.

I then augment the baseline NK model with heterogeneous firms and oligopolistic competition only. To facilitate comparison, I use the same firm productivity draws across all three iterations of the heterogeneous-firm model. Incorporating oligopolistic interaction (red dashed line) alone further strengthens the counter-cyclicality of markups in the CES-oligopoly setting I study. The intuition is that due to the granularity of firms in an oligopoly, firms would act strategically and absorb cost shocks into markups so as to protect their market shares, and for a given degree of nominal rigidities, the greater this strategic complementary, the more cost shocks would be absorbed into markups and less would pass-through to price and inflation. The oligopoly model thus produces a more counter-cyclical markup response and a less pro-cyclical inflation response relative to the benchmark NK model.[44]

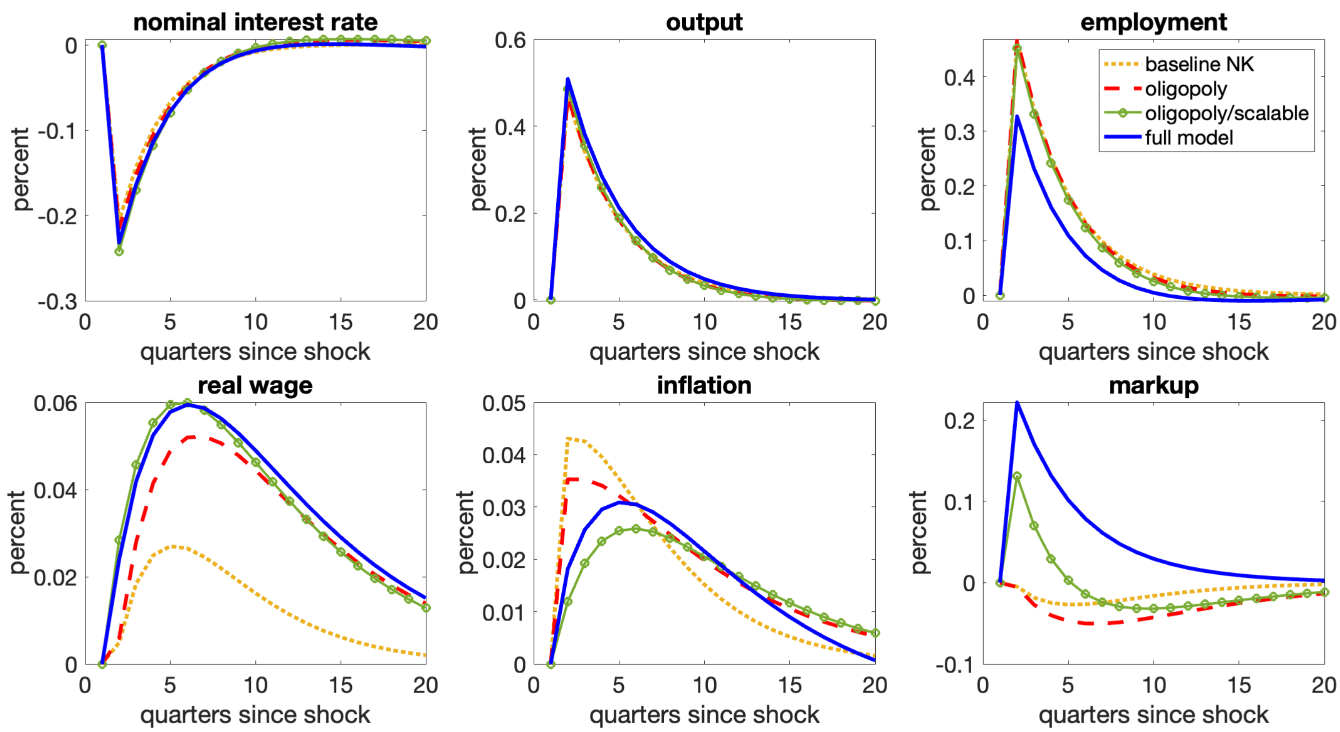


Figure 3: Model-implied aggregate IRFs to an expansionary monetary policy shock

Note: The figure plots the model-implied impulse responses of key aggregate variables to an expansionary monetary policy shock. The IRFs represent the percentage deviation of each variable from their respective steady state values. Four iterations of the model are displayed in the following order: (1) a baseline representative firm NK model with monopolistic competition as a benchmark; (2) A heterogeneous firm NK model with oligopolistic competition; (3) A heterogeneous firm NK model with oligopolistic competition and a scalable production technology, and (4) A heterogeneous firm NK model with oligopolistic competition, a scalable production technology, and sales/advertising/marketing as the full model.

I then introduce one type of intangible inputs, namely, the scalable production technology, into the CES-oligopoly model such that firms now operate a scalable technology and could endogenously change their economies of scale in response to shocks. Adding the scalable production technology to the oligopoly model (green line with circles) helps generating a mildly pro-cyclical aggregate markup response, as it breaks down the direct relationship between markup and real wage through differentiating average and marginal cost. As such, despite increasing real wages, firms are able to cut marginal cost by adopting more scalable production

[44]Due to the presence of heterogeneous firms, aggregate markup is computed as a harmonic mean of the firm-level markups weighted by firm market shares.

technologies (as it is more profitable to adopt them in expansions), and as a result the endogenous increase in economies of scale turns aggregate markup responses pro-cyclical.

Finally, I further augment the oligopoly/scalable model with the second type of intangible inputs such that firms could also engage in sales, marketing, and advertising to directly influence their idiosyncratic demands. In the full model with oligopoly and both types of intangible inputs (blue solid line), the aggregate markup is unequivocally pro-cyclical conditional on a monetary easing, and at the same time, the model reproduces pro-cyclical real wages and inflation.

Larger firms become larger. Figure 4 displays the theoretical IRFs of key firm-level variables - markups, sales, and adoption of the two intangible inputs - for the most productive (in the top row) and the least productive (in the bottom row) firm in a sector. In steady state, the more productive firm is also the larger firm in terms of market share and charges higher markup, and vice versa.[45] I omit the firm-level responses for the baseline NK model, as there is no heterogeneity involved.

In the oligopoly model, firms' dynamic responses show minimal heterogeneity. Both the larger, more productive firm and the smaller, less productive firm exhibit counter-cyclical markups and pro-cyclical sales following a monetary easing.[46] In the oligopoly model augmented with the scalable production technology only, the larger, more productive firms have higher steady state value for this new input, and would respond less to a monetary policy shock by virtue of the convex adoption cost structure. In fact, following a monetary easing, the smaller firms end up investing aggressively in the scalable production technology and would be able to produce more and charge a more pro-cyclical markup relative to the larger firms in the sector.

Finally, in the full model augmented with both intangible inputs such that firms could influence both types of market power, highly productive firms have higher steady state values for both intangible inputs, and would be in fact "larger" in terms of steady state market shares relative to their counterparts in the oligopoly; i.e., the modelled economy is more concentrated in the steady state.[47] Following an expansionary monetary shock, the larger, more productive firms respond more aggressively by increasing expenditures on both types of intangible inputs. Production resources are re-allocated towards them, and they are able to increase sales and exhibit pro-cyclical markups at the same time. This re-allocation effect is so strong that in this iteration of the model, a monetary easing appears to be contractionary for the smaller, less

[45] The notion that the more efficient firm would have higher market share and greater market power is a standard property in many macro models with heterogeneous firms and endogenous markups. It is supported empirically by studies such as Nakamura and Zerom (2010) and Berman, Martin and Mayer (2012).

[46] In theory, the larger, more productive firms should exhibit greater strategic complementarities, since their price is more sensitive to prices of their rivals when they have higher market share. As a result they would absorb more cost shocks into markups and pass-through less into prices. This leads to a reallocation of production resources towards larger firms such that their market share increases while the market shares of smaller firms fall, and they would exhibit more counter-cyclical markups compared to the smaller, less productive firms. This heterogeneity is not pronounced, especially when I compare the three cases that necessitates re-scaling, but is nevertheless present in a figure that plots only the firm-level IRFs of the CES-oligopoly model.

[47] The steady state implication is consistent with the finding in Bessen (2017), who shows that the adoption of proprietary IT systems is strongly correlated with industry concentration.

productive firms, who display conditionally counter-cyclical markups and sales as they cut their spending on both intangible inputs. At the aggregate, the harmonic mean of markups weighted according to market shares goes up following a monetary expansion despite the firm-level heterogeneous responses. This is because the larger firms' pro-cyclical responses are weighted more following the re-allocation of market shares towards them.

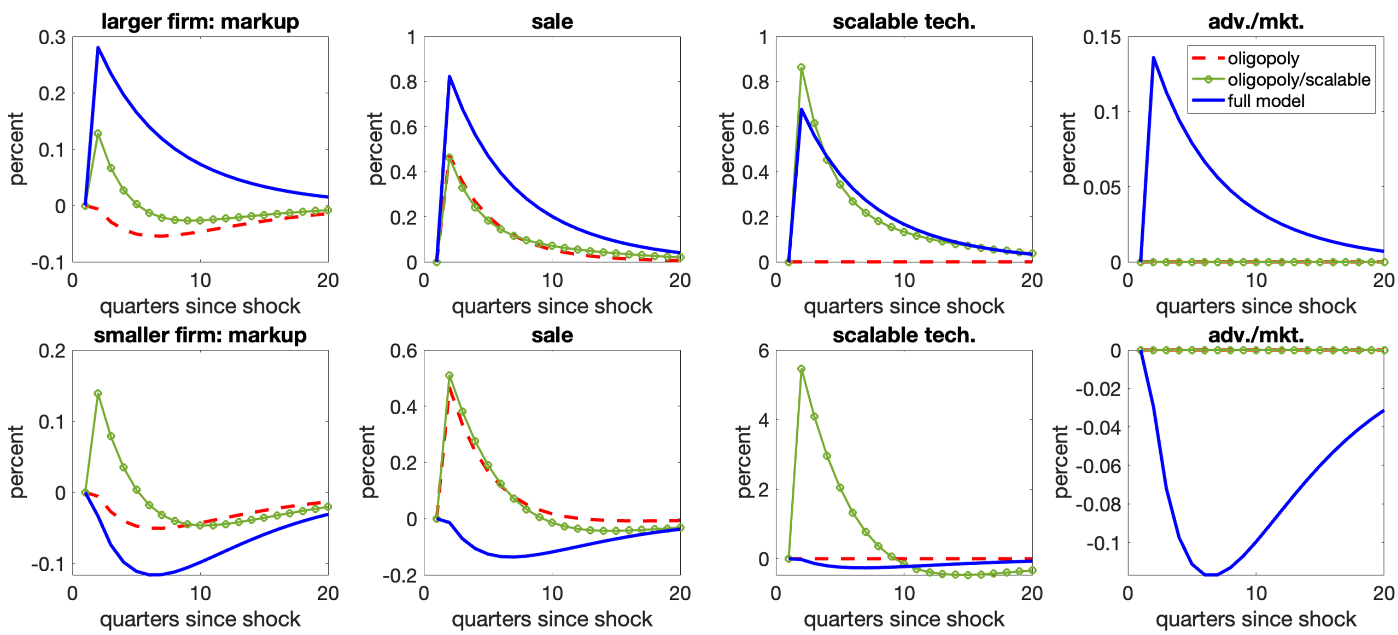


Figure 4: Model-implied firm-level IRFs to an expansionary monetary policy shock

Note: The figure plots the model-implied impulse responses of key firm-level variables to an expansionary monetary policy shock. The IRFs represent the percentage deviation of each variable from their respective steady state values. Three iterations of a heterogeneous firm NK model are displayed in the following order: (1) with oligopolistic competition; (2) with oligopolistic competition and a scalable production technology, and (3) with oligopolistic competition, a scalable production technology, and sales/advertising/marketing as the full model.

3.5 Inspecting the Mechanism

In order to illustrate the mechanism proposed, I approximate the dynamics of the economy by log-linearizing the normalized competitive equilibrium around a zero-inflation steady state. The formal definition of the log-linearized normalized competitive equilibrium is given in Appendix A. For illustrative purpose, I discuss the firm-level responses before turning to the the aggregate implications.

Endogenous pro-cyclical scale elasticities. An endogenous increase in the scale elasticities induced by the adoption of a scalable production technology leads to a pro-cyclical markup. The intuition is that in the presence of upfront adjustment costs, markup is function of both profit share of revenue s^{π} and scale elasticity υ such that:

$$\mu = \frac{P}{AC}\frac{AC}{MC} = \frac{1}{1-s^{\pi}}\upsilon, \text{ where } s^{\pi} = \frac{\text{Revenue} - \text{Total Cost}}{\text{Revenue}}, \tag{16}$$

and the scale elasticity of the function υ, as the ratio between average and marginal cost, is the inverse of the elasticity of production costs to quantity produced. Log-linearizing the equation yields:

$$\hat{\mu}_t = \hat{s}^{\pi}_t + \hat{\upsilon}_t = -\hat{w}_t, \text{ if } \hat{\upsilon}_t = 0 \text{ and } \hat{z}_t = 0 \tag{17}$$

Cyclical movement in markups reflects both movements in profit share $\hat{s}^{\pi}_t$ and scale elasticity $\hat{\upsilon}_t$. In most models with a constant returns to scale (CRS) production technology, the movement in scale elasticities $\hat{\upsilon}$ is fixed at zero and that the cyclicality of markups is pinned downed exclusively by movements of profit shares $\hat{s}^{\pi}$, which is inversely related to movements in real wage in the absence of a even more pro-cyclical labor productivity (i.e., $\hat{z}_t = 0$ or more precisely, the relationship holds as long as real wage is more pro-cyclical than labor productivity) and with labor as the only production input.[48]

This is the reason why in a standard NK model, it is impossible to generate both pro-cyclical markups (profit margins) and pro-cyclical real wages simultaneously, unless there are mechanisms in place to break-down this direct mapping.[49] Incorporating a scalable production technology into the production process naturally breaks down the relationship. Even though following a monetary easing, the dynamic response of real wage $\hat{w}$ is still positive, if $\hat{f}$ goes up sufficiently (i.e., firm increases its adoption of the scalable production technology sufficiently), real marginal cost would fall for some firms such that their markups would increase. This is evident from log-linearizing the real marginal cost equation:

$$\hat{mc}_t(z) = \hat{w}_t - \frac{f(z)}{1-f(z)}\hat{f}_t(z) - \hat{p}_t(z) = -\hat{\mu}_t(z) \tag{18}$$

The scalable property of intangible inputs thus helps reconciling higher real wages and higher markups.

[48] In a standard NK model with labor as the only input, the labor share of income is inversely related to markup:

$$\text{the labor share of income} = \frac{wL}{Y} = \frac{w}{Y/L} = \frac{\text{real wage}}{\text{labor productivity}} \simeq \frac{w}{z} = \frac{1}{\mu}$$

[49] Cantore et al. (2021) show that in a baseline NK model with labor as the sole input in production, the labor share, which equals to real marginal cost, would be conditionally pro-cyclical to monetary policy shocks, regardless of the relative strength of wage and price rigidities. It follows that markup would be conditionally counter-cyclical. They go on to show numerically that in a medium-scale NK model that incorporates various features (e.g., fixed production cost, working capital, CES production) to break down the direct mapping between marginal cost and the labor share, the structural model struggles to reproduce counter-cyclical labor share following a monetary policy shock without inducing a counterfactual counter-cyclical real wage response. Even though they focus on the cyclicality of the labor share, their results show that a pro-cyclical marginal cost (hence counter-cyclical markups) still quantitatively dominates the response of the labor share in models that have fixed production cost, CES production function, or in an open economy. It implies that unless there exists mechanism to induce a counter-cyclical marginal cost, conditionally counter-cyclical markups would likely persist in a medium-scale model.

Firms could pay more to each worker, and at the same time, see an increase in scale elasticities following a monetary easing when their marginal cost falls. The "lowered marginal cost" contributes to a rise in the price-cost markups.[50]

Importantly, a pro-cyclical markup does *not* necessarily imply higher price as it reflects a reduction in marginal cost, and hence all else equal, firms would face a lessened cost push pressure. Whether or not the rise in markup also reflects an "increase in market power" is related to the movement in a firm's relative price $\hat{p}_t(z)$. It turns out that the scalable production technology could interact with the second source of market power in the economy through a firm's ability to influence its demand, and this interaction breaks down the direct mapping between relative price and market share (market power), and at the core of the cross-sectional heterogeneity is the notion that larger firms can enjoy higher markup and less price-sensitive demand at the same time.

Synergies between the two intangibles inputs. The synergies between the two intangible inputs lead to a re-allocation of production resources towards the larger firms in a sector in a monetary policy-induced expansion. As shown in the firm-level impulse responses in Figure 4, when there is only one intangible input, namely, the scalable production technology, the model implies a re-allocation towards smaller firms as they respond by aggressively adopting the technology and see correspondingly a greater reduction in marginal cost. As a result, smaller firms display more pro-cyclical markups following a monetary easing. In Section 4, I show that this is not consistent with data. The presence of the second source of market power and the synergies between the ability to influence consumer demand and the adoption of a scalable production technology rationalize the cross-sectional heterogeneity.[51]

Intuitively, scalability embedded in the technology reduces marginal cost and allows firms to charge higher markups; meanwhile, higher markups incentivize firms to spend more on sales, advertising, and marketing, because the marginal profits of gaining a marginal consumer are increasing in the markup of the firm. This in turn shifts more demand towards the firm such that it produces more, incentivizing the firm to further reduce its marginal cost by increasing its adoption of the scalable production technology (with higher demand, the marginal benefit of deploying such a technology is higher). This positive feedback loop is the key mechanism underlying the "Matthew effect" - a capacity for larger firms to grow even larger and more profitable in a monetary easing.[52] Importantly, the reverse occurs for the smaller firms, who are facing the polar opposite in a negative feedback loop and are discouraged from expanding.

[50] Because the convex adjustment costs are paid regardless of sales, firms operating a scalable production technology would have an average profit margin below markup. Rising markups (marginal profitability) therefore do not necessarily reflect rising average profitability. It could reflect an increase in scale elasticities, a rising profit share of income, or a combination of both.

[51] In Section 4, I show that following a monetary easing, larger firms invest aggressively in intangible inputs and experience a stronger growth in sales, consistent with the mechanism of the model.

[52] The "Matthew effect" occurs if small exogenous differences are amplified by agents' endogenous responses to those small differences. It embodies the idea of "the well-endowed receiving further privilege", or more colloquially, "the rich getting richer".

The mechanism is discernible from log-linearizing the first order conditions with respect to the two intangible inputs and rearranging:

$$\hat{f}_t(z) = \underbrace{\frac{1-f(z)}{f(z)}\frac{\theta(1-\rho)}{1+\psi_F}}_{\oplus\text{ direct effect}}\hat{b}_t(z) - \underbrace{\frac{1-f(z)}{f(z)}\frac{\rho}{1+\psi_F}}_{\oplus\text{ indirect effect}}\underbrace{\left(\hat{p}_t(z)+\theta\hat{b}_t(z)-\hat{\tilde{p}}_t\right)}_{\Delta\text{relative price (market share)}} + \frac{1-f(z)}{f(z)(1+\psi_F)}\hat{Y}_t \tag{19}$$

$$\hat{b}_t(z) = \underbrace{\frac{\left(\frac{\mu(z)^{-1}}{1-\mu(z)^{-1}}+1\right)}{1+\frac{b(z)}{1-b(z)}(1+\psi_B)}}_{\oplus\text{ indirect effect}}\hat{\mu}_t(z) + \underbrace{\frac{f(z)}{1-f(z)}\left[(1+\psi_F) - \frac{1}{1+\frac{b(z)}{1-b(z)}(1+\psi_B)}\right]}_{\oplus\text{ direct effect}}\hat{f}_t(z) \tag{20}$$

where $\hat{mc}_t(z) = -\hat{\mu}_t(z)$ and $\theta < 0$. Notably, the weights on each contributing factor are firm-specific and are pinned down by the steady state values of the relevant firm-level variables together with structural parameters.

The two log-linearized equations stipulate a positive feedback loop for larger firms to adopt both intangible inputs following a monetary easing. An increase in aggregate demand leads to an increase in the scalable production technology adoption. In the presence of scalability, a sufficient increase in $\hat{f}_t(z)$ implies that marginal costs go down and markups go up. The direct effect of adopting the technology and the indirect effect through its impact on markups both contribute to an increase in advertising and marketing as shown in Equation 20. This increase in $\hat{b}_t(z)$ in turn shifts more demand towards the firm both directly and indirectly through its impact on firm market share via a lowered ideal price, leading to more production, and further incentives to adopt the scalable production technology as shown in Equation 19.

In contrast, in the absence of the second source of market power such that a firm cannot influence its demand through sales/advertising/marketing, the two log-linearized equations above reduce to one:

$$\hat{f}_t(z) = \underbrace{\frac{1-f(z)}{f(z)}\frac{1}{1+\psi_F}}_{\text{Higher for smaller firms}}\hat{Y}_t - \underbrace{\frac{1-f(z)}{f(z)}\frac{1}{1+\psi_F}\rho}_{\oplus\text{ indirect effect}}\underbrace{\hat{p}_t(z)}_{\Delta\text{relative price}}$$

Importantly, the weights on each component (aggregate demand and relative price) are firm-specific. A monetary easing leads to an increase in aggregate demand, this has a stronger direct effect on smaller firms with lower steady state level of scalable production technology $f(z)$. Scalability implies lower marginal cost and hence lower cost-driven pressure for smaller firms. This in turn leads to a lowered relative price and hence an indirect effect that reinforces the direct impact of an increase in demand. In contrast, larger firms increase relative prices and the indirect effect works against the direct effect of a rising demand. On net, the smaller firms end up adopting more scalable technologies and see a re-allocation of market shares towards them.

Dynamic responses of markups. Log-linearizing a firm's optimal pricing equation decomposes the dynamic response of markups into that of flexible-price and sticky-price components:

$$\hat{\mu}_t(z) = \underbrace{\frac{-1}{\Theta(z)-1}\hat{\Theta}_t(z)}_{\text{flexible-price markup}} \underbrace{-\frac{\phi_p}{\Theta(z)-1}\hat{\Pi}_t(z) + \frac{\beta\phi_p}{\Theta(z)-1}\mathbf{E}_t\hat{\Pi}_{t+1}(z),}_{\text{sticky-price markup}} \tag{21}$$

where $\hat{\Theta}_t(z) = \frac{\Theta(z)^{-1}-\rho^{-1}}{\Theta(z)^{-1}}(\rho-1) \underbrace{\left(\hat{p}_t(z)+\theta\hat{b}_t(z)-\hat{\tilde{p}}_t\right)}_{\Delta\text{relative adj. price (market share)}}$

Following an expansionary monetary shock, the larger, more productive firms increase expenditures on both types of intangible inputs aggressively, this leads to a pro-cyclical scale elasticity, and that marginal cost goes down despite increasing real wages (as long as the increase in the scalable production technology adoption f out-weights the increase in real wage). The relative sticker price p also goes up even though market share increases, and accordingly, the perceived demand elasticity Θ goes down as it is inversely related to market share. This is because the increase in demand shifter b out-weights the increase in the relative sticker price, leading to a decline in the relative ideal price $\tilde{p}$ (and by definition an increase in market share). A decline in the perceived demand elasticity implies a pro-cyclical flexible price markup that more than counteracts the counter-cyclical sticky-price markup. As a result, following a monetary easing, my calibration implies that the larger firms could have higher markups, higher market share, and higher relative sticker price at the same time.[53] Furthermore, the pro-cyclical markups now reflect both "lower marginal cost" and "increased market power".

Conversely, for the smaller, less productive firms, they cut spending on both intangible inputs, counter-cyclical scale elasticity and pro-cyclical real wages collectively imply that marginal cost goes up, and markup goes down. The relative sticker price goes down even though market share decreases. Again, the decrease in the demand shifter b out-weights the fall in the relative sticker price.

3.6 Aggregate Implications

The model has several implications at the aggregate level for allocative efficiency of the economy, as well as the inflation-output trade-off and the distributional consequences of monetary policy. In terms of efficiency, the re-allocation of market shares from low- to high-efficiency firms typically leads to efficiency gains in the economy as a whole (Baqaee, Farhi and Sangani, 2021; Meier and Reinelt, 2020). In a standard variable markup model, re-allocation occurs as larger firms' relative prices fall and thus market shares increase. In

[53] In a standard variable markup model, reallocation happens because larger firms have more elastic markups to absorb cost shocks such that following an increase in marginal cost, larger firm's relative price falls relative to smaller firms. In my model, re-allocation of market shares towards larger firms happens, despite their relative sticker price going up. This is because incorporating the demand-shifter breaks down the direct inverse mapping between relative price and firm market share, and now there is a distinction between the actual sticker price charged by the firm and the ideal price as perceived by customers.

my model, re-allocation of market shares towards larger firms happens in spite of increasing relative price.

At the aggregate level, incorporating intangible inputs leads to a similar output response as in the baseline NK benchmark, despite muting the response of the hours worked. This is because pro-cyclical markups lead to increased profits and a positive wealth effect that work against the substitution effect of an increased real wage. The muted labor response is almost entirely compensated for by the countervailing increase in allocative efficiency through reallocation of production resources towards the larger, more productive firms. However, this also implies that the labor share of income may fall during booms as profit margins widen.

Rearranging the log-linearized optimal pricing equation leads to a firm-level Phillips Curve:

$$\hat{\Pi}_t(z) = -\frac{1}{\phi_p}\hat{\Theta}_t(z) + \frac{\Theta(z)-1}{\phi_p}\hat{mc}_t(z) + \beta \mathbf{E}_t \hat{\Pi}_{t+1}(z) \tag{22}$$

Firm-level inflation is positively related to changes in real marginal cost and inflation expectation, and negatively related to change in perceived elasticity of demand. In other words, holding inflation expectation constant, firm-level inflation can be driven by rising cost and/or rising firm market power (characterized by a fall in the perceived demand elasticity). Despite an increasing real wage following interest rate cuts, larger firms in the model are facing lower cost-driven pressure, because their measured productivity is affected by the endogenous scalable production technology choices and is increasing faster than the real wage. The opposite happens to the smaller firms. As a result, the larger firm is experiencing greater profit-driven pressure that dominates the lowered cost-driven pressure, and ends up charging a higher relative sticker price. Conversely, the smaller firm faces lowered profit-driven pressure that dominates greater cost-driven pressure, and would charge a lower relative price.

Given the firm-level Phillips curve, the aggregate inflation is the weighted sum of firm-level inflation by its market share: $\hat{\Pi}_t = \sum_z s(z)\hat{\Pi}_t(z)$, where $s(z) = \left(\frac{\tilde{p}(z)}{\tilde{p}}\right)^{1-\rho}$ is the marker share of firm z. The aggregate New-Keynesian Phillips curve is:

$$\hat{\Pi}_t = -\sum_{z\in S}\frac{s(z)}{\phi_p}\hat{\Theta}_t(z) + \sum_{z\in S} s(z)\frac{\Theta(z)-1}{\phi_p}\hat{mc}_t(z) + \beta \mathbf{E}_t \hat{\Pi}_{t+1} \tag{23}$$

At the aggregate, inflation still responds pro-cyclically, albeit to a lesser extent compared to the NK benchmark. This is because for both larger and smaller firms, two counteracting forces of market power and marginal cost both affect inflation. The aggregate inflation is thus driven by the relative strength of these two forces weighted by firm market shares. Importantly, relative to the baseline NK case in which inflation is entirely driven by rising cost, holding inflation expectation constant, inflation in the full model occurs for a different reason - it is driven by both rising cost and rising market power. Curiously, the inflation response in the full model is muted relative to both the baseline NK and a model with oligopoly only; i.e., the full model features stronger monetary non-neutrality. This is again due to the presence of the counteracting forces in the full model that attenuates the inflation response.

Finally, my finding has implications for the redistribution consequence of monetary policy. The conventional view is that a monetary easing leads to a redistribution of income from firm profits to labor. The model implies the opposite that aggregate profit margins respond pro-cyclically, whereas the labor share of income responds counter-cyclically, consistent with the empirical findings in Cantore et al. (2021). In other words, a monetary easing leads to a redistribution of income from workers to firm profits.

4 Testing the Mechanism

In this section, I test the proposed mechanism by checking first if the data support some of the key testable predictions of the model, and second if the data support this mechanism rather than a competing explanation. Specifically, the model predicts that (1) larger firms in a sector that use intangible inputs would display more pro-cyclical markups than smaller firms, and that (2) at the firm level, intangible input expenditures and sales should co-move in response to a monetary policy shock. I first divide firms into four groups along two dimensions, namely larger (smaller) in market share and high (low) in intangible intensity; then I look at group-specific empirical impulse responses according to these characteristics. Second, I examine the dynamic responses of intangible inputs and sales based to the same grouping criteria to show that their movements are by and large aligned with the model's predictions. Finally, I assess the marginal contribution of market share/intangible intensity status conditioning on alternative firm-level characteristics to show that the cross-sectional heterogeneity is not driven by the alternative mechanism.

4.1 Testable Predictions

Larger firms respond more. The model predicts that larger firm in a sector that use intangible inputs would display more pro-cyclical markups. In order to empirically test the prediction, I redefine the groupings of firms based on firm characteristics $z_{i,t-1}$ along two dimensions such that the effects of monetary policy vary across the distributions of both firm market share and intangible intensity. Firms fall into four groups based on high and low market shares, and high and low intangible intensity. I define firm market share as the fraction of firm-level sale to the total sales at the 3-digit sectoral level.[54] Smaller (larger) refers to firms with market share below (above) the median of its distribution in the previous quarter; similarly, low (high) intensity corresponds to firms with intangible intensity below (above) the median of its distribution in the previous quarter. I then estimate impulse response functions in the second stage LP-IV panel regression in Equation 2.

[54] Taking a backward-looking (current and previous three quarters) moving averages to smooth the measure of market shares does not change my results. Nor does defining market shares at the 4-digit sectoral level.

Figure 5 plots the IRFs based on the refined grouping: the firms that are characterized by greater market shares and higher intangible intensity in cost structure are showing the most pro-cyclical markup response, which reaches a peak of 2.1% after one year, whereas the firms with smaller market share and lower intangible intensity display the least pro-cyclical markups that show no statistical significance at the 90% confidence level and turn counter-cyclical after two years. Meanwhile, those with high market share and low intangible intensity or low market share and high intangible intensity fall somewhere in between. For expository purpose, I report the impulse responses for the most and the least responsive groups in the main text and relegate the full set of impulse responses to Appendix E.2.3.

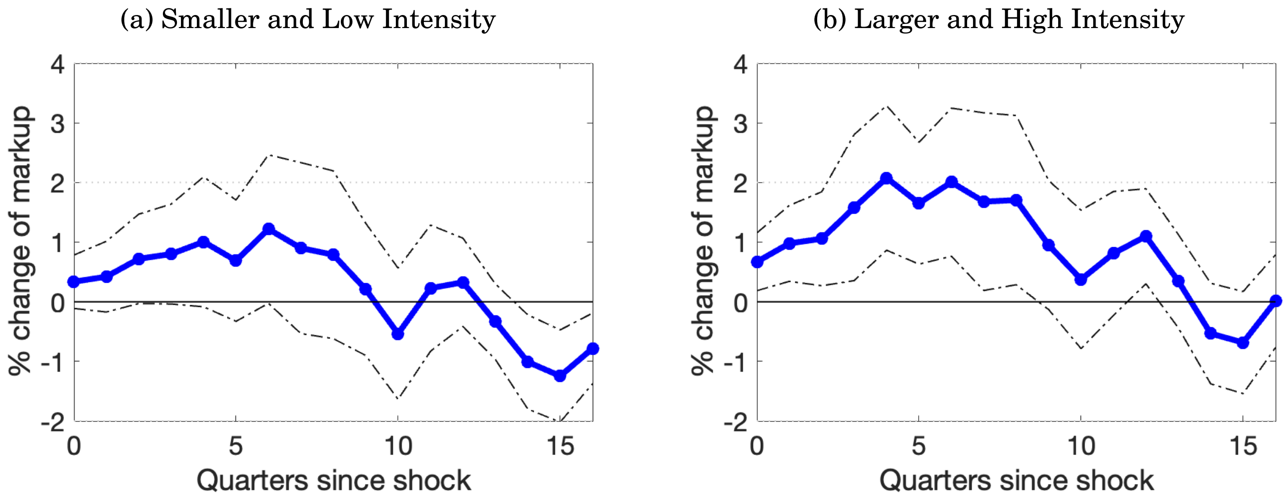


Figure 5: Heterogeneous Effects of Monetary Policy Shocks on Markups

Note: The figure shows the impulse responses of the baseline measure of markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The sampling period is 1990Q1-2009Q4, and the forecast horizon is 16 quarters. The interest rate is interacted with binary indicators based on distributions of firm market share and intangible intensity, in which smaller (larger) refers to firms with market share below (above) the median of its distribution in the previous quarter, and low (high) intensity corresponds to firms with intangible intensity below (above) the median of its distribution in the previous quarter. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

An alternative way to examine the cross-sectional heterogeneity is to assess the relative difference of group-specific markup response with respect to the least responsive base group. This method allows for a formal assessment of the statistical differences between the impulse responses across groups at each horizon. Formally, I re-run the second stage baseline regression with the Fed Funds rate (instrumented by the GSS shocks) added as an additional regressor and designate the smaller and low intangible intensity firms as the base group. In Appendix E.2.3, I show that the markup responses of larger and high intangible intensity firms are significantly greater than those of the base group at almost all forecast horizons.

In Appendix E.2.3, I also use alternative measures of markups and the narratively-identified Romer and

Romer (2004) shocks as in the robustness section of Section 2, to show that the cross-sectional results based on both market share and intangible intensity are robust.

Co-movement of intangible inputs and sales. The mechanism relies on the co-movement of intangible input expenditures and sales at the firm-level following interest rate changes. In particular, the model predicts that the dynamics of intangible inputs are closely linked to that of sales, and that they co-move after a monetary policy shock. The natural test of this prediction is to show that firm-level intangible input expenditures as proxied by firms' spending on SG&A net of R&D are empirically consistent with the conditional movements in sales according to the same distribution of firm characteristics. To do so, I use the same empirical specification as in Equation 2, but change the dependent variables to log of real SG&A expenditures and log of real sales, respectively.

Keeping the same grouping of firms as before, I allow the effects of monetary policy to vary across the distributions of both market share and intangible intensity. Figure 6 reports the IRFs for the two baseline groups of firms. Intangible expenditures go up significantly for larger firms that operate intangible-intensive technologies, and reach a peak of approximately 6% 12 quarters after the one percentage cut in the Fed Fund rates. In contrast, smaller firms with low intangible intensity cut their spending on SG&A significantly initially after the shock and reach a trough of about -2% three quarters later. The cumulative response turns positive further down the horizon but is nevertheless not statistically significant.

(a) Smaller and Low Intensity (b) Larger and High Intensity

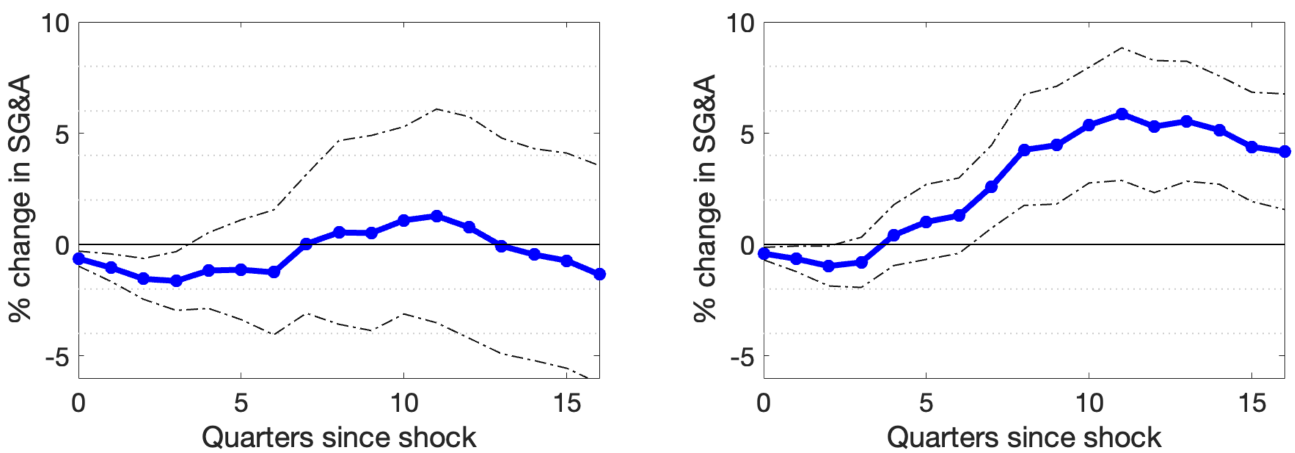


Figure 6: Heterogeneous Effects of Monetary Policy Shocks on SG&A

Note: The figure shows the impulse responses of log real SG&A in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The sampling period is 1990Q1-2009Q4, and the forecast horizon is 16 quarters. The interest rate is interacted with binary indicators based on distributions of firm market share and intangible intensity, in which smaller (larger) refers to firms with market share below (above) the median of its distribution in the previous quarter, and low (high) intensity corresponds to firms with intangible intensity below (above) the 50th percentile of its distribution in the previous quarter. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

In comparison to the IRFs of markups, one interesting feature of the IRFs of SG&A is that they lag behind the pro-cyclical markup responses. The timing seems to suggest that the pro-cyclical profit margins may contribute to the pro-cyclical SG&A expenses, providing supportive evidence for one aspect of the positive feedback loop laid down in Section 3; namely, higher markups incentivize firms to increase spending on intangible inputs. The initial contractionary response is a puzzling yet robust feature of the micro data that is also present in the average response. The lumpiness and volatility of the data partially account for the initial contractionary effect.[55] To the extent that the initial contractionary response is not a peculiar feature of the data, the timing of response in SG&A and the fact that pro-cyclical markups are found across all firms suggest the possible presence of other mechanism. I come back to this point in Section 5 when I discuss alternative mechanisms.

Turning to the impulse responses of sales, in Figure 7, I show that a similar picture emerges as that of SG&A: sales go up significantly for high market share firms with high intangible intensity, and reach a peak of 6% 12 quarters after the initial cut of the interest rate. In contrast, low market share firms with low intangible intensity see a significant reduction in sales initially after the shock, even though the cumulative response turns positive further down the horizon and is statistically significant.

(a) Smaller and Low Intensity (b) Larger and High Intensity

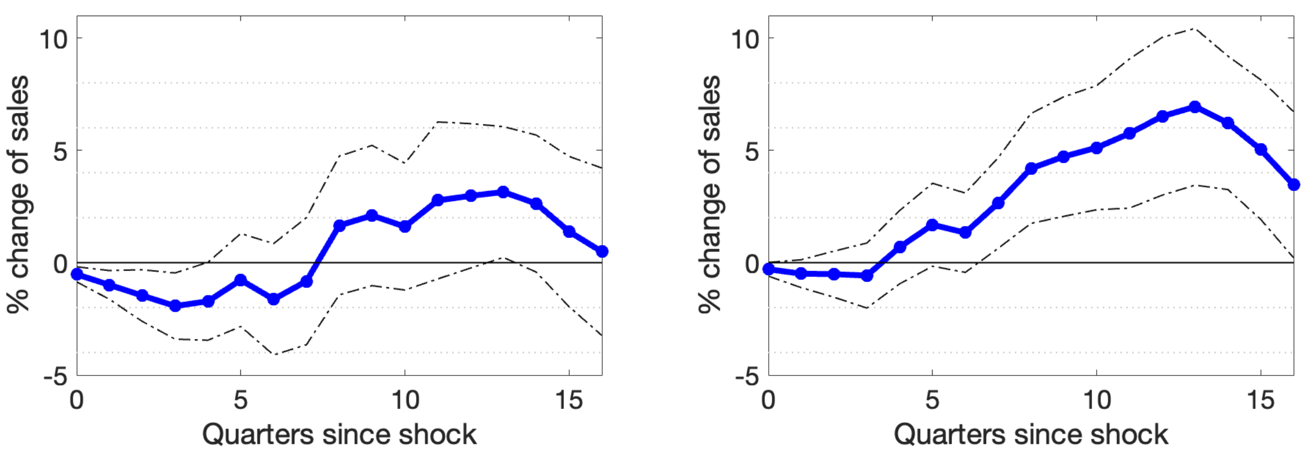


Figure 7: Heterogeneous Effects of Monetary Policy Shocks on Sales

Note: The figure shows the impulse responses of log real sales in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The sampling period is 1990Q1-2009Q4, and the forecast horizon is 16 quarters. The interest rate is interacted with binary indicators based on distributions of firm market share and intangible intensity, in which smaller (larger) refers to firms with market share below (above) the median of its distribution in the previous quarter, and low (high) intensity corresponds to firms with intangible intensity below (above) the 50th percentile of its distribution in the previous quarter. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

[55] I smooth the data by taking a moving average of the present period and previous three quarters. In order to reduce volatility, I trim the dependent variable based on growth rates (dropping those with growth rates above and below 5% by period).

The initial contractionary response from the smaller/low intangible intensity firms is puzzling, and at first glance stands in sharp contrast with the empirical finding in Gertler and Gilchrist (1994) that "small" firms' sales respond more to a monetary policy shock.[56] In addition to the different sampling period and samples used, we focus on different features of the data. My focus on market share (a relative measure) centers on the re-allocation happening within each sector, whereas theirs is about the absolute measures of large versus small, based on distribution of total assets or sales in a period. As a robustness check, I adopt a standard measure of firm size (total assets) as the conditioning variable and find that sales of small firms (those with assets below 50th percentile each period) respond slightly more to a monetary policy shock, but the heterogeneity is not pronounced. The bottom line is: based on the absolute measure of firm size, sales of large firms do not respond more to a monetary easing, unlike high market share/high intangible firms.

Similar to SG&A, the initial contractionary response is a robust feature of the micro data that remains after smoothing and trimming based on growth rates. It seems that this puzzling feature is more pronounced in the group of firms that have low market share and low intangible intensity to the point that the point estimates are statistically significant in the first year after the shock. The cumulative response does turn positive after two years, but the contractionary effect lasts much longer than the contrasting group. My interpretation of the finding is that there exist evidence for a contractionary response for sales of this particular type of firms.

4.2 Alternative Firm-level Characteristics

The evidence so far suggests that high market share/high intangible intensity is a strong predictor of a more pro-cyclical markup response to a monetary policy shock. Nevertheless, there are alternative firm characteristics that correlate with firm market share/intangible intensity and could potentially induce a pro-cyclical markup response. In particular, throughout the analysis, I abstract from the notion that firms' pro-cyclical innovation effort to boost productive efficiency could also affect the dynamics of markups. In the empirical section, R&D is removed from my definition of intangible intensity, whereas in the theoretical section, firm-level productivity is exogenous and fixed over time. I look at R&D intensity as an alternative conditioning variable to show that the baseline characteristics remain a strong predictor.

Conditioning on R&D intensity. In order to assess the the marginal impact of market share/intangible intensity conditional on R&D intensity, I further interact the four baseline groups with the distribution of

[56] Specifically, in Gertler and Gilchrist (1994), the sales of small firms (those with sales below 30th percentile each period) respond more to monetary policy shocks. It should be noted that the sample and time period we study differ significantly: I focus on publicly-listed firms from a wide variety of sectors from Compustat for 1990Q1-2009Q4, whereas they study the universe of U.S. manufacturing firms from Quarterly Financial Report (QFR) of the Census Bureau for 1958-1994. Compustat data is known to biased towards large firms, and it is likely that the smaller firms in my sample would fall into the large category in their sample. There could also exists considerable sectoral heterogeneity, and that manufacturing firms' responses are marked by responses of non-manufacturing firms.

the third firm characteristic in a triple-sorting of the data. My identification strategy exploits an exogenous monetary shock common for all firms in order to identify heterogeneous responses among sub-groups of firms. The nature of the "experiment" necessitates the interaction of the additional characteristic with the common shock, instead of being simply added as an additional regressor. I adopt this triple-sorting strategy to condition on the additional characteristic in a semi-parametric manner.[57]

Because firm-level R&D is sparsely populated, I instead focus on sectoral means in order to maximize the number of observations. The idea is to leverage the considerable sectoral heterogeneity in research and innovation efforts to test if the cross-sectional heterogeneity is instead driven by high R&D firms. Some sectors (for instance, pharmaceutical and high-tech) are intrinsically better suited to engage in own innovation than others (for instance, mining and construction), and therefore the goal is to test if market share/intangible intensity remains a strong predictor of the cross-sectional heterogeneity in the high R&D intensive sectors. I re-run the regressions in Equation 2 in two sub-samples based on sectoral mean R&D intensity, in which I sort firms into the same four groups as before.[58]

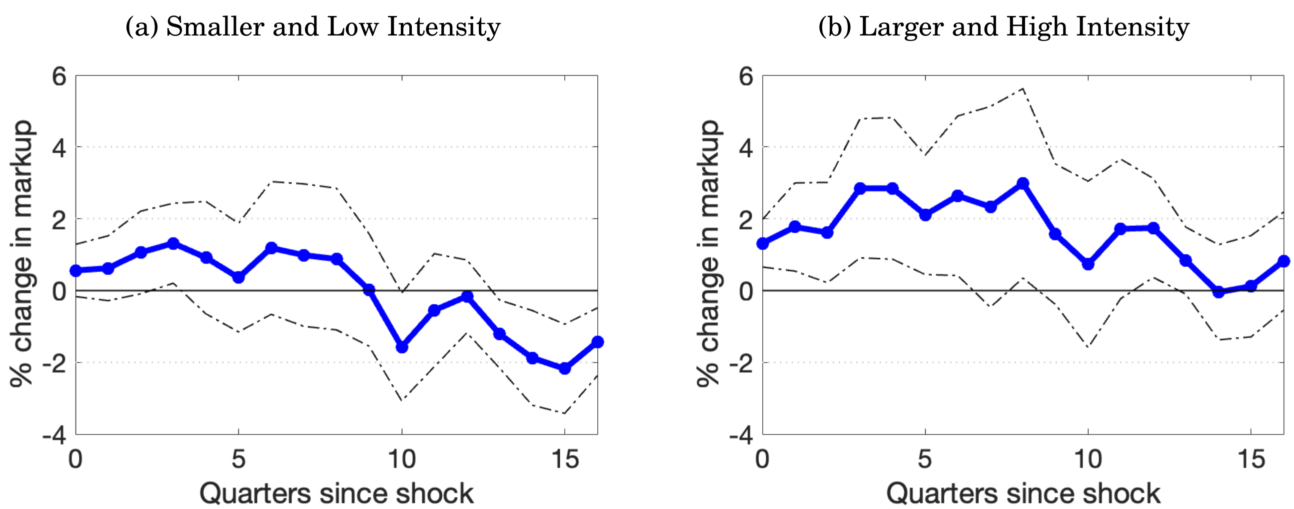


Figure 8: Heterogeneous Effects of Monetary Policy Shocks on Markups in High R&D Intensity Sectors

Note: The figure shows the impulse responses of the baseline markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate in a sub-sample of high R&D intensity firms. The sampling period is 1990Q1-2009Q4, and the forecast horizon is 16 quarters. The interest rate is interacted with binary indicators based on distributions of firm market share and intangible intensity, in which smaller (larger) refers to firms with market share below (above) the median of its distribution in the previous quarter, and low (high) intensity corresponds to firms with intangible intensity below (above) the 50th percentile of its distribution in the previous quarter. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

[57]The triple-sorting strategy is highly demanding of the data, to maximize numbers of observations per group, I use medians as cutoffs for each distribution of firm characteristics.

[58]R&D intensity is computed as the ratio between firm-level R&D soundings to sales. The cut-off is fixed at the median of the distribution of sectoral means. For alternative firm-level charateristics that are well-populated (e.g., leverage ratio), I apply the tripe-sorting stratefy at the firm-level, in which firms are sorted into eight sub-groups based on three characteristics.

Figure 8 reports the IRFs for the two baseline groups, conditioning on high sectoral R&D intensity. Market share and intangible intensity remain robust in predicting the cross-sectional heterogeneity. In fact, the heterogeneity appears to be more pronounced in sectors with high R&D intensity. The most responsive sub-group of high market share/high intangible intensity/high sectoral R&D intensity firms displays a peak effect that is slightly greater than 3% in two years after the shock. R&D is considered the third type of business intangible alongside computerized information and brand value by Corrado et al. (2005), and it is plausible that a pro-cyclical R&D and the synergies among the three types of intangibles also play a role in driving the dynamic responses of markups, in particular, the pro-cyclical responses further down the horizon.

In summary, I find empirical support for the mechanism proposed by testing the model's key predictions, and the empirical evidence is robust to controlling for R&D intensity.

5 Discussion: Alternative Mechanisms

In this section, I discuss the roles of financial frictions, exporter status, and search frictions as either an alternative mechanism that could induce pro-cyclical markups following a monetary stimulus or constitutes a channel that could interact with the proposed mechanism to affect the cross-sectional heterogeneity of markups. I revisit R&D as the third type of intangible input. My findings that markups are consistently pro-cyclical to monetary shocks across firms suggest that there may be alternative mechanisms that are not captured in the model. Nevertheless, I argue that it is difficult to explain my empirical findings, particularly those in the cross-section, from the viewpoint of any of these channels alone. I do not find empirical support for the roles of financial frictions and exporter status; the dynamics of the responses suggest that R&D intensity might be driving the average response over a longer horizon, but I find it implausible to explain the short-run effects.

Financial frictions. Financial friction *per se* does not lead to pro-cyclical markups, nevertheless it could interact with the credit channel of monetary policy to induce pro-cyclical markups through the assumption of working capital. Firms need to borrow in advance to finance labor costs. An interest rate cut reduces the cost of borrowing, and if this reduction out-weights the increase in real wage, firms would face lower marginal costs and have pro-cyclical markups (Ravenna and Walsh, 2006; Surico, 2008). The credit channel of monetary policy implies that firms that are financially constrained, would have more pro-cyclical markups conditional on a monetary easing. In Appendix E.2.3, I show that market share and intangible intensity remain strong predictors of markup dynamics among firms that are likely to be financially constrained (i.e., those small in size as measured by total asset, high in leverage, and low in liquidity). In fact, these additional empirical results collectively suggest that once market share and intangible intensity are taken into account, there is little difference in markup responses across firms based on measures of financial frictions, implying that financial frictions are unlikely to be driving the cross-sectional heterogeneity of markups through the credit channel of monetary policy.

Could financial frictions affect the dynamics of markups through intangible adoption decisions? Throughout the theoretical analysis, I abstract from firm-level financial frictions, and assume all firms, regardless of their levels of productivity, are not financially constrained when making decisions. This is unlikely to be true in the real world, and financial frictions may very well influence firms' intangible input adoption decisions, and hence the dynamics of markups.[59] Morlacco and Zeke (2020) show that a monetary easing leads to greater responses of SG&A for large firms.[60] To the extent that the monetary easing is alleviating financial constraints to a greater extent for the smaller firms, it is unlikely that financial friction is interacting with intangible input adoption to generate relatively more pro-cyclical markups for larger firms.[61]

R&D. Firms could engage in process innovation through Research & Development (R&D) to improve technical efficiency *per se*, and a pro-cyclical R&D investment potentially provides an alternative mechanism that could explain the pro-cyclical responses of markups. Intuitively, if a firm's technical productivity is endogenous and is more pro-cyclical than real wages, then the firm is effectively facing a counter-cyclical marginal cost after a monetary easing. Therefore R&D, either conducted in-house or purchased directly through acquisitions, is yet another important candidate that may influence both the firm's marginal cost (hence pricing decisions and markups) and likely product quality (which may behave isomorphically to a demand shifter, albeit now consumers can indeed benefit from an actual improvement of product quality). Larger firms could plausibly invest more aggressively in R&D following a monetary easing, resulting in pro-cyclical technical efficiency and/or product quality, and are thus able to charge higher markups.[62]

The reason I think pro-cyclical R&D has little bearing on my cross-sectional results is that empirically, as shown in Section 4, in high R&D intensity sectors, the heterogeneous responses of markups by market share and intangible intensity remain robust. Furthermore, the effects of R&D typically manifest with a lag. The finding that markups respond immediately to monetary policy shocks in a pro-cyclical manner implies the existence of alternative mechanism that influences current period profits, and arguably has little to do with firms' research and innovation. Nevertheless, as shown in Appendix E.2.3, firms in high R&D sectors

[59] Altomonte et al. (2021) show that financial constraints distort firms' decision to adopt scalable production technology, which in turn lead to heterogeneous markups and pass-through elasticities in a static partial equilibrium model.

[60] They show that the heterogeneity remains robust after controlling for proxies of financial frictions. Their measure of size (by sales relative to industry mean) is closer to my notion of relative firm size, than the standard, absolute measure.

[61] For financial frictions to drive the cross-sectional dynamic responses through more aggressive adoption of intangible inputs, larger firms should be experiencing more binding financial constraints. There is also the argument that the intangible-intensive firms are less affected by the traditional credit constraints, since intangibles suffer from greater information asymmetry and alack of collateralizability (Hall and Lerner, 2010; Lim, Macias and Moeller, 2020). Therefore, intangible-intensive firms often resort to retained earning or equity issuance, in place of the traditional debt issuance, to fund intangible investment.

[62] Recently, in a dynamic patent race model, Liu, Mian and Sufi (2018) show that a decline in the long-term interest rate induces a stronger R&D investment response from market leaders, resulting in higher profits and increased industry concentration.

have, by and large, more pro-cyclical markups than their counterparts in low R&D sectors, and the most responsive sub-group in the high R&D sector shows a higher peak further down the forecast horizon.[63] I interpret the finding as suggestive that pro-cyclical R&D could be driving the pro-cyclical average response of markup over a longer horizon.

Exporter in an open economy. Another consideration concerns the role of exporters in an open economy. A large literature links firm-level markup to exporter status and finds that exporters tend to have higher markups on average. Moreover, new exporters often see markup increases upon entry into export markets (Bernard, Eaton, Jensen and Kortum, 2003; Melitz and Ottaviano, 2008; De Loecker and Warzynski, 2012). Following a monetary easing, exporters may benefit from a weakened currency and are able to increase markups relative to the non-exporters. In Appendix E.2.3, I show that market share/intangible intensity remains a robust predictor of cross-sectional heterogeneity even among multinational firms in Compustat, and that exporter status is unlikely to be driving my results.[64]

Labor market frictions. Another route to disentangle the inverse relationship between the real wage and markup is to relax the assumption of a competitive labor market and instead allow search and matching frictions such that the real wages are contingent upon workers' bargaining power (Galí, 2010; Christiano, Eichenbaum and Trabandt, 2016). In particular, when coupled with search frictions in product market, Kaplan and Menzio (2016) show that in such a framework, workers' counter-cyclical shopping intensity for lower-priced goods could result in pro-cyclical markups for firms.

While the mechanism could induce pro-cyclical responses in average markups, it is not straightforward to argue that it would bring about systematic heterogeneous responses at the firm level in a given sector, especially after a common exogenous shock (monetary policy shock) across all consumers and producers alike.[65] Fundamentally, this line of argument centers on the role of consumer shopping behaviors and largely treats producers as passively facing a less elastic demand curve in booms. I view my argument as complementary, in which I emphasize the producer's potential role in actively influencing the elasticity of its own demand curve. In reality, the truth probably falls somewhere in between, and the cyclical properties of price elasticity (and hence markup) are jointly shaped by the interaction between consumers and producers. That said, this type of interaction is outside the scope of this paper.

[63] The baseline sub-group of firms with high market share/high intangible intensity shows a peak effect of 2.1% in a year after the shock, whereas the same group in high R&D sector shows a peak effect of above 3% in two years.

[64] In order to maximize number of observations, I designate a firm is as a multinational firm if it ever reports pre-tax foreign income (Compustat item *pifo*) at any point during the sampling period.

[65] Consumers could respond differently to the monetary stimulus and change their shopping behaviors accordingly, such that producers, in particular local producers, would experience heterogeneous shopping intensity and hence exhibit heterogeneous markup responses. The channel, while possible, is outside the scope of the paper.

6 Conclusion

How do firms' price-cost markups respond to monetary policy shocks? Empirically, they are conditionally pro-cyclical, and firms that use intangible inputs intensively have more pro-cyclical markups. I show that the properties of scalability and synergies associated with intangible inputs can help rationalizing both findings in a heterogeneous firms NK model with sticky prices.

Intuitively, a firm's market power is ultimately determined by the perceived consequences of increasing prices: first, how would its customers react? And second, how would its competitors in the same sector react? Through the adoption of intangible inputs, larger firms are able to have customers tolerate higher prices with less elastic demand, and find their competitors discouraged to expand, as such, the increase in market power from influencing consumer demand reinforces the reduction in marginal cost from adopting the scalable technology and leads to more a pro-cyclical markup for larger firms.

Firms' delicate balancing act of protecting market share and reaping greater business profits are shaped by the cyclical movements of firms' pricing power, which is in turn influenced by the cyclical aggregate demand pressure. Even though my analysis focuses on markup dynamics following a monetary policy shock, its implications are relevant for understanding business cycles in general. In booms, firms are more likely to have their customers accepting higher prices. Thus it is plausible that the feedback loops identified here is operative when there is aggregate demand pressure in general.

My findings have several implications for the inflation-output trade-off and the distributional consequences of monetary policy. On the one hand, my results show that there could be unintended consequences of increasing larger firms' market power through their more aggressive adoption of a cost-reducing technology and engagement in marketing campaigns to capitalize on the rising demand in the wake of a monetary easing. This in turn leads to a redistribution of income from workers to firm profits as aggregate profit margins respond pro-cyclically to interest rate cuts. On the other hand, this re-allocation towards larger firms could be efficiency-enhancing, as larger firms are often more efficient. Therefore I do not intent to argue for the notion of resisting bigness *per se*, nor for the use of anti-trust policy to manage inflation.[66] Instead, I offer a more nuanced view and my results call for more attention to the modelling of the cyclical properties of firms' profit margins, consumers' price sensitivity, and their interaction over the business cycles.

Another implication of my findings concerns the role of business intangibles and their aggregate implications. The notion that that intangible inputs are both a good and a bad source of concentration is not new. My results highlight that the source of firm market power matters for the implication of intangible input adoption, especially given that firms need some market power in equilibrium to deploy such a technology in

[66]The idea is at the core of the popular narrative that large, greedy corporations are responsible for inflation, the very myth that mainstream economists seek to debunk, as summarized by a tweet from Lawrence Summers in December, 2021: "Resisting bigness per se, even when it comes from efficiency or seeking to protect competitors from efficient rivals, is a prescription for higher not lower prices."

the first place. The focal point is not about larger firms having market power, is rather about if such market power being utilized to the detriments of consumer welfare. My findings suggest that if a firm's ability to influence consumer demand constitutes a significant source of its market power, then the synergies between intangible inputs could negate the benefits of a cost reduction associated with its scalability, and impair customer welfare with increased prices.[67] An interesting direction for future research would be to quantify the sources of corporate market power and clarify the channels through which they potentially interact, and assess the welfare consequences.

References

Afrouzi, Hassan, Andres Drenik, and Ryan Kim. 2020. "Growing by the Masses: Revisiting the Link between Firm Size and Market Power." *Available at SSRN 3703244*.

Akcigit, Ufuk, and Sina T Ates. 2019. "Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory." *National Bureau of Economic Research, Working paper No. 25756*.

Akcigit, Ufuk, and William R Kerr. 2018. "Growth through heterogeneous innovations." *Journal of Political Economy*, 126(4): 1374–1443.

Altomonte, Carlo, Domenico Favoino, Monica Morlacco, and Tommaso Sonno. 2021. "Markups, intangible capital and heterogeneous financial frictions."

Amiti, Mary, Oleg Itskhoki, and Jozef Konings. 2019. "International shocks, variable markups, and domestic prices." *The Review of Economic Studies*, 86(6): 2356–2402.

Anderson, Eric, Sergio Rebelo, and Arlene Wong. 2018. "Markups across space and time." National Bureau of Economic Research.

Arkolakis, Costas. 2010. "Market penetration costs and the new consumers margin in international trade." *Journal of political economy*, 118(6): 1151–1199.

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

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. 2020. "The fall of the labor share and the rise of superstar firms." *The Quarterly Journal of Economics*, 135(2): 645–709.

Baqaee, David, Emmanuel Farhi, and Kunal Sangani. 2021. "The Supply-Side Effects of Monetary Policy." National Bureau of Economic Research Working Paper 28345.

[67]Note that I abstract from the notion that firm's idiosyncratic demand could also arise from quality differences, which constitute another source of market power, in addition to productive efficiency and the ability to shift consumer demand discussed here.

Basu, Susanto. 1995. “Intermediate Goods and Business Cycles: Implications for Productivity and Welfare.” *The American Economic Review*, 85(3): 512.

Benigno, Pierpaolo, and Ester Faia. 2010. “Globalization, pass-through and inflation dynamic.” National Bureau of Economic Research.

Berman, Nicolas, Philippe Martin, and Thierry Mayer. 2012. “How do different exporters react to exchange rate changes?” *The Quarterly Journal of Economics*, 127(1): 437–492.

Bernard, Andrew B, Jonathan Eaton, J Bradford Jensen, and Samuel Kortum. 2003. “Plants and productivity in international trade.” *American economic review*, 93(4): 1268–1290.

Bessen, James. 2017. “Information technology and industry concentration.”

Bhandari, Anmol, and Ellen R McGrattan. 2021. “Sweat equity in US private business.” *The Quarterly Journal of Economics*, 136(2): 727–781.

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

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

Blanchard, Olivier, and Jordi Galí. 2007. “Real wage rigidities and the New Keynesian model.” *Journal of Money, Credit and Banking*, 39(s1): 35–65.

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

Broda, Christian, and David E Weinstein. 2006. “Globalization and the Gains from Variety.” *The Quarterly journal of economics*, 121(2): 541–585.

Broer, Tobias, Niels-Jakob Harbo Hansen, Per Krusell, and Erik Öberg. 2020. “The New Keynesian transmission mechanism: A heterogeneous-agent perspective.” *The Review of Economic Studies*, 87(1): 77–101.

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

Calvo, Guillermo A. 1983. “Staggered prices in a utility-maximizing framework.” *Journal of Monetary Economics*, 12(3): 383–398.

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

Cavenaile, Laurent, Murat Alp Celik, Pau Roldan-Blanco, and Xu Tian. 2021. “Style Over Substance? Advertising, Innovation, and Endogenous Market Structure.” *Advertising, Innovation, and Endogenous Market Structure (March 13, 2021)*.

Christiano, Lawrence J, Martin S Eichenbaum, and Mathias Trabandt. 2016. "Unemployment and business cycles." *Econometrica*, 84(4): 1523–1569.

Christiano, L.J., M. Eichenbaum, and C.L. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy*, 113(1).

Cloyne, James, Clodomiro Ferreira, Maren Froemel, and Paolo Surico. 2019. "Monetary Policy, Corporate Finance and Investment."

Corhay, Alexandre, Howard Kung, and Lukas Schmid. 2020. "Q: risk, rents, or growth?" *USC Marshall School of Business Research Paper Sponsored by iORB, No. Forthcoming*.

Corrado, Carol, Charles Hulten, and Daniel Sichel. 2005. "Measuring capital and technology: an expanded framework." In *Measuring capital in the new economy*. 11–46. University of Chicago Press.

Corrado, Carol, Charles Hulten, and Daniel Sichel. 2009. "Intangible capital and US economic growth." *Review of income and wealth*, 55(3): 661–685.

Covarrubias, Matias, Germán Gutiérrez, and Thomas Philippon. 2020. "From Good to Bad Concentration? US Industries over the past 30 years." *NBER Macroeconomics Annual*, 34(1): 1–46.

Crouzet, Nicolas, and Janice C Eberly. 2019. "Understanding weak capital investment: The role of market concentration and intangibles." National Bureau of Economic Research.

Crouzet, Nicolas, and Neil R. Mehrotra. 2020. "Small and Large Firms over the Business Cycle." *American Economic Review, Forthcoming*.

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

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

De Ridder, Maarten. 2019. "Market power and innovation in the intangible economy."

Dornbusch, Rudiger. 1987. "Exchange Rates and Prices." *American Economic Review*, 77(1): 93–106.

Döttling, Robin, and Lev Ratnovski. 2021. "Monetary policy and intangible investment."

Driscoll, John C., and Aart C. Kraay. 1998. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics*, 80(4): 549–560.

Drozd, Lukasz A, and Jaromir B Nosal. 2012. "Understanding international prices: Customers as capital." *American Economic Review*, 102(1): 364–95.

Erceg, Christopher J, Dale W Henderson, and Andrew T Levin. 2000. "Optimal monetary policy with staggered wage and price contracts." *Journal of Monetary Economics*, 46(2): 281–313.

Etro, Federico, and Lorenza Rossi. 2015. “New-Keynesian Phillips curve with Bertrand competition and endogenous entry.” *Journal of Economic Dynamics and Control*, 51: 318–340.

Ewens, Michael, Ryan H Peters, and Sean Wang. 2019. “Measuring intangible capital with market prices.” National Bureau of Economic Research.

Fitzgerald, Doireann, and Anthony Priolo. 2018. “How Do Firms Build Market Share?” National Bureau of Economic Research.

Galí, Jordi. 2010. “Monetary policy and unemployment.” In *Handbook of Monetary Economics*. Vol. 3, 487–546. Elsevier.

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

Galí, Jordi, Mark Gertler, and J David López-Salido. 2007. “Markups, gaps, and the welfare costs of business fluctuations.” *The Review of Economics and Statistics*, 89(1): 44–59.

Gertler, Mark, and Peter Karadi. 2015. “Monetary Policy Surprises, Credit Costs, and Economic Activity.” *American Economic Journal: Macroeconomics*, 7(1): 44–76.

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

Gilchrist, Simon, and Charles Himmelberg. 1998. “Investment: fundamentals and finance.” *NBER macroeconomics annual*, 13: 223–262.

Gilchrist, Simon, and Egon Zakrajšek. 2012. “Credit spreads and business cycle fluctuations.” *American Economic Review*, 102(4): 1692–1720.

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

Goldberg, Pinelopi, and Rebecca Hellerstein. 2011. “How rigid are producer prices?” *Rebecca and Hellerstein, Rebecca, How Rigid are Producer Prices*.

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

Gorodnichenko, Yuriy, and Michael Weber. 2016. “Are sticky prices costly? Evidence from the stock market.” *American Economic Review*, 106(1): 165–199.

Gourio, Francois, and Leena Rudanko. 2014. “Customer capital.” *Review of Economic Studies*, 81(3): 1102–1136.

Guilloux-Nefussi, Sophie. 2020. “Globalization, market structure and inflation dynamics.” *Journal of International Economics*, 123: 103292.

Gürkaynak, Refet S, Brian Sack, and Eric Swanson. 2005. "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking*, 1(1).

Hall, Bronwyn H, and Josh Lerner. 2010. "The financing of R&D and innovation." In *Handbook of the Economics of Innovation*. Vol. 1, 609–639. Elsevier.

Hall, Robert E. 2012. "The cyclical response of advertising refutes counter-cyclical profit margins in favor of product-market frictions." National Bureau of Economic Research.

Haskel, Jonathan, and Stian Westlake. 2018. *Capitalism without capital: the rise of the intangible economy.* Princeton University Press.

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

Hyun, Jay, Ryan Kim, and Byoungchan Lee. 2021. "Business Cycles with Cyclical Returns to Scale." *Available at SSRN 3184050*.

Jeenas, Priit. 2019. "Firm balance sheet liquidity, monetary policy shocks, and investment dynamics." *Work*, 5.

Jordà, Òscar, and Alan M Taylor. 2016. "The time for austerity: estimating the average treatment effect of fiscal policy." *The Economic Journal*, 126(590): 219–255.

Justiniano, Alejandro, Giorgio E Primiceri, and Andrea Tambalotti. 2010. "Investment shocks and business cycles." *Journal of Monetary Economics*, 57(2): 132–145.

Kaplan, Greg, and Guido Menzio. 2016. "Shopping externalities and self-fulfilling unemployment fluctuations." *Journal of Political Economy*, 124(3): 771–825.

Kimball, Miles S. 1995. "The quantitative analytics of the basic neomonetarist model."

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

Krugman, Paul R. 1986. "Pricing to market when the exchange rate changes."

Lim, Steve C, Antonio J Macias, and Thomas Moeller. 2020. "Intangible assets and capital structure." *Journal of Banking & Finance*, 118: 105873.

Liu, Ernest, Atif Mian, and Amir Sufi. 2018. "Low Interest Rates, Market Power, and Productivity Growth." mimeograph.

Li, Wendy CY, and Bronwyn H Hall. 2020. "Depreciation of business R&D capital." *Review of Income and Wealth*, 66(1): 161–180.

Meier, Matthias, and Timo Reinelt. 2020. "Monetary policy, markup dispersion,and aggregate TFP." *European Central Bank Working Paper Series*.

Melitz, Marc J, and Gianmarco IP Ottaviano. 2008. "Market size, trade, and productivity." *The review of economic studies*, 75(1): 295–316.

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

Morlacco, Monica, and David Zeke. 2020. "Monetary Policy, Customer Capital, and Market Power." USC Manuscript.

Nakamura, Emi, and Dawit Zerom. 2010. "Accounting for incomplete pass-through." *The review of economic studies*, 77(3): 1192–1230.

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

Nekarda, Christopher J, and Valerie A Ramey. 2013. "The cyclical behavior of the price-cost markup." National Bureau of Economic Research.

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

Ottonello, Pablo, and Thomas Winberry. 2017. "Financial Heterogeneity and the Investment Channel of Monetary Policy." Working paper, July.

Paciello, Luigi, Andrea Pozzi, and Nicholas Trachter. 2019. "Price dynamics with customer markets." *International Economic Review*, 60(1): 413–446.

Peters, Ryan H, and Lucian A Taylor. 2017. "Intangible capital and the investment-q relation." *Journal of Financial Economics*, 123(2): 251–272.

Qiu, Zhesheng, and José-Víctor Ríos-Rull. 2022. "Procyclical productivity in new keynesian models." National Bureau of Economic Research.

Ravenna, Federico, and Carl E Walsh. 2006. "Optimal monetary policy with the cost channel." *Journal of Monetary Economics*, 53(2): 199–216.

Ravn, Morten O, Stephanie Schmitt-Grohė, and Lenno Uuskula. 2010. "Deep habits and the dynamic effects of monetary policy shocks." *Journal of the Japanese and International Economies*, 24(2): 236–258.

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

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

Romer, Christina D., and David H. Romer. 2004. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review*, 94(4): 1055–1084.

Rotemberg, Julio J. 1982. “Sticky prices in the United States.” *Journal of political economy*, 90(6): 1187–1211.

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

Sbordone, Argia. 2007. “Globalization and inflation dynamics: the impact of increased competition.”

Smets, Frank, and Rafael Wouters. 2007. “Shocks and frictions in US business cycles: A Bayesian DSGE approach.” *American Economic Review*, 97(3): 586–606.

Stock, James H., and Mark W. Watson. 2018. “Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments.” *The Economic Journal*, 128(610): 917–948.

Stroebel, Johannes, and Joseph Vavra. 2019. “House prices, local demand, and retail prices.” *Journal of Political Economy*, 127(3): 1391–1436.

Surico, Paolo. 2008. “The cost channel of monetary policy and indeterminacy.” *Macroeconomic Dynamics*, 12(5): 724–735.

Traina, James. 2018. “Is aggregate market power increasing? production trends using financial statements.” *Production Trends Using Financial Statements (February 8, 2018)*.

Wang, Olivier, and Iván Werning. 2020. “Dynamic oligopoly and price stickiness.” National Bureau of Economic Research.

Wieland, Johannes F, and Mu-Jeung Yang. 2016. “Financial dampening.” National Bureau of Economic Research.

A Competitive Equilibrium

Definition A.1 (Normalized Competitive Equilibrium). The normalized competitive equilibrium for the economy is defined as

1. Sets of aggregate quantities $\{Y_t, C_t, N_t\}$ and prices $\{p_{jt}, \tilde{p}_t, \tilde{p}_{jt}, Q_{t,t+1}, F_t, i_t, w_t, w_t^*, \Pi_t\}$
2. Sets of firm-level quantities $\{y_{ijt}, f_{ijt}, b_{ijt}\}$, labor demand $\{l_{p,ijt}, l_{f,ijt}, l_{s,ijt}\}$, relative prices $\{p_{ijt}, \tilde{p}_{ijt}\}$, price adjustments $\{\pi_{ijt}\}$, markups $\{\mu_{ijt}\}$, and price elasticity of demands $\{\Theta_{ijt}\}$

such that:

- Given aggregate variables, the optimality conditions are satisfied for all firms;
- Given aggregate prices, representative household's consumption optimization problem and is satisfied; Given labor demand curve, households optimally set their wages;
- All markets (bond, good, and labor) clear;
- The aggregate price index is consistent with firm-level pricing functions;
- The central bank sets interest rates following a Taylor rule.

The equilibrium equations are summarized as:
Euler equation and Stochastic Discount Factor:

$$C_t^{-\sigma} = \beta \mathbf{E}_t \left[C_{t+1}^{-\sigma} \frac{P_t}{P_{t+1}} (1 + i_t) \right] \tag{24}$$

$$Q_{t,t+1} = \beta \frac{C_{t+1}^{-\sigma}}{C_t^{-\sigma}} \tag{25}$$

$$\Pi_t = \frac{P_t}{P_{t-1}} \tag{26}$$

Wage setting frictions:

$$F_t = \frac{\eta_L - 1}{\eta_L} (w_t^*)^{1-\eta_L} C_t^{-\sigma} w_t^{\eta_L} N_t + \beta \gamma_w \mathbf{E}_t \left(\frac{\Pi_t^{\chi_w}}{\Pi_{t+1}} \right)^{1-\eta_L} \left(\frac{w_{t+1}^*}{w_t^*} \right)^{\eta_L - 1} F_{t+1} \tag{27}$$

$$F_t = \theta_L \left(\frac{w_t^*}{w_t} \right)^{-\eta_L(1+\psi_L)} N_t^{1+\psi_L} + \beta \gamma_w \mathbf{E}_t \left(\frac{\Pi_t^{\chi_w}}{\Pi_{t+1}} \right)^{-\eta_L(1+\psi_L)} \left(\frac{w_{t+1}^*}{w_t^*} \right)^{\eta_L(1+\psi_L)} F_{t+1} \tag{28}$$

$$w_t^{1-\eta_L} = \gamma_w \left(\frac{\Pi_{t-1}^{\chi_w}}{\Pi_t} \right)^{1-\eta_L} w_{t-1}^{1-\eta_L} + (1 - \gamma_w) w_t^{*1-\eta_L} \tag{29}$$

Firm-level pricing equation:

$$p_{ijt} = \frac{P_{ijt}}{P_t} = \mu_t(z)(1 - f_{ijt}) \frac{w_t}{z}, \text{ where } w_t = \frac{W_t}{P_t} \tag{30}$$

Firm-level ideal pricing equation:

$$\tilde{p}_{ijt} = p_{ijt} b_{ijt}^{\theta} \tag{31}$$

Firm-level markups:

$$\begin{aligned} \mu_{ijt} &= \frac{\Theta_{ijt}}{(\Theta_{ijt}-1)\left[1-\frac{\phi_p}{2}(\Pi_t \pi_{ijt}-1)^2\right] + \phi_p \Pi_t \pi_{ijt}(\Pi_t \pi_{ijt}-1) - \Gamma_{ijt}}, \text{where} \\ \Gamma_{ijt} &= \phi_p \mathbf{E}_t \left[Q_{t,t+1} \Pi_{t+1} \pi_{ij,t+1}^2 (\Pi_{t+1} \pi_{ij,t+1} - 1) \frac{y_{ij,t+1}}{y_{ijt}} \right] \end{aligned} \tag{32}$$

$$\pi_{ijt} = \frac{p_{ijt}}{p_{ij,t-1}} \tag{33}$$

Firm-level price-elasticity of demand:

$$\Theta_{ijt} = \left[\frac{1}{\rho} + (\frac{1}{\eta} - \frac{1}{\rho}) s_{ijt} \right]^{-1} = \left[\frac{1}{\rho} + (\frac{1}{\eta} - \frac{1}{\rho}) \left(\frac{\tilde{p}_{ijt}}{\tilde{p}_{jt}} \right)^{1-\rho} \right]^{-1} \tag{34}$$

Firm-level demand for good z:

$$y_{ijt} = \left(\frac{p_{ijt}}{\tilde{p}_{jt}} \right)^{-\rho} b_{ijt}^{\theta(1-\rho)} \left(\frac{\tilde{p}_{jt}}{\tilde{p}_t} \right)^{-\eta} Y_t, \tag{35}$$

Firm-level scalable production technology adoption:

$$\psi_F (1-\phi_F)(1-f_{ijt})^{-(\psi_F+1)} = \frac{y_{ijt}}{z_{ij}} \tag{36}$$

Firm-level advertising and marketing:

$$\psi_B (1-\phi_B)(1-b_{ijt})^{-(\psi_B+1)} w_t = \theta(1-\rho) \left[1 - \frac{1}{\mu_{ijt}} \right] \frac{p_{ijt} y_{ijt}}{b_{ijt}} \tag{37}$$

Firm-level demand for production labor demand:

$$l_{p,ijt} = \frac{y_{ijt}(1-f_{ijt})}{z_{ij}} \tag{38}$$

Firm-level demand for scalable production technology labor:

$$l_{f,ijt} = (1-\phi_F) \left[\left(\frac{1}{1-f_{ijt}} \right)^{\psi_F} - 1 \right] \tag{39}$$

Firm-level for advertising and marketing labor:

$$l_{s,ijt} = (1-\phi_B) \left[\left(\frac{1}{1-b_{ijt}} \right)^{\psi_B} - 1 \right] \tag{40}$$

Labor market clearing:

$$N_t = \int_0^1 \left(\sum_i^{N_j} l_{p,ijt} + \sum_i^{N_j} l_{f,ijt} + \sum_i^{N_j} l_{s,ijt} \right) dj \tag{41}$$

Resource constraint:

$$Y_t = C_t + \int_0^1 \left(\sum_i^{N_j} \frac{\phi_p}{2} (\Pi_t \pi_{ijt} - 1)^2 p_{ijt} y_{ijt} \right) dj \tag{42}$$

Aggregate pricing index:

$$1 = \int_0^1 p_{jt}^{1-\eta} dj, \text{where } p_{jt} = \left[\sum_i^{N_j} p_{ijt}^{1-\rho} \right]^{\frac{1}{1-\rho}} \tag{43}$$

Ideal pricing index:

$$\tilde{p}_{jt} = \left[\sum_i^{N_j} (p_{ijt} b_{ijt}^{\theta})^{1-\rho} \right]^{\frac{1}{1-\rho}} = \left[\sum_i^{N_j} \tilde{p}_{ijt}^{1-\rho} \right]^{\frac{1}{1-\rho}} \tag{44}$$

$$\tilde{p}_t = \left(\int_0^1 \tilde{p}_{jt}^{1-\eta} dj \right)^{\frac{1}{1-\eta}} \tag{45}$$

Monetary policy rule:

$$1 + i_t = (1 + i_{ss}) \left(\frac{\Pi_t}{\Pi_{ss}} \right)^{\phi^\pi} \left(\frac{Y_t}{Y_{ss}} \right)^{\phi^y} \epsilon_t^m, \phi^\pi > 0, \phi^y > 0 \tag{46}$$

Definition A.2 (Steady State Equilibrium). The zero-inflation steady state equilibrium for the economy with fully symmetric sectors are characterized by constant values a set of aggregate quantities $\{Y, C, N\}$ and prices $\{w, w^*, \tilde{p}\}$, and firm-level quantities $\{y(z), b(z), f(z)\}$, prices $\{p(z), \tilde{p}(z)\}$, labor demand $\{l_p(z), l_f(z), l_s(z)\}$, markups $\{\mu(z)\}$ and elasticity of demand $\{\Theta(z)\}$ that satisfy:
Pricing equation for all z:

$$p(z) = \mu(z)(1 - f(z)) \frac{w}{z} \tag{47}$$

Ideal pricing equation for all z:

$$\tilde{p}(z) = p(z) b(z)^{\theta} \tag{48}$$

Desired markups for all z:

$$\mu(z) = \frac{\Theta(z)}{\Theta(z) - 1} \tag{49}$$

Price-elasticity of demand:

$$\Theta(z) = \left[\frac{1}{\rho} + (\frac{1}{\eta} - \frac{1}{\rho}) s(z) \right]^{-1}, \text{ where } s(z) = \left(\frac{\tilde{p}(z)}{\tilde{p}} \right)^{1-\rho}, \tilde{p} = \left[\sum_z \tilde{p}(z)^{1-\rho} \right]^{\frac{1}{1-\rho}} \tag{50}$$

Demand for good z:

$$y(z) = \left(\frac{p(z)}{\tilde{p}} \right)^{-\rho} b(z)^{\theta(1-\rho)} Y \tag{51}$$

Scalable production technology adoption for all z:

$$\psi_F (1 - \phi_F)(1 - f(z))^{-(\psi_F + 1)} = \frac{y(z)}{z} \tag{52}$$

Advertising and marketing for all z:

$$w \psi_B (1 - \phi_B)(1 - b(z))^{-(\psi_B + 1)} = \theta(1 - \rho) \left(1 - \frac{1}{\mu(z)} \right) \frac{p(z) y(z)}{b(z)} \tag{53}$$

Production labor demand for all z:

$$l_p(z) = \frac{y(z)(1-f(z))}{z} \tag{54}$$

Scalable production technology labor for all z:

$$l_f(z) = (1-\phi_F)\left[\left(\frac{1}{1-f(z)}\right)^{\psi_F} - 1\right] \tag{55}$$

Advertising and marketing labor demand for all z:

$$l_s(z) = (1-\phi_B)\left[\left(\frac{1}{1-b(z)}\right)^{\psi_B} - 1\right] \tag{56}$$

Labor market clearing:

$$N = \sum_z l_p(z) + \sum_z l_f(z) + \sum_z l_s(z) \tag{57}$$

Optimal wage setting:

$$w^* = \frac{\psi_L}{\psi_L - 1}\frac{\theta_L N^{\psi_L}}{C^{-\sigma}} \tag{58}$$

$$w = w^* \tag{59}$$

Resource constraint:

$$Y = C \tag{60}$$

Aggregate pricing index:

$$1 = \sum_z p(z)^{1-\rho} \tag{61}$$

Ideal pricing index:

$$\tilde{p} = \left[\sum_z \tilde{p}(z)^{1-\rho}\right]^{\frac{1}{1-\rho}} \tag{62}$$

Solving for steady state. Solving for the zero-inflation steady state (SS) of the economy amounts to a two-dimensional fixed point problem in real wage rate w and aggregate demand Y:

1. Given an initial guess of w and Y, and a given draw of firm-level productivity, simultaneously solve for steady state levels of firm-level relative prices and ideal prices, scalable production technology and customer acquisition expenditure, desired markups, and price-elasticity of demand.

2. Given optimal firm-level relative prices, use the aggregate pricing equation

$$1 = \sum_z p(z)^{1-\rho}$$

to verify that the guessed real wage rate and aggregate demand are consistent. Update the guess, and iterate until the two fixed points are reached.

3. Recover the remaining endogenous firm-level (labor demand, quantity of variety produced) and aggregate variables (aggregate labor, ideal pricing index) using the remaining SS equations.

A.1 Log-linearization

I log-linearize the normalized competitive equilibrium around the zero-inflation steady state. For each variable, I define $\hat{var}_t = \log var_t - \log var$, where var is the steady-state value for the variable var_t.

Definition. The log-linearized competitive equilibrium of the economy is defined as a sequence of aggregate-level variables $\{\hat{Y}_t, \hat{C}_t, \hat{N}_t, \hat{F}_t, \hat{w}_t, \hat{w}_t^*, \hat{i}_t, \hat{\Pi}_t, \hat{\tilde{p}}_t\}$, and firm-level variables $\left\{\{\hat{p}_t(z)\}, \{\hat{\tilde{p}}_t(z)\}, \{\hat{\Pi}_t(z)\}, \{\hat{\mu}_t(z)\}, \{\hat{\Theta}_t(z)\}, \{\hat{f}_t(z)\}, \{\hat{b}_t(z)\}, \{\hat{y}_t(z)\}\right\}_{z\in\Omega_z}$ and labor demand $\left\{\{\hat{l}_{p,t}(z)\}, \{\hat{l}_{f,t}(z)\}, \{\hat{l}_{s,t}(z)\}\right\}_{z\in\Omega_z}$ that satisfy the following equations, for a given sequence of exogenous shocks $\{\hat{\epsilon}_t^m\}$.

Euler equation:

$$-\sigma \hat{C}_t = -\sigma \mathbf{E}_t \hat{C}_{t+1} + \hat{i}_t - \mathbf{E}_t \hat{\Pi}_{t+1} \tag{63}$$

Wage setting frictions:

$$\begin{aligned}\hat{F}_t = &(1-\beta\gamma_w)\left((1-\eta_L)\hat{w}_t^* - \sigma\hat{C}_t + \eta_L \hat{w}_t + \hat{N}_t\right) \\ &+ \beta\gamma_w \mathbf{E}_t\left(\hat{F}_{t+1} - (1-\eta_L)(\hat{\Pi}_{t+1} - \chi_w\hat{\Pi}_t + \hat{w}_{t+1}^* - \hat{w}_t^*)\right)\end{aligned} \tag{64}$$

$$\begin{aligned}\hat{F}_t = &(1-\beta\gamma_w)\left(\eta_L(1+\psi_L)(\hat{w}_t - \hat{w}_t^* + (1+\psi_L)\hat{N}_t)\right) \\ &+ \beta\gamma_w \mathbf{E}_t\left(\hat{F}_{t+1} + \eta_L(1+\psi_L)(\hat{\Pi}_{t+1} - \chi_w\hat{\Pi}_t + \hat{w}_{t+1}^* - \hat{w}_t^*)\right)\end{aligned} \tag{65}$$

$$\gamma_w(\hat{\Pi}_t - \chi_w\hat{\Pi}_{t-1} + \hat{w}_t - \hat{w}_{t-1}) = (1-\gamma_w)(\hat{w}_t^* - \hat{w}_t) \tag{66}$$

Pricing equation for all z:

$$\hat{p}_t(z) = \hat{\mu}_t(z) + \hat{w}_t - \frac{f(z)}{1-f(z)}\hat{f}_t(z) \tag{67}$$

Ideal pricing equation for all z:

$$\hat{\tilde{p}}_t(z) = \hat{p}_t(z) + \theta\hat{b}_t(z) \tag{68}$$

Markups for all z:

$$\hat{\mu}_t(z) = \frac{-1}{\Theta(z)-1}\hat{\Theta}_t(z) - \frac{\phi_p}{\Theta(z)-1}\hat{\Pi}_t(z) + \frac{\beta\phi_p}{\Theta(z)-1}\mathbf{E}_t\hat{\Pi}_{t+1}(z) \tag{69}$$

Firm-level inflation for all z:

$$\hat{\Pi}_t(z) = \hat{\Pi}_t + \hat{p}_t(z) - \hat{p}_{t-1}(z) \tag{70}$$

Price-elasticity of demand:

$$\hat{\Theta}_t(z) = -\left(\frac{\Theta(z)^{-1} - \rho^{-1}}{\Theta(z)^{-1}}\right)(1-\rho)(\hat{\tilde{p}}_t(z) - \hat{\tilde{p}}_t) \tag{71}$$

Firm-level scalable production technology adoption:

$$\hat{f}_t(z) = \frac{1-f(z)}{f(z)}\frac{1}{\psi_F+1}\hat{y}_t(z) \tag{72}$$

Firm-level advertising and marketing:

$$\left[1+\frac{b(z)}{1-b(z)}(\psi_B+1)\right]\hat{b}_t(z)=\left(\frac{\mu(z)^{-1}}{1-\mu(z)^{-1}}\right)\hat{\mu}_t(z)+\hat{p}_t(z)+\hat{y}_t(z)-\hat{w}_t \tag{73}$$

Demand for good z:

$$\hat{y}_t(z)=-\rho\left(\hat{\tilde{p}}_t(z)-\hat{\tilde{p}}_t\right)+\theta(1-\rho)\hat{b}_t(z)+\hat{Y}_t \tag{74}$$

Firm-level demand for production labor demand:

$$\hat{l}_{p,t}(z)=\hat{y}_t(z)-\frac{f(z)}{1-f(z)}\hat{f}_t(z) \tag{75}$$

Firm-level demand for scalable production technology labor:

$$\hat{l}_{f,t}(z)=\left(\psi_F+\frac{\psi_F(1-\phi_F)}{l_f(z)}\right)\frac{f(z)}{1-f(z)}\hat{f}_t(z) \tag{76}$$

Firm-level for advertising and marketing labor:

$$\hat{l}_{s,t}(z)=\left(\psi_B+\frac{\psi_B(1-\phi_B)}{l_s(z)}\right)\frac{b(z)}{1-b(z)}\hat{b}_t(z) \tag{77}$$

Labor market clearing:

$$N\hat{N}_t=\sum_z l_p(z)\hat{l}_{p,t}(z)+\sum_z l_f(z)\hat{l}_{f,t}(z)+\sum_z l_s(z)\hat{l}_{s,t}(z) \tag{78}$$

Resource constraint:

$$\hat{Y}_t=\hat{C}_t \tag{79}$$

Price index:

$$0=\sum_z(1-\rho)p(z)^{1-\rho}\hat{p}_t(z) \tag{80}$$

Ideal pricing index:

$$\hat{\tilde{p}}_t=\sum_z\left(\frac{\tilde{p}(z)}{\tilde{p}}\right)^{1-\rho}\hat{\tilde{p}}_t(z) \tag{81}$$

Monetary policy rule:

$$\hat{i}_t=\phi^\pi\hat{\Pi}_t+\phi^y\hat{C}_t+\hat{\epsilon}_t^m \tag{82}$$

B Data Description and Variable Constructions

Aggregate data. I obtain the following series of aggregate data from their respective sources and aggregate them to quarterly frequency if necessary.

- I obtain macroeconomic time series data, including the Fed Fund rates, real GDP, CPI (the Consumer Price Index: Total All Items for the United States), Unemployment Rate (UNRATE), median real wage (Employed full time: Median usual weekly real earnings: Wage and salary workers: 16 years and over (LES1252881600Q)) from the Federal Reserve Bank of St. Louis FRED database. Credit spread is the excess bond premium and is obtained from Gilchrist and Zakrajšek (2012).

- From NIPA Table 1.1.9. (Implicit Price Deflators for Gross Domestic Product: Quarterly), I obtain GDP deflator (line 1) and nonresidential fixed investment good deflator (line 9).

- For aggregate markups, I use the estimated series from Nekarda and Ramey (2020).

- For high frequency identification, I obtain two monetary policy shock instruments from Gürkaynak et al. (2005) with an extension by Gorodnichenko and Weber (2016)(Monthly, 1990 - 2009).

- For robustness, I use Romer and Romer (2004)'s narrative instruments to identify monetary policy shocks (extended by Wieland and Yang (2016), Monthly, 1969 - 2007).

Firm-level data. The primary source of data draws from the quarterly firm-level balance sheet of U.S. publicly listed firms from Compustat for the period 1980 - 2009 from the Wharton Research Data Services (WRDS). The data-set contains information on firm-level revenues, operating and capital expenditures, as well as industry classification. I use a standard cleaning procedure to construct the baseline sample used in the empirical analysis. The industry classification is based on the North American Industry Classification System (NAICS), and the focus is on industries at the three-digit NAICS level. Firms in utilities, finance, insurance, and real estate, and public administration sectors (two-digit NAICS codes 52, 53,22,and 99) are dropped as a standard practice. To deal with missing values in the panel, for key variables (e.g., sales, costs of goods sold, sales general and administration expenses, and (net) property, plant, and equipment), I use a linear interpolation of their neighboring values to impute one-quarter gaps. I keep the missing values if there are two-quarter or greater gaps. I then deflate non-investment-related variables such as sales, cost of goods sold, and SG&A by GDP deflator (NIPA Table 1.1.9. GDP deflator (line 1)) to form real measures of revenue and variable input.

Following the standard procedure in the production function estimation literature, I use the perpetual inventory method to create a proxy for the firm-level capital stock. To do so, I initialize the capital stock by identifying the first available item of (gross) property, plant, and equipment for a given firm as $K_{i0} =$ gross PPE_{i0}, then iterate forward to construct capital stock as $K_{it} = K_{it-1} +$ net $PPE_{it} -$ net PPE_{it-1}. I deflate net PPE by the nonresidential fixed investment good deflator (NIPA Table 1.1.9. GDP deflator (line 9)) to construct real measures of capital stock.

I closely follow Peters and Taylor (2017) to make adjustment to the measures of SG&A net of R&D. Specifically, I construct the adjusted SG&A as SG&A minus R&D expenditure minus in-process R&D (Compustat item *xsgaq* - *xrdq* - *rdipq*). If *xrdq* exceeds *xdgaq* but is less than *cogsq*, I measure SG&A net of R&D as *xsgaq*, the rationale behind is to check if R&D expenditures are reported in COGS instead of SG&A. Finally, I set *xsgaq*, *xrdq*, and *rdipq* to zero when missing.

For variables that proxy for financial frictions I use: (i) size measured as total asset (*atq*), (ii) liquidity (*cheq*/*atq*), and (iii) leverage (*(dlcq + dlttq)*/*atq*).

To obtain a baseline sample, I drop observations with negative sales, costs of goods sold, and capital. To remove outlier bias, I also drop observations with either (i) sales less then 1, (ii) percentage growth in sales less than -67%, or (iii) percentage growth in sales more than 100%.

A complementary data-set is the annual firm-level balance sheet of U.S.publicly listed firms from Compustat for the period 1980 - 2019 from WRDS. I apply a similar cleaning procedure to the data-set. I obtain measures for advertising expenditures (*xad*), R&D expenditure (*xrd*), and pre-tax foreign incomes (*pifo*), in order to construct measures of firm-level and sectoral mean advertising and R&D intensities (measured as a ratio between the expenditures and either sales or total assets), and firm-level exporter status (if the firm reports pre-tax foreign incomes at any point in the sampling period that corresponds to the actual regression).

Interpreting SG&A. The uncertainty in the classification of SG&A arises from the fact that firm-level balance-sheet data do not capitalize intangible investment and instead report it as an expenditure under the tag Sales, General, and Administrative expenses (SG&A). Meanwhile, SG&A constitutes the second main component of firms' total operating costs after Costs of Goods Sold (COGS), and encompasses almost all internal intangible creating expenditures (for instance, R&D, advertising and marketing, and management expenses, among others).[68] US firms have experienced an increasing trend in SG&A as a fraction of total operating costs (OPEX) since the 1980s, meanwhile, COGS share of total costs has been declining (Traina, 2018; Covarrubias et al., 2020). This technological shift in firms' production is consonant with the underlying advancements in Information and Communication Technology (ICT) in the wider economy.

The nature of SG&A expenses, however, is hard to ascertain. While 30% of SG&A spending is generally classified as an investment in the intangible capital stock, Peters and Taylor (2017) interpret the remaining 70% as part of operating costs in support of the current period's profits, i.e., broadly speaking, intangible input. I follow their interpretation and consider the former as an investment in future profitability, and the latter as an expenditure to support current profitability. To compound matters further, SG&A is an encompassing variable that reflect spending on various functions of firm. High level of SG&A expenditures could be indicative that the firm spends significant resources in cultivating and maintaining a strong customer base (Gourio and Rudanko, 2014; Morlacco and Zeke, 2020), it could also imply high fixed costs of operation and hence a potential increase in scale economies (Gilchrist and Himmelberg, 1998). Quarterly balance sheet data do not differentiate expenditure on advertising and customer acquisition versus expenditure on fixed-cost intensive technologies. In fact, even if quarterly advertising expenditure is available, it may not be the best proxy for customer acquisition and maintenance expenditures as the latter, broadly speaking, should include firms' spending on salespeople, marketing, customer relation, branding, and advertising to influence consumer demand. Therefore I do not attempt to differentiate them in the data and treat an increase in SG&A as indicative of an increase in the use of intangible inputs in general.

C Markup Estimation

I closely follow De Loecker and Warzynski (2012) to estimate firm-level markups using quarterly balance sheet data in a production function. The economy is populated with N firms that are heterogeneous in productivity but operate under the same production technology. At time t, each firm i optimally minimizes its cost of production subject to producing a level of output $Q_{it} = Q_{it}(\Omega_{it}, \mathbf{V}_{it}, K_{it})$, where $\mathbf{V}_{it} = (V_{it}^1, ..., V_{it}^J)$

[68] Standard accounting practices dictates that a firm's total expenses can be either classified as Capital Expenses (incurred to build up the firm's capital stock) or as Operating expenses (under Compustat item *xopr*, and are incurred through the firm's routine business operation). Operating expenses (OPEX) can be further broken down into COGS (that accounts for the direct inputs to production, and include labor costs and material inputs, among others) and SG&A (that measures the indirect input of production, including costs incurred through R&D, advertising and marketing, and IT staff expenses).

corresponds a vector of J variable inputs, K_{it} is the capital input, and Ω_{it} is the firm-level productivity. Firms enjoy market power to set price in their own product market solely and therefore would take all input prices as given. They set up the following Lagrangian to solve the dual problem of cost-minimization as:

$$\mathscr{L}(V_{it}^{1},...,V_{it}^{J},K_{it},\Omega_{it}) = \sum_{j=1}^{J} P_{it}^{j} V_{it}^{j} + r_{it}K_{it} + \lambda_{it}(Q_{it} - Q_{it}(\cdot)),$$

where P_{it}^{j} is the price of variable input j, and r_{it} is the rental rate of capital. The first order condition of the cost-minimization problem with respect to a variable input j leads to an expression for markup as the ratio between the output elasticity of the variable input and the its factor share of output:

$$\mu_{it} = \frac{\theta_{it}^{j}}{\alpha_{it}^{j}} \quad \forall j = 1,...,J,$$

where θ_{it}^{j} is to the output elasticity with respect to any of the variable input j, and α_{it}^{j} measures the cost share of the said input over total revenue. In general, α_{it}^{j} can be easily computed from firm-level data, whereas θ_{it}^{j} requires estimation.

In theory, output elasticity can be estimated based on any variable input (for instance, energy, materials, and labor). In practice, data obtained from Compustat pose a practical challenge, because there is no breakdown of various inputs, instead, two board items - Cost of Goods Sold (COGS) and Operating expenses (OPEX) - capture the total operating costs of production that is not capital expenses. De Loecker, Eeckhout and Unger (2020) focus on COGS as a their main proxy for variable input. Recently, this approach has drawn criticism from Traina (2018) and Covarrubias et al. (2020), who argue that the increasing importance of SG&A as a share of total operating expenses results in an upward bias in the COGS-based markups. Given the ongoing debate in the empirical literature that is beyond the scope of this paper, I adopt both approaches and estimate two sets of markups based on both COGS and OPEX.

Taking COGS as the baseline proxy for variable input, I estimate its output elasticity in the following industry-specific production function that use capital k_{it} (measured by capital stock) and a composite variable input v_{it} (measured by COGS) in the production process:

$$q_{it} = f_{st}(v_{it},k_{it};\theta_{st}) + \omega_{it} + \epsilon_{it},$$

where lower case letters denote logs of each variable, ω_{it} is log of the firm-level productivity, and ϵ_{it} represents a shock to productivity that can only be observed after input choices were made.

I use a three-digit industry-specific Cobb-Douglas production function in my baseline estimation:

$$f_{s}(v_{it},k_{it};\theta_{s}^{V},\theta_{s}^{K}) = \theta_{s}^{V} v_{it} + \theta_{s}^{K} k_{it}.$$

The estimation procedure occurs in two stages. In the first stage, I regress log real sales q_{it} on a non-paramedic function proxied by a third-order polynomial of variable input v_{it} and capital k_{it} to obtain a predicted (purged) level of output, denoted by $\hat{q}_{it}$. In the second stage, I then estimate output elasticity θ_{s} in an iterative procedure using Generalized method of moments (GMM) methods. Specifically, I obtain firm productivity as $\omega_{it}(\theta_{st}) = \hat{q}_{it} - f_{st}(v_{it}, k_{it}; \theta_{st})$, then compute the shock to productivity ξ_{it} as the residuals of regressing ω_{it} on its first lag (that is assuming ω_{it} follows an AR(1) process whereby $\omega_{it} = g_t(\omega_{it-1}) + \xi_{it}$). The moment condition hinges on the assumption that for inputs choices that are predetermined, they are orthogonal to the shock to firm productivity:[69]

$$\mathbf{E}\left[\xi_{it}(\theta_s^V, \theta_s^K)\begin{pmatrix} v_{it-4} \\ k_{it} \end{pmatrix}\right] = 0.$$

Minimizing the relevant GMM objective functions as implied the moment condition defined above yields the sectoral output elasticities θ_s^V and θ_s^K. As a robustness check, I also use a two-digit industry-time specific Cobb-Douglas production function to estimate sectoral-time specific output elasticities with a similar procedure in a sequence of five-year backward-looking rolling window regressions. I then calculate firm-level markups μ_{it} as the ratio of the output elasticity and the revenue share of COGS. I use the range of firm-level data that corresponds to the range of GSS shocks in my baseline estimation that focuses on a 3-digit sectoral production function to minimize the impacts of time-varying production function. For the rolling-window regressions, I make use of the full range of firm-level data.

As a robustness check, I consider an alternative to the Cobb-Douglas production function - a translog production function that is more flexible in both a three-digit industry-specific and a two-digit industry-time-specific setting as follows:[70]

$$f_s(v_{it}, k_{it}; \theta_s) = \beta_s^V v_{it} + \beta_s^K k_{it} + \beta_s^{VV} v_{it}^2 + \beta_s^{KK} k_{it}^2.$$

I adopt an analogous estimation procedure as before, in which the second stage moment condition is given by:

$$\mathbf{E}\left[\xi_{it}(\theta_{st}^V, \theta_{st}^K)\begin{pmatrix} v_{it-4} \\ v_{it-4}^2 \\ k_{it} \\ k_{it}^2 \end{pmatrix}\right] = 0.$$

I then calculate the firm-time-specific output elasticities as follow:

$$\theta_{it}^V = \beta_s^V + 2\beta_s^{VV} v_{it},$$

and compute the COGS-based markups in a tranlog production function as before.

For the OPEX-based markups, I deploy a similar method, in which I use OPEX in place of COGS throughout the estimation procedure to recover the OPEX-based markups in both a Cobb-Douglas and a translog production function.

Finally, in my baseline regressions, I drop by quarter the bottom and top 5% of the estimated markups to remove outliers. In Table 3, I report the summary statistics for each type of markups.

[69] De Loecker et al. (2020) use the first lag of the variable input as an instrument in annual data. Because I am using a quarterly data, I use the fourthly lag instead to factor in the impact of seasonality.

[70] Cross-terms are omitted to prevent measurement errors in capital stock to directly affect the estimation of the elasticity of the variable input.

Table 3: Summary Statistics of Estimated Markups

	Mean	Std.Dev	Median	10th Pct.	90th Pct.	Observations
3-digit sectoral production function						
COGS Cobb-Douglas	1.498	0.256	1.460	1.198	1.833	184,047
COGS translog	1.496	0.269	1.453	1.194	1.851	184,047
OPEX Cobb-Douglas	1.143	0.107	1.130	1.019	1.300	186,912
OPEX translog	1.145	0.110	1.131	1.018	1.306	186,912
2-digit sectoral-time production function						
COGS Cobb-Douglas	1.459	0.196	1.447	1.220	1.715	193,548
COGS translog	1.451	0.194	1.449	1.213	1.683	193,548
OPEX Cobb-Douglas	1.095	0.070	1.094	1.006	1.182	175,635
OPEX translog	1.099	0.070	1.097	1.011	1.187	175,635

Note: The table reports the summary statistics for 1990-2009, the sampling period corresponds to the coverage of the GSS shocks.

D Theoretical Extension

Internalizing pricing decision in the first stage. In the baseline model, individual firm's problems are solved in two stages without internalizing the impacts of the firm's pricing decision has on the first stage choice of intangible inputs. In other words, the two partial derivatives of intangible inputs with respect to price are assumed to be zero ($\frac{\partial f^*_{ijt}}{\partial P_{ijt}} = 0, \frac{\partial b^*_{ijt}}{\partial P_{ijt}} = 0$).

In this section, I relax this myopic assumption such that firms solves their problem fully. In the first stage, individual firm still solves the same static problem as in the baseline model, taking prices as given. This leads to optimal choices of intangible inputs $f^*_{ijt}(P_{ijt})$ and $b^*_{ijt}(P_{ijt})$ for the scalable production technology and sales/advertising/marketing, respectively.

In the second stage, individual firms solve the following maximization problem subject to the inverse demand function:

$$\max_{\{P_{ij\tau}, y_{ij\tau}\}_{\tau \geq t}} \quad \mathbf{E}_t \sum_{\tau \geq t} \frac{Q_{t,\tau}}{P_\tau} \left\{ \left(P_{ij\tau} - (1 - f^*_{ij\tau}(P_{ijt})) \frac{W_\tau}{z_{ij}} \right) y_{ij\tau} - \underbrace{\frac{\phi_p}{2} \left(\frac{P_{ij\tau}}{P_{ij,\tau-1}} - 1 \right)^2 P_{ij\tau} y_{ij\tau}}_{\text{Rotemberg costs}} \right\} \tag{83}$$

$$\text{s.t. } P_{ij\tau} = \left(\frac{y_{ij\tau} b^*_{ij\tau}(P_{ijt})^{\theta(\rho-1)}}{X_{j\tau}} \right)^{-\frac{1}{\rho}} \left(\frac{X_{j\tau}}{Y_\tau} \right)^{-\frac{1}{\eta}} \tilde{P}_\tau,$$

where the optimal choices for intangible inputs now enter the maximization problem as a function of prices. Solving the second-stage dynamic problem leads to an optimal price as a time-varying markup over the firm's marginal cost:

$$P_{ijt} = \mu_{ijt}(1 - f^*_{ijt}) \frac{W_t}{z_{ij}}, \tag{84}$$

where

$$\mu_{ijt} = \frac{\Theta_{ijt}}{(\Theta_{ijt} - \mathscr{W}_{ijt})\left[1 - \frac{\phi_p}{2}(\pi_{ijt}-1)^2\right] - \mathscr{W}_{ijt}\left[\frac{W_t}{z_{ij}}\frac{\partial f^*_{ijt}}{\partial P_{ijt}} - \phi_p \pi_{ijt}(\pi_{ijt}-1) + \Gamma_{ijt}\right]},$$

$$\mathscr{W}_{ijt} = \frac{1}{1 + \frac{\theta(1-\rho)}{\rho}\frac{P_{ijt}}{b^*_{ijt}}\frac{\partial b^*_{ijt}}{\partial P_{ijt}}}$$

The output-pricing decision leads to optimal time-varying markups μ_{ijt} that depends on both the firm's market share s_{ijt} and the degree of nominal rigidities. Internalizing pricing decision further introduces a wedge $\mathscr{W}_{ijt}$ into the markup. Notice that the wedge equals to one when both $\frac{\partial f^*_{ijt}}{\partial P_{ijt}} = 0$ and $\frac{\partial b^*_{ijt}}{\partial P_{ijt}} = 0$ hold and the baseline case is recovered.

Applying the Implicit Function Theorem, $\frac{\partial f^*_{ijt}}{\partial P_{ijt}}$ and $\frac{\partial b^*_{ijt}}{\partial P_{ijt}}$ can be computed and are given below:

$$\frac{\partial b^*_{ijt}}{\partial P_{ijt}} = \frac{-\mathscr{B}_{ijt,1} - \mathscr{B}_{ijt,3}\mathscr{F}_{ijt,1}}{\mathscr{B}_{ijt,2} + \mathscr{B}_{ijt,3}\mathscr{F}_{ijt,2}} \tag{85}$$

$$\frac{\partial f^*_{ijt}}{\partial P_{ijt}} = \mathscr{F}_{ijt,1} + \mathscr{F}_{ijt,2}\frac{\partial b^*_{ijt}}{\partial P_{ijt}}, \tag{86}$$

where

$$\mathscr{B}_{ijt,1} = \frac{1}{\psi_B + 1}\frac{1 - b^*_{ijt}}{P_{ijt}}\left[\frac{\mu_{ijt}}{\mu_{ijt}-1} - \rho + (\rho - \eta)s_{ijt}\right]$$

$$\mathscr{B}_{ijt,2} = -1 + \frac{1}{\psi_B + 1}\frac{1 - b^*_{ijt}}{b^*_{ijt}}\left[\theta(1-\rho) - 1 + \theta(\rho-\eta)s_{ijt}\right]$$

$$\mathscr{B}_{ijt,3} = \frac{1}{\psi_B + 1}\frac{1 - b^*_{ijt}}{P_{ijt}}\left[\frac{\mu_{ijt}}{\mu_{ijt}-1}\frac{W_t}{z_{ij}}\right]$$

$$\mathscr{F}_{ijt,1} = \frac{1}{\psi_F + 1}\left[\frac{1}{\psi_F(1-\phi_F)z_{ij}}\right]^{-\frac{1}{\psi_F+1}}\left(\frac{y_{ijt}^{-\frac{1}{\psi_F+1}}}{P_{ijt}}\right)\left[-\rho + (\rho-\eta)s_{ijt}\right]$$

$$\mathscr{F}_{ijt,2} = \frac{1}{\psi_F + 1}\left[\frac{1}{\psi_F(1-\phi_F)z_{ij}}\right]^{-\frac{1}{\psi_F+1}}\left(\frac{y_{ijt}^{-\frac{1}{\psi_F+1}}}{b^*_{ijt}}\right)\left[\theta(1-\rho) + \theta(\rho-\eta)s_{ijt}\right]$$

I likewise impose symmetric sector assumptions and drop the subscript ij and use firm productivity z to denote individual firm in subsequent analysis.

Focusing on the zero-inflation steady state, I log-linearize the optimal pricing decision to derive the firm-level inflation equation:

$$\begin{aligned}\hat{\Pi}_t(z) = &-\frac{1}{\phi_p}\left[1 + \frac{w}{z}\frac{\partial f^*(z)}{\partial p(z)}\right]\left(\hat{\Theta}_t(z) + \frac{\mathscr{W}(z)-1}{\mathscr{W}(z)}\hat{\epsilon}_{b_t(z)}\right)\\ &+ \frac{1}{\phi_p}\left(\frac{\Theta(z) - \mathscr{W}(z)}{\mathscr{W}(z)}\right)\hat{mc}_t(z) + \frac{1}{\phi_p}\frac{w}{z}\frac{\partial f^*(z)}{\partial p(z)}\left[\hat{\epsilon}_{f_t(z)} - \frac{f^*(z)}{1-f^*(z)}\hat{f}^*_t(z)\right]\\ &+ \beta\mathbf{E}_t\hat{\Pi}_{t+1}(z),\end{aligned} \tag{87}$$

where $\hat{mc}_t(z) = -\hat{\mu}_t(z) = \hat{w}_t - \frac{f(z)}{1-f(z)}\hat{f}_t(z) - \hat{p}_t(z)$ is the log deviation of real marginal cost from its steady-state value, and $\hat{\epsilon}_{b_t(z)}$ and $\hat{\epsilon}_{f_t(z)}$ are log deviation of price elasticity of sales/advertising/marketing and scalable production technology from their respective steady states.

Similarly, log-linearizing the two first order conditions for optimal intangible inputs and rearranging leads to:

$$\begin{aligned}\hat{b}^*_t(z) = &\frac{1}{1+\frac{b^*(z)}{1-b^*(z)}(1+\psi_B)}\left(\frac{\mu(z)^{-1}}{1-\mu(z)^{-1}}+1\right)\hat{\mu}_t(z) \\ &+\frac{f^*(z)}{1-f^*(z)}\left[(1+\psi_F)-\frac{1}{1+\frac{b^*(z)}{1-b^*(z)}(1+\psi_B)}\right]\hat{f}^*_t(z)\end{aligned} \tag{88}$$

The goal of the analysis is to show that the mechanism proposed in the main text still holds in the more involved way of solving the model, namely, to demonstrate that first, inflation is driven by pro-cyclical firm market power, and second, there exits synergy between the two intangible inputs and their co-movement following an expansionary monetary policy shock leads to pro-cyclical markup and firm-level inflation.

Notice that the augmented inflation equation depends on steady state firm-level variables that include the perceived elasticity of demand, choice of scalable production technology, and the wedge, among others. I solve for the steady state equilibrium of the model numerically and discuss the short-run pricing dynamics in relation to the steady state market power of firms.

Proposition 1. *Individual firm inflation depends **positively** on changes in real marginal cost (and by definition, depends **negatively** on changes in price-cost markup, $\hat{\mu}_t(z)$), inflation expectation, changes in scalable production technology adoption, and changes in price elasticity of sales / advertising / marketing, and depends **negatively** on changes in price elasticity of scalable production technology and changes in perceived elasticity of demand $\hat{\Theta}_t(z)$. A decline in $\hat{\Theta}_t(z)$ represents a strengthening of the firm's market power, which in turn drives up inflation.*

Proof. See Equation 87. □

Proposition 2. *Changes in sales / advertising / marketing $\hat{b}^*_t(z)$ depends **positively** on changes in markup $\hat{\mu}_t(z)$ and changes in scalable production technology $\hat{f}^*_t(z)$.*

Proof. From Equation 88, change in scalable production technology $\hat{f}^*_t(z)$ depends positively on changes in customer base $\hat{b}^*_t(z)$ and aggregate demand $\hat{Y}_t$. A sufficiently large increase in $\hat{f}^*_t(z)$ implies a pro-cyclical movement in markup, both the pro-cyclical movements in markups (indirect effect) and in scalable production technology imply a pro-cyclical movement in sales/advertising/marketing. □

It implies that if there is pro-cyclical movement in customer acquisition, there would be pro-cyclical movement in scalable production technology adoption as the latter depends positively on changes in customer acquisition and aggregate demand. A sufficiently pro-cyclical movement in scalable production technology that out-weights the pro-cyclical real wage would imply a pro-cyclical markup, which further aids the pro-cyclical movement in customer acquisition.

Corollary 1. *If $\hat{b}_t^*(z)$ and $\hat{f}_t^*(z)$ both go up following an expansionary monetary shock sufficiently, the firm would have pro-cyclical markup (counter-cyclical marginal cost) and counter-cyclical perceived elasticity of demand (i.e., pro-cyclical market power) and firm-level inflation is driven by the pro-cyclical movement of market power, holding inflation expectation constant, and assuming that the log deviation of the price-elasticity of intangible inputs are negligible.*

Internalizing pricing decision in a flexible price model. Another way to internalize the pricing decision is to solve the two-stage problem backwards in a flexible price model. In the second stage of this fully static firm's problem, for given optimal choices of intangible inputs, firms choose prices and quantities flexibly in a static Cournot game to maximize per-period profit:

$$\max_{P_{ijt},y_{ijt}} \left(P_{ijt} - (1-f_{ijt}^*)\frac{W_t}{z_{ij}}\right) y_{ijt}$$
$$\text{s.t. } P_{ijt} = \left(\frac{y_{ijt} b_{ijt}^{*\,\theta(\rho-1)}}{X_{jt}}\right)^{-\frac{1}{\rho}} \left(\frac{X_{jt}}{Y_t}\right)^{-\frac{1}{\eta}} \tilde{P}_t$$

The optimal pricing equation takes the familiar shape:

$$P_{ijt} = \mu_{ijt}(1-f_{ijt}^*)\frac{W_t}{z_{ij}}, \text{ where} \tag{89}$$

$$\mu_{ijt} = \frac{\Theta_{ijt}}{(\Theta_{ijt}-1)}, \Theta_{ijt} = \left[\frac{1}{\rho} + (\frac{1}{\eta} - \frac{1}{\rho})s_{ijt}\right]^{-1}, s_{ijt} = \frac{P_{ijt}y_{ijt}}{\sum_{l=1}^{n_j} P_{ljt}y_{ljt}} = \left(\frac{P_{ijt} b_{ijt}^{*\,\theta}}{\tilde{P}_{jt}}\right)^{1-\rho}$$

Solving the maximization problem leads to the static profits before intangible inputs costs:

$$\Phi_{ijt} = \left(1 - \frac{1}{\mu_{ijt}}\right) P_{ijt}y_{ijt} = \left(1 - \frac{1}{\mu_{ijt}}\right) s_{ijt}\tilde{P}_{jt}Y_t, \tag{90}$$

In the first stage, I denote the post-intangible inputs profits as $\Phi_{ijt}^{int} = \Phi_{ijt} - W_t F(f_{ijt}) - W_t F(b_{ijt})$. As with output choices, individual good firms internalize that their intangible input decisions affect sectoral price distributions through their effects on markups, market shares, and the sectoral ideal price. The optimal level of intangible inputs satisfies $\frac{\partial \Phi_{ijt}}{\partial f_{ijt}} = W_t F'(f_{ijt})$, and $\frac{\partial \Phi_{ijt}}{\partial b_{ijt}} = W_t F'(b_{ijt})$, respectively, whereby:

$$\frac{\partial \Phi_{ijt}}{\partial f_{ijt}} = s_{ijt}\tilde{P}_{jt}Y_t \frac{\partial\left(1-\frac{1}{\mu_{ijt}}\right)}{\partial f_{ijt}} + \left(1-\frac{1}{\mu_{ijt}}\right)\tilde{P}_{jt}Y_t\frac{\partial s_{ijt}}{\partial f_{ijt}} + \left(1-\frac{1}{\mu_{ijt}}\right) s_{ijt}Y_t \frac{\partial \tilde{P}_{jt}}{\partial f_{ijt}}, \tag{91}$$

and symmetric equation holds for optimal choice of b_{ijt}. It can then be shown that the first order conditions with respect to both intangible inputs are:

$$\psi_F(1-\phi_F)(1-f_{ijt}^*)^{-\psi_F} W_t = s_{ijt}\tilde{P}_{jt}Y_t \left[\left(\frac{1}{\eta}-\frac{1}{\rho}\right)(\rho-1)s_{ijt} + \left(1-\frac{1}{\mu_{ijt}}\right)(\rho-1-s_{ijt})\right]$$
$$\psi_B(1-\phi_B)(1-b_{ijt}^*)^{-(\psi_B+1)} W_t = \frac{s_{ijt}\tilde{P}_{jt}Y_t}{b_{ijt}^*}\left[\left(\frac{1}{\eta}-\frac{1}{\rho}\right)\theta(1-\rho)s_{ijt} + \left(1-\frac{1}{\mu_{ijt}}\right)\theta(1-\rho+s_{ijt})\right]$$

I then proceed to solve the full dynamic equilibrium as in the main text, with the relevant equations replaced. At the aggregate, as shown in Figure 9, the model is still capable of producing pro-cyclical markups (even though the magnitude is much smaller than the baseline case), and pro-cyclical real wage. Interestingly, the model features some Fisher effect when the initial response of price inflation is negative, before turning positive. Introducing nominal interest rate smoothing partially alleviates the effect.

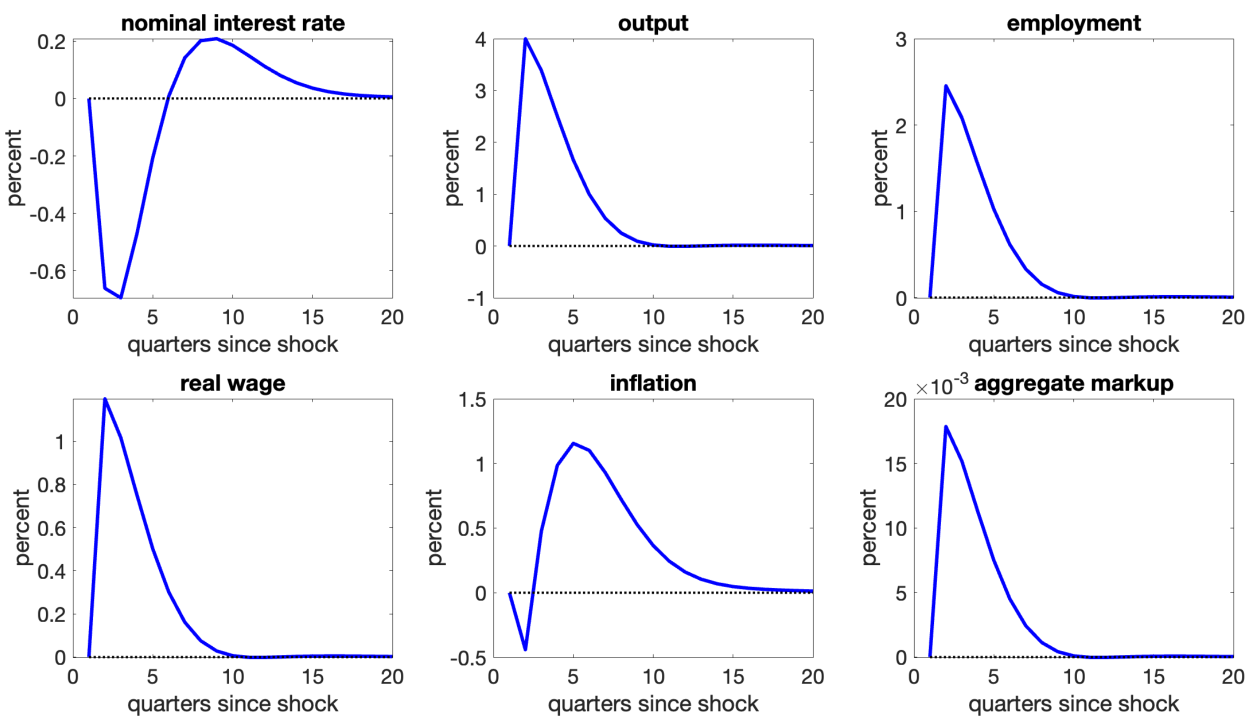


Figure 9: Model-implied aggregate IRFs to an expansionary monetary policy shock

Note: The figure plots the model-implied impulse responses of key aggregate variables to an expansionary monetary policy shock. The IRFs represent the percentage deviation of each variable from their respective steady state values.

E Additional Empirical Results

E.1 Aggregate-level Evidence

In order to benchmark the magnitudes of my empirical results in the main text against the empirical macroeconomics literature, I provide some evidence base on aggregate-level time series data. It would also be helpful to establish the conditional cyclicality of aggregate markups to interest rate changes as a benchmark, and to sense check that the empirical specification I adopt is producing appropriate impulse responses for standard macroeconomic outcome variables such as real GDP, unemployment, and credit spread, among others. To keep the specification as close to the panel regressions as possible, I estimate the following second stage regressions for horizon h using data from national statistics:

$$y_{t+h} - y_{t-1} = \alpha^h + \beta^h \Delta \hat{R}_t + \Gamma_l^h L(p).Y_t + \epsilon_{i,t+h}, \tag{92}$$

where the nominal interest ΔR_t is the change in the Fed Funds rate, and the instrumented $\Delta\hat{R}_t$ is based on GSS shocks summed to quarterly frequency. On the right hand side, I include four lags of macroeconomic controls in Y_t that include the Fed Fund rate, the monetary policy shocks, changes in log real GDP, credit spread, unemployment rate, and changes in log CPI and log commodity price index, and the standard errors are adjusted with the Newy-West method.

Figure 10 shows that a one percent decrease in the Fed Fund rate leads to an increase in real GDP of around 3% after two years, a fall in unemployment rate of 1% after two years and a half, and a fall in credit spread. Overall, these impulse responses are consistent with the notion that interest rate cuts expand economic activity.

Figure 11 looks at the movements of inflation (CPI), aggregate markup (log of markups estimated in a Cobb-Douglas production function from Nekarda and Ramey (2020)) , and unit labor cost (median real wage) following a one percent decrease in the Fed Fund rates. The three aggregate series are all conditionally pro-cyclical: CPI increases by approximately 1.5% after three years; markups increase by almost 4% at the peak one year after the initial shock; median real wage displays a persistent increase with a peak of 2% after two years.

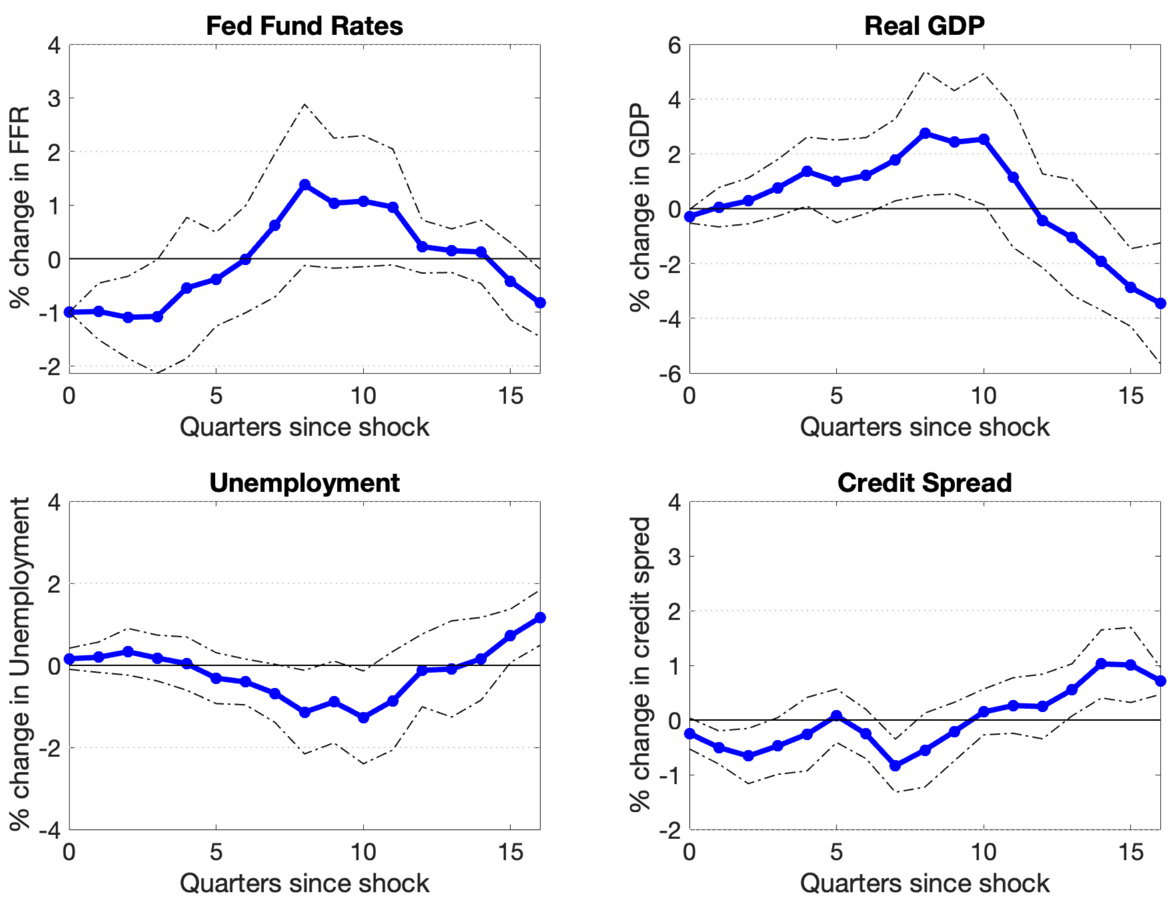


Figure 10: Effects of Monetary Policy Shocks on Aggregate Time-series Data (1)

Note: The figure shows the impulse responses of the Fed Funds rates, log real GDP, unemployment rate, and credit spread in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The sampling period is 1990Q1-2009Q4, and the forecast horizon is 16 quarters. Dotted lines plot the 90% standard error bands, with robust standard errors adjusted for serial correlation.

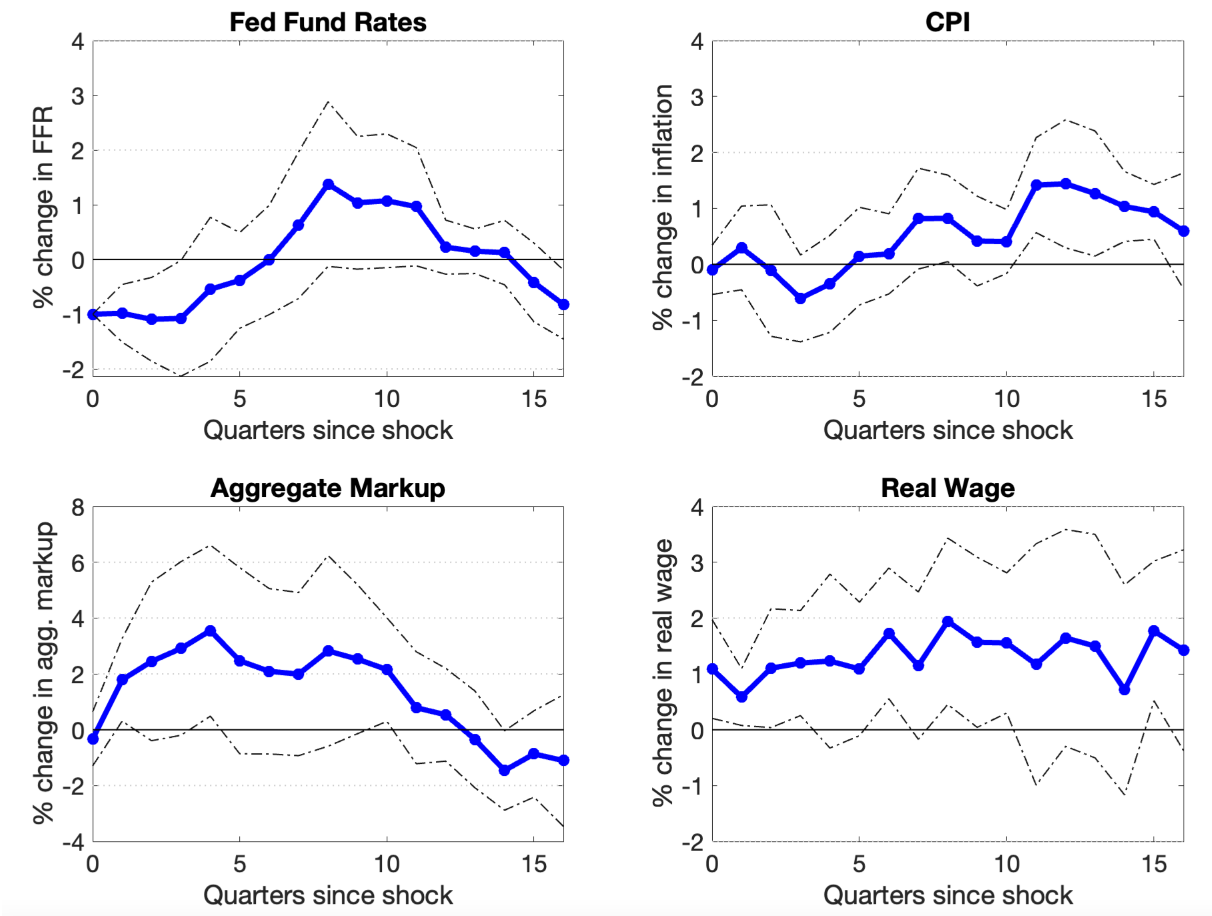


Figure 11: Effects of Monetary Policy Shocks on Aggregate Time-series Data (2)

Note: The figure shows the impulse responses of the Fed Funds rates, CPI, aggregate markup, and real wage in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The sampling period is 1990Q1-2009Q4, and the forecast horizon is 16 quarters. Dotted lines plot the 90% standard error bands, with robust standard errors adjusted for serial correlation.

Figure 12 and 13 further show the impulse responses of alternative measures of markups from Nekarda and Ramey (2013) following a one percent decrease in the Fed Fund rate. Contingent on details of empirical implementation, the peak effect is typically reached after one year and ranges between 2-4%. Out of the 12 markup series shown here, two display counter-cyclical responses, but neither has statistically significance at the 90% confidence interval. The overall picture seems to indicate that markups are conditionally pro-cyclical to monetary policy shocks.

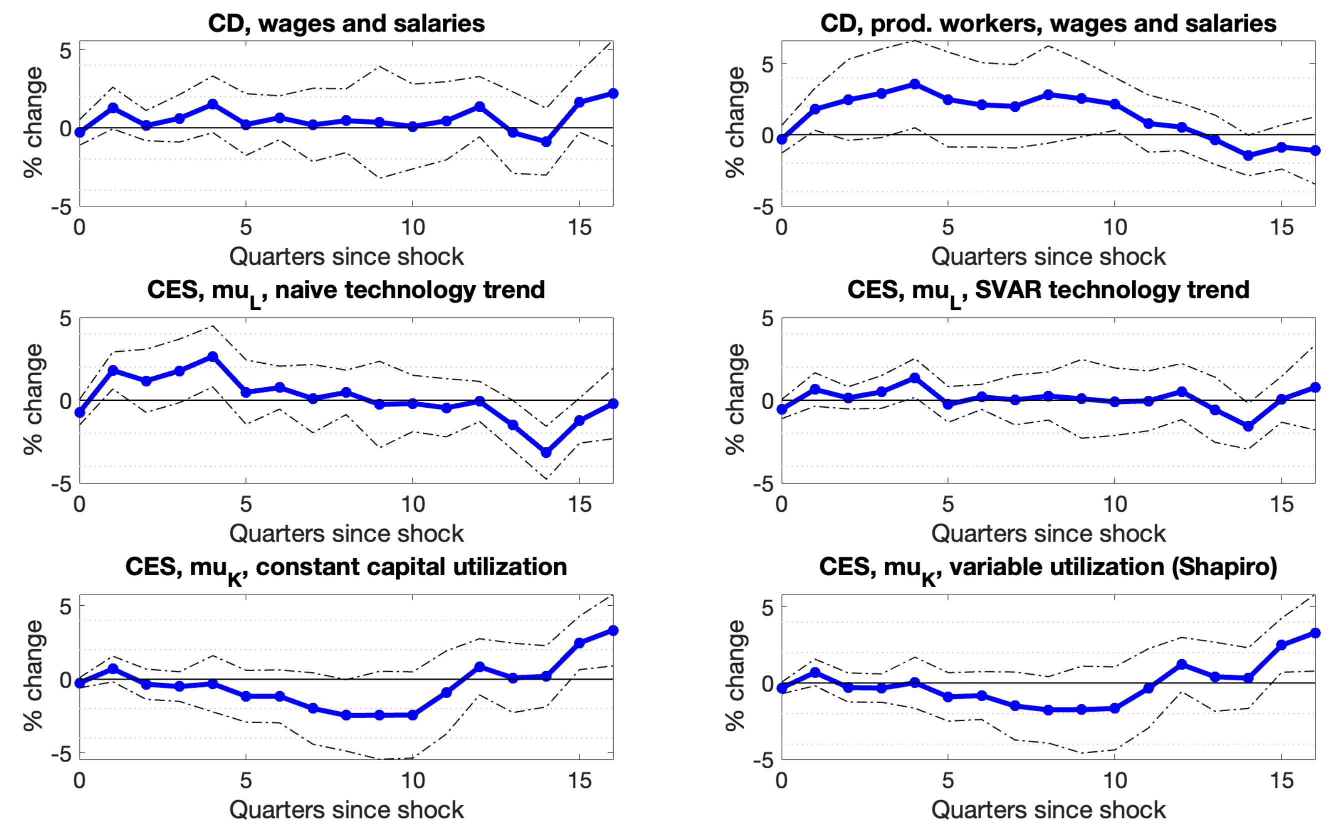


Figure 12: Effects of Monetary Policy Shocks on Aggregate Markups (1)

Note: The figure shows the impulse responses of various markup series from Nekarda and Ramey (2020) in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The sampling period is 1990Q1-2009Q4, and the forecast horizon is 16 quarters. Dotted lines plot the 90% standard error bands, with robust standard errors adjusted for serial correlation.

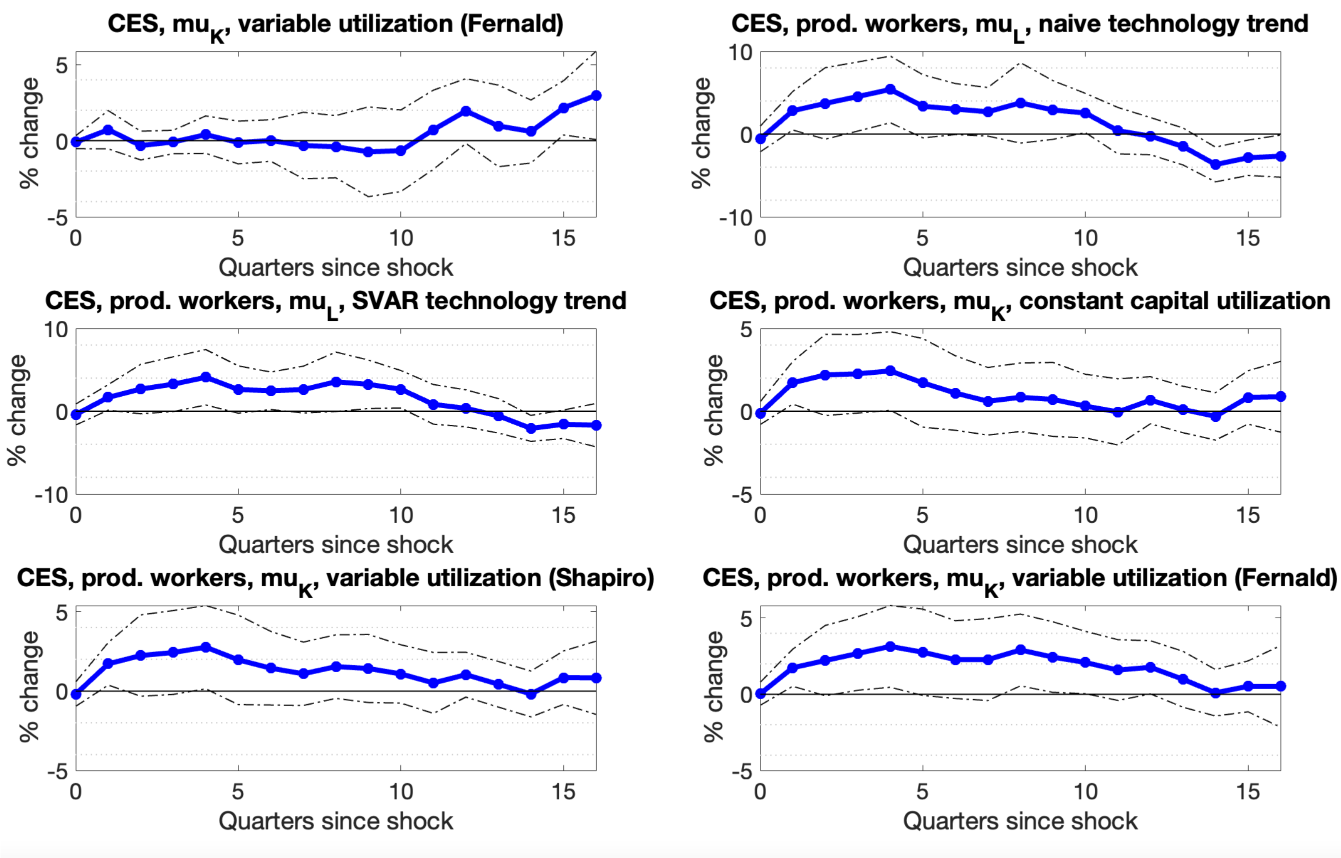


Figure 13: Effects of Monetary Policy Shocks on Aggregate Markups (2)

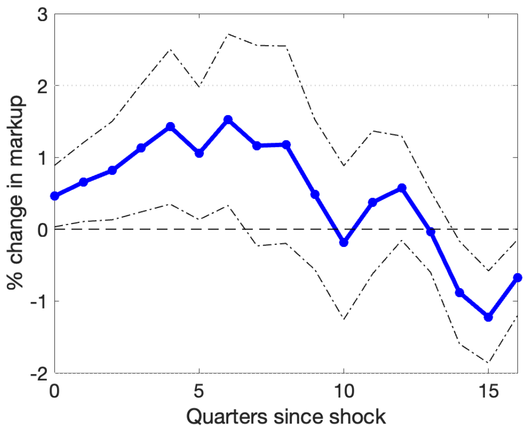
Note: The figure shows the impulse responses of various markup series from Nekarda and Ramey (2020) in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The sampling period is 1990Q1-2009Q4, and the forecast horizon is 16 quarters. Dotted lines plot the 90% standard error bands, with robust standard errors adjusted for serial correlation.

E.2 Firm-level Evidence

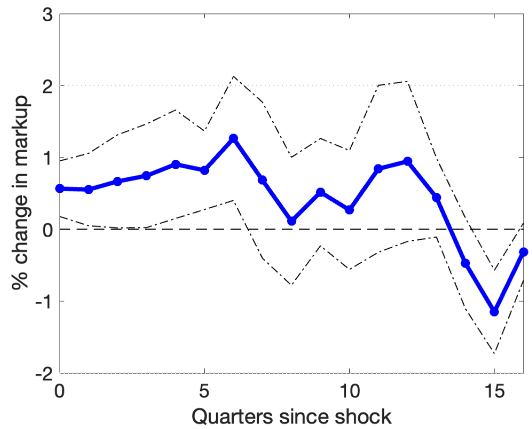
In this section, I provide additional empirical evidence based on the firm-level data. I first provide robustness checks for the pro-cyclical average response of markups based on alternative measures of markups and monetary policy shock; I then show that the cross-sectional heterogeneity identified in the main text holds for these alternative markups and monetary policy shock. Finally, I provide additional evidence supporting the proposed mechanism when I control for a third firm characteristic that correlates with market share/intangible intensity. Overall, how markups are estimated quantitatively affects the dynamic responses, but qualitatively, the big picture lends support to my baseline results.

E.2.1 Firm-level average responses

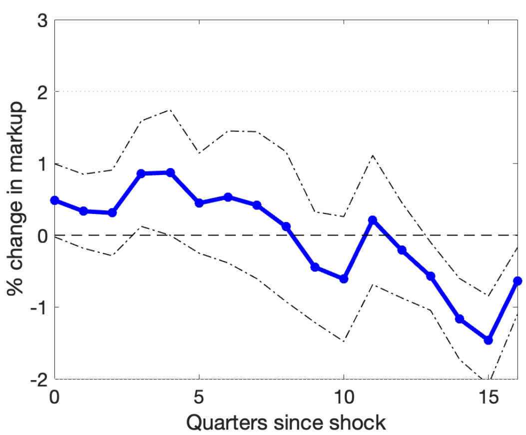
Alternative markups. In Figure 14, I first show that the average pro-cyclical response of markups to a monetary easing is robust to alternative measures of markups that include: (a) COGS-based markups in a 3-digit sectoral translog production function, (b) OPEX-based markups in a 3-digit sectoral Cobb-Douglas production function, (c) COGS-based markups in a 2-digit sectoral-time Cobb-Douglas production function, and (d) OPEX-based markup in a 2-digit sectoral-time Cobb-Douglas production function. With the exception of the last measure, the remaining measures of markups display a statistically significant pro-cyclical response. I show in the next section that deals with the cross-sectional heterogeneity that the volatile response displayed in Figure 14 (d) is driven by highly heterogeneous responses across firms.



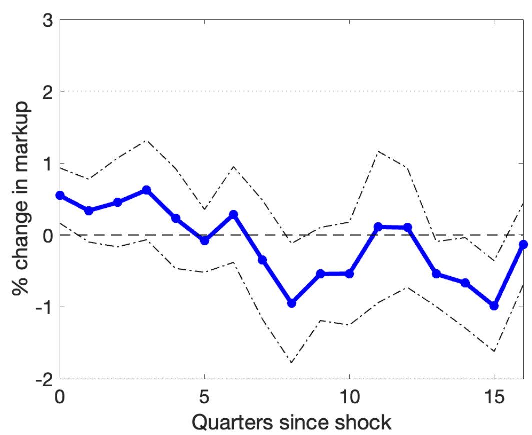
(a) COGS translog



(b) OPEX Cobb-Douglas



(c) COGS Cobb-Douglas with RW



(d) OPEX Cobb-Douglas with RW

Figure 14: Average Effects of Monetary Policy Shocks on Alternative Markups

Note: The figure shows the impulse responses of the alternative measure of markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The sampling period is 1990Q1-2009Q4, and the forecast horizon is 16 quarters. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

Alternative monetary policy shock. I also consider alternative identification of monetary policy shocks. In Figure 15, I plot the average impulse responses of the baseline measure of markups in an alternative LP-IV setup with the narratively-identified monetary policy shocks from Romer and Romer (2004) as the instrument in the first stage regressions to replace the high-frequency identified GSS shocks. The pro-cyclical average response is robust to this alternative identification method.

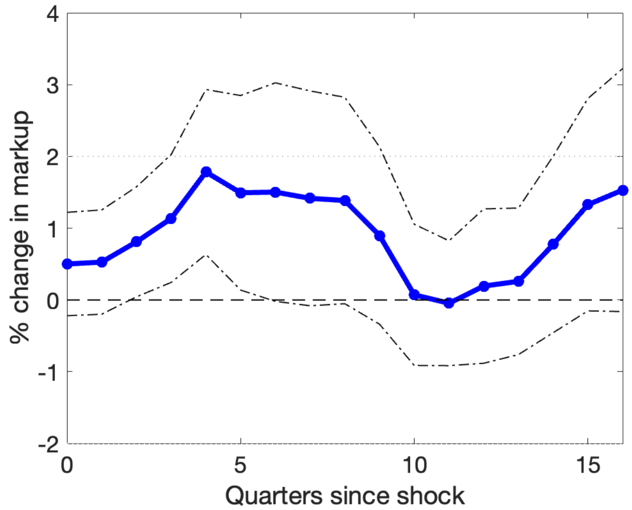


Figure 15: Average Effects of Alternative Monetary Policy Shocks on Markups

Note: The figure shows the impulse responses of the baseline measure of markups (estimated in a 3-digit sector-specific Cobb-Douglas production function based on COGS) in LP-IV regressions with narratively-identified shocks from Romer and Romer (2004) instrumenting for one percentage decrease in the Fed Funds rate. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

E.2.2 Motivational evidence: cross-sectional heterogeneity

Before turning to the cross-sectional results based on alternative measures of markups and monetary policy shocks, I first report the full set of group-specific impulse responses based on the baseline groups of low/medium/high intangible intensity. I then zoom in on the two groups of low intangible intensity and high intangible intensity as in the baseline case for alternative measures of markups and monetary policy shocks.

Full set of baseline results. In Figure 16 I show the full set of impulse responses of baseline measure of markups by intangible intensity. The distribution of intangible intensity is divided by terciles. The bottom two groups display IRFs that are broadly in line with each other, whereas the cross-sectional heterogeneity is most pronounced in the top tercile.

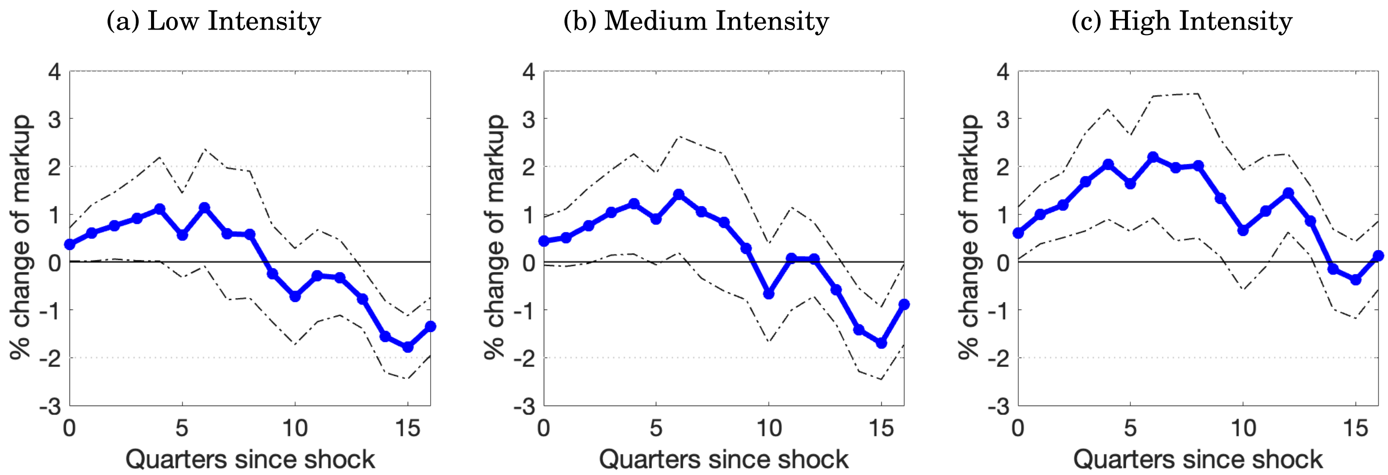


Figure 16: Heterogeneous Effects of Monetary Policy Shocks on Markups

Note: The figure shows the impulse responses of the baseline measure of markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The sampling period is 1990Q1-2009Q4, and the forecast horizon is 16 quarters. The interest rate is interacted with binary indicators based on distributions of firm market share and intangible intensity, in which low (high) intensity corresponds to firms with intangible intensity below the 33th percentile (above the 66th percentile) of its distribution in the previous quarter. Medium intensity corresponds to firms fall between the top and bottom terciles. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

Alternative markups. In Figures 17 - 20, I show that the cross-sectional heterogeneity is largely robust to alternative measures of markups that include: (a) COGS-based markups in a 3-digit sectoral translog production function, (b) OPEX-based markups in a 3-digit sectoral Cobb-Douglas production function, (c) COGS-based markups in a 2-digit sectoral-time Cobb-Douglas production function, and (d) OPEX-based markup in a 2-digit sectoral-time Cobb-Douglas production function. For expository purpose, I report the IRFs for the two groups based on low and high intangible intensity as in the main text.The overall picture is consistent with the baseline finding of heterogeneity in markup responses along the market share/intangible intensity dimensions.

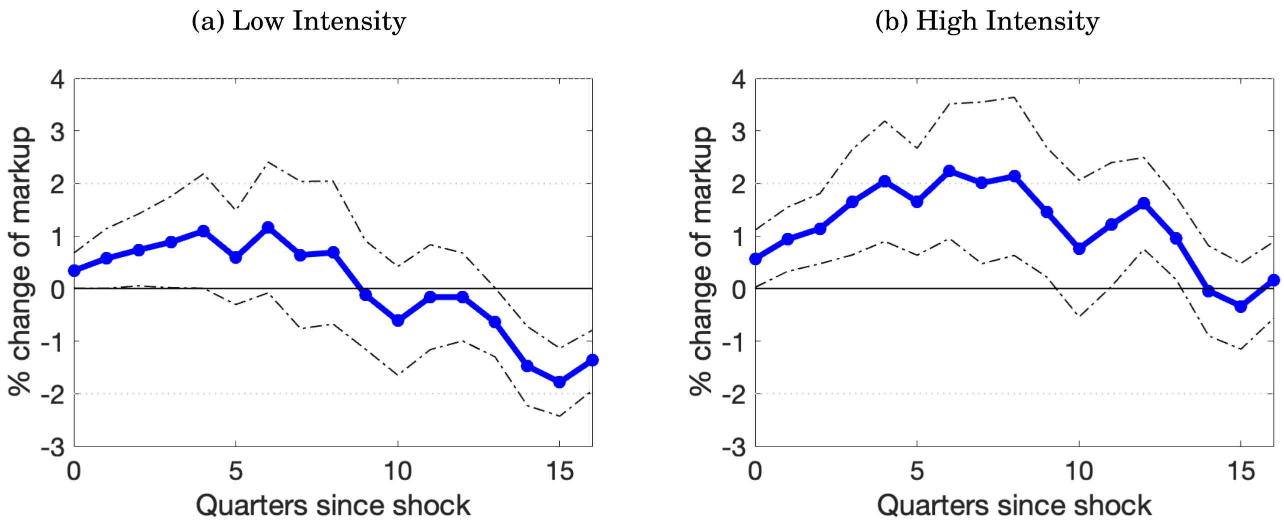


Figure 17: Heterogeneous Effects of COGS markups in translog production function

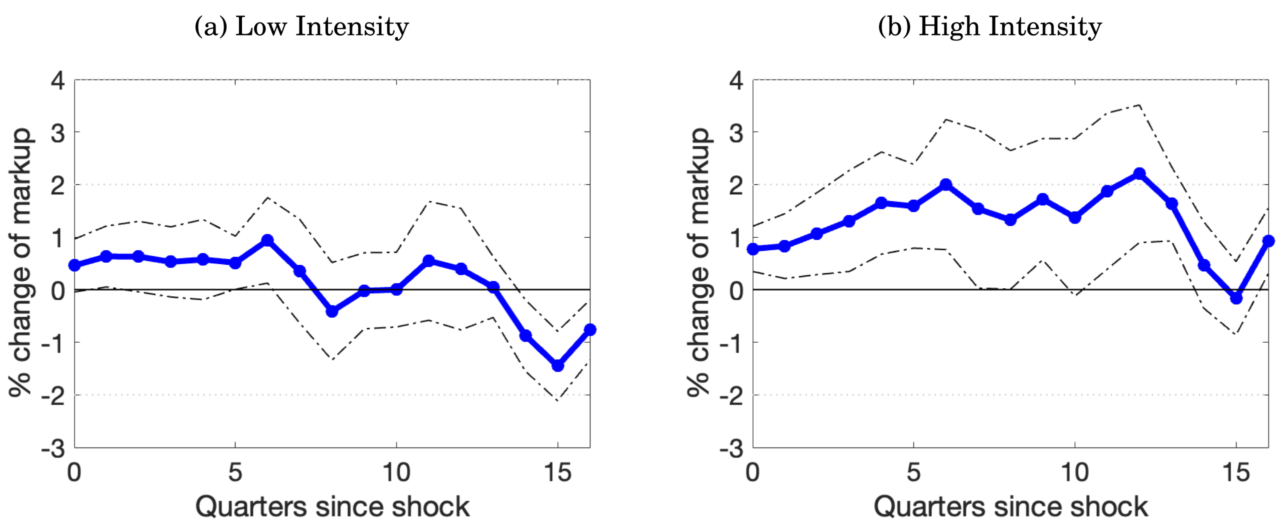


Figure 18: Heterogeneous Effects of OPEX markups in Cobb-Douglas production function

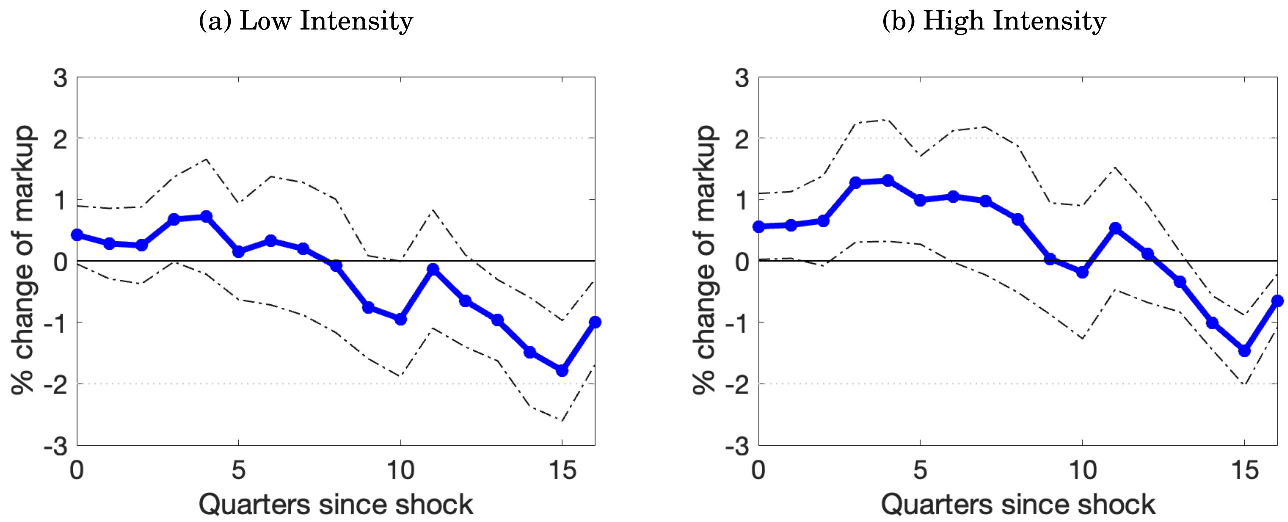


Figure 19: Heterogeneous Effects of COGS markups in Cobb-Douglas RW production function

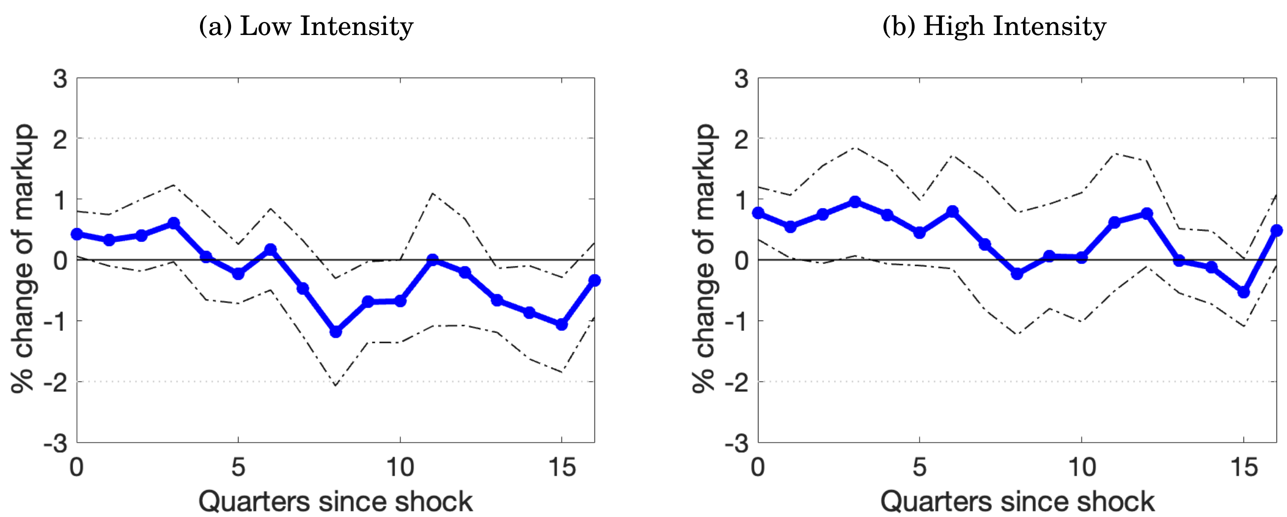


Figure 20: Heterogeneous Effects of OPEX markups in Cobb-Douglas RW production function

Note: These figure shows the impulse responses of different measure of markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with binary indicators based on distributions of intangible intensity as in the main text. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

Alternative monetary policy shock. I likewise show that the cross-sectional heterogeneity is robust to alternative identification of monetary policy shocks. Specifically, I use the Romer-Romer shocks to replace GSS shocks as the instrument in the second stage regression and report the findings for the two baseline groups in Figure 21.

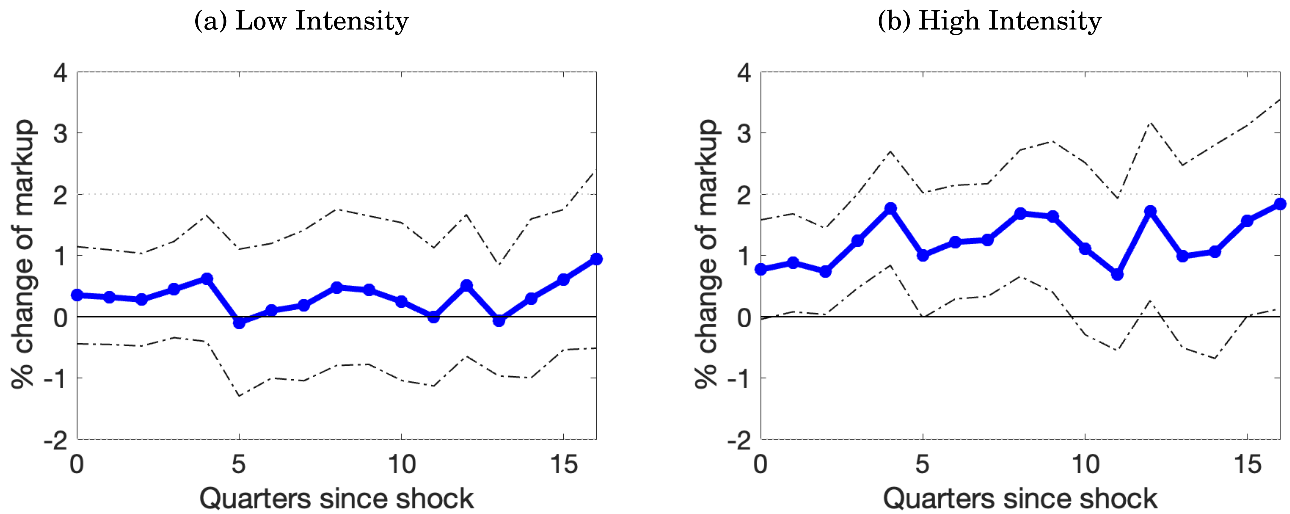


Figure 21: Heterogeneous Effects of COGS markups to Romer-Romer shocks

Note: These figure shows the impulse responses of baseline measure of markups in LP-IV regressions with Romer-Romer shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with binary indicators based on distributions of firm intangible intensity as in the main text. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

Linear interaction. Alternatively, in Equation 93, I impose a linear interaction such that a continuous measure of intangible intensity interact with interest rate changes to assess how do markups respond to monetary policy shocks depending on the firm's intangible intensity. In this end, I interact intangible intensity from the previous quarter $x_{i,t-1}$ with change in interest rates; β^h is the coefficient that captures the cross-sectional heterogeneity based on intangible intensity. Figure 22 shows that firms with high intangible intensity exhibit more pro-cyclical markups.

$$\Delta \log y_{i,t+h} = \alpha_i^h + \gamma^h \Delta \hat{R}_t + \beta^h x_{i,t-1} \Delta \hat{R}_t + \Gamma_l^h L(p).Y_t + \epsilon_{i,t+h} \tag{93}$$

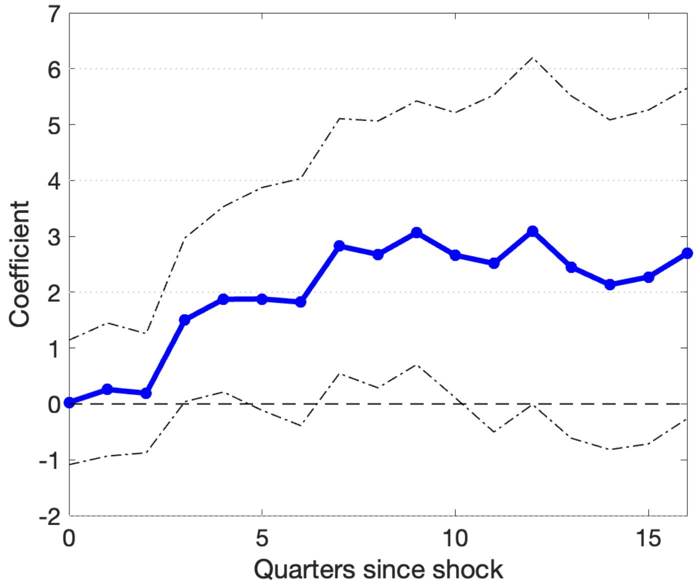


Figure 22: State-dependent Effects of Monetary Policy Shocks on Markups: Continuous Measure

Note: The figure shows the state-dependent impulse responses of the baseline measure of markups (estimated in a 3-digit sector-specific Cobb-Douglas production function based on COGS) in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The conditioning variable (interacting with interest rate change) is a continuous measure of firm-level intangible intensity. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

Alternative firm characteristics: financial frictions. I test if the financially constrained firms are showing more pro-cyclical markups as implied by the working capital assumption. In the presence of working capital, firms need to borrow to finance a fraction or all of their labor costs, such that an interest rate cut reduces the cost of borrowing, and if the benefit out-weights the increase in real wage, firms face a lower marginal cost and charge a pro-cyclical markups. To this end, I allow the effects of monetary policy to vary across the distributions of both firm-level intangible intensity and financial frictions such that firms fall into four groups based on high and low intangible intensity and high and low financial friction. I proxy for financial frictions with two common variables: leverage and liquidity, in which high leverage (low liquidity) indicates high possibility of being financially constrained. In Figure 23 and 24 I report the IRFs based on the four groups for intangible intensity/leverage and intangible intensity/ liquidity, respectively.

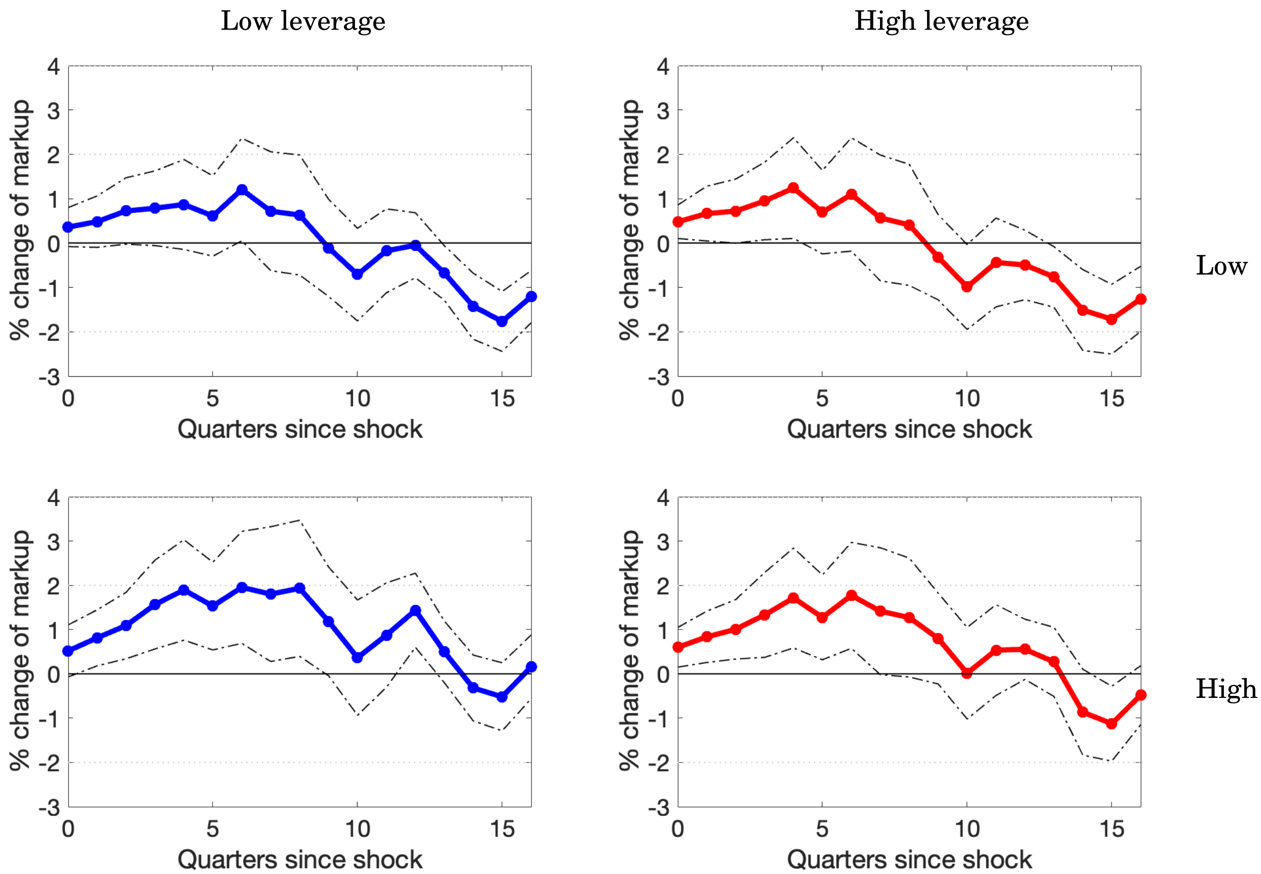


Figure 23: Heterogeneous Effects of Monetary Policy Shocks on Markups

Note: The figure shows the impulse responses of the baseline measure of markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with binary indicators based on distributions of firm-level leverage and intangible intensity as described in the main text. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

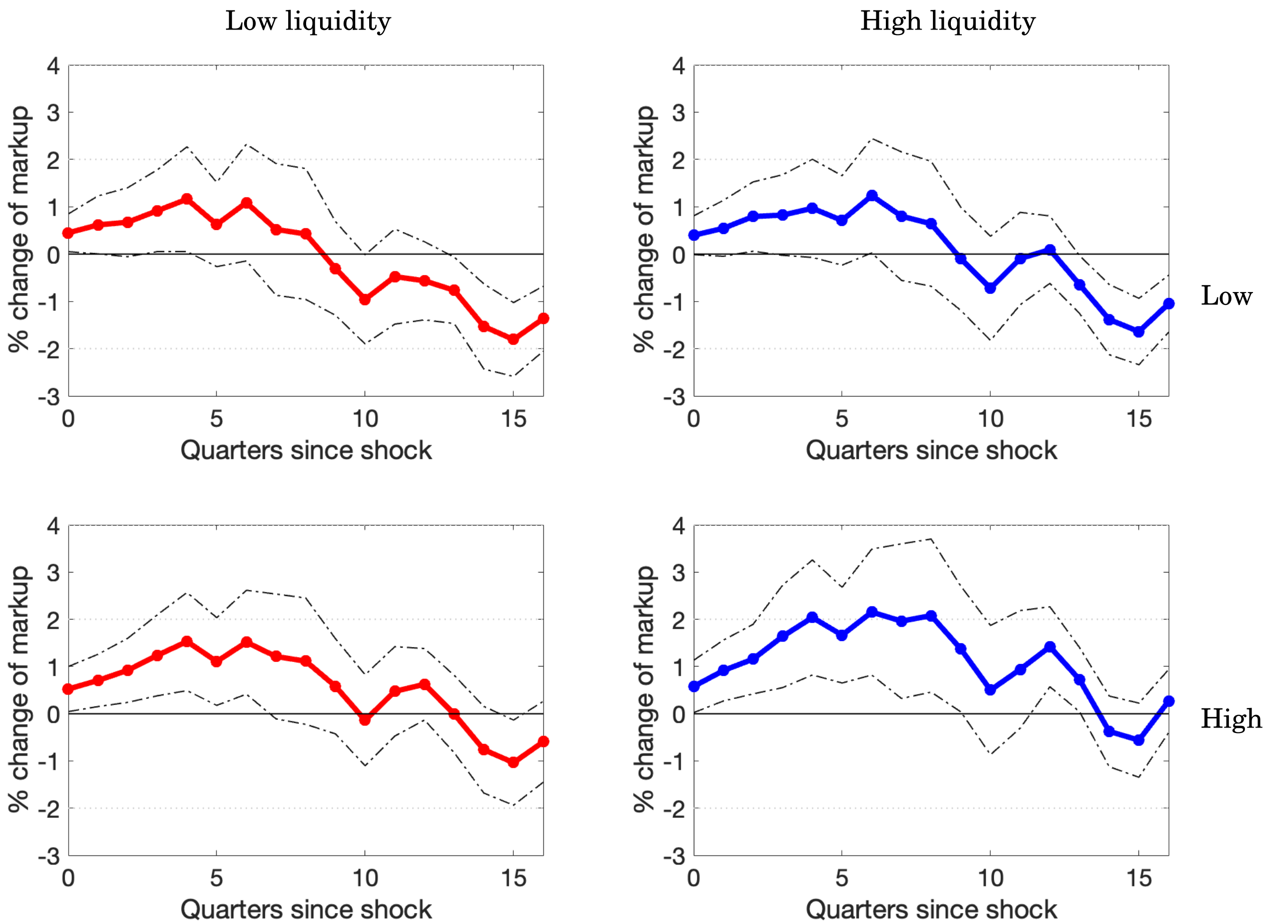


Figure 24: Heterogeneous Effects of Monetary Policy Shocks on Markups

Note: The figure shows the impulse responses of the baseline measure of markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with binary indicators based on distributions of firm-level liquidity and intangible intensity as described in the main text. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

I show that intangible intensity remains a robust predictor of markups' heterogeneous responses to a monetary easing, i.e., comparing cross rows, the bottom rows (high intangible intensity) display more pro-cyclical markups. Importantly, I do not find evidence that financially constrained firms (red IRFs) are showing more pro-cyclical markups. In fact, the results point to the opposite, conditioning on intangible intensity, i.e., comparing cross columns, financially unconstrained firms (those with low leverage or high liquidity, blue IRFs) are showing more pro-cyclical markups.

E.2.3 Mechanism testing: cross-sectional heterogeneity

Before turning to the cross-sectional results based on alternative measures of markups and monetary policy shocks, I first report the full set of group-specific impulse responses based on the baseline groups of larger (smaller) market share and high (low) intangible intensity. I then zoom in on the two groups of smaller and low intangible intensity and larger and high intangible intensity as in the baseline case for alternative measures of markups and monetary policy shocks. Additionally, to assess statistical significance of the heterogeneity, I report the relative response of larger and high intangible intensity firms to that of smaller and low intangible intensity. Finally, I consider a number of alternative firm characteristics in a triple-sorting of the data to check the marginal contribution of market share/intangible intensity conditioned on a relevant third firm characteristic that include R&D intensity, various measures of financial frictions, and exporter status. The estimates provide support, by and large, to the cross-sectional heterogeneity in markup responses along the lines of market share/intangible intensity.

Full set of group-specific results. In Figure 25 I show the full set of impulse responses of baseline measure of markups by market share and intangible intensity. Smaller (larger) refers to lower (higher) market share, whereas low (high) refers to lower (higher) intangible intensity. A comparison across rows (columns) for a given column (row) shows the responses based on intangible intensity (market share) for a given status of market share (intangible intensity). Most of the heterogeneity occurs along the intangible intensity dimension for this baseline measure of markups, and the effect is more pronounced at the top of the distribution for intangible intensity as shown in Figure 26, in which the cut-off for the distribution of intangible intensity is at the 70th percentile.

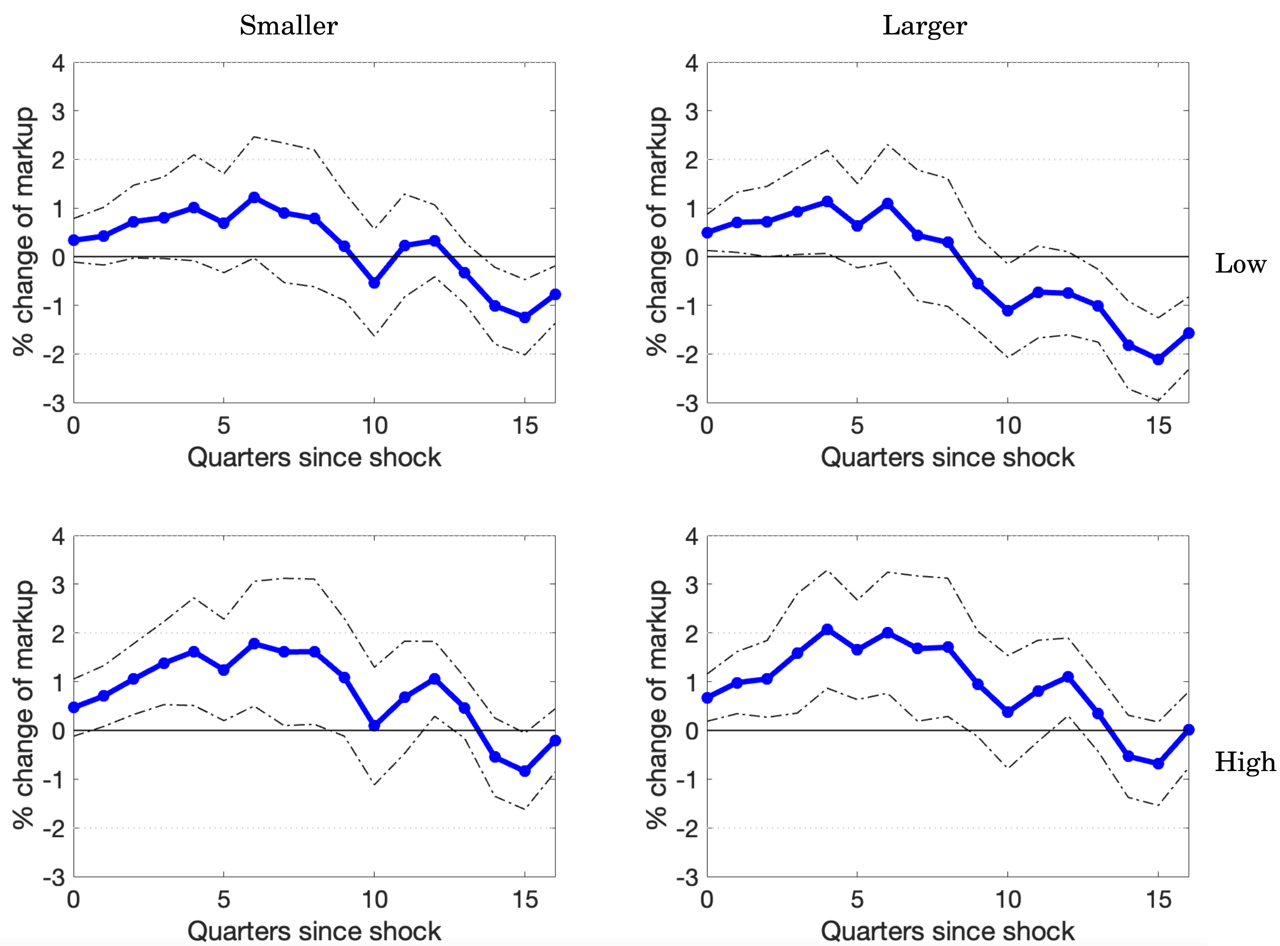


Figure 25: Heterogeneous Effects of Monetary Policy Shocks on Markups

Note: The figure shows the impulse responses of the baseline measure of markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with binary indicators based on distributions of firm market share and intangible intensity as described in the main text. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

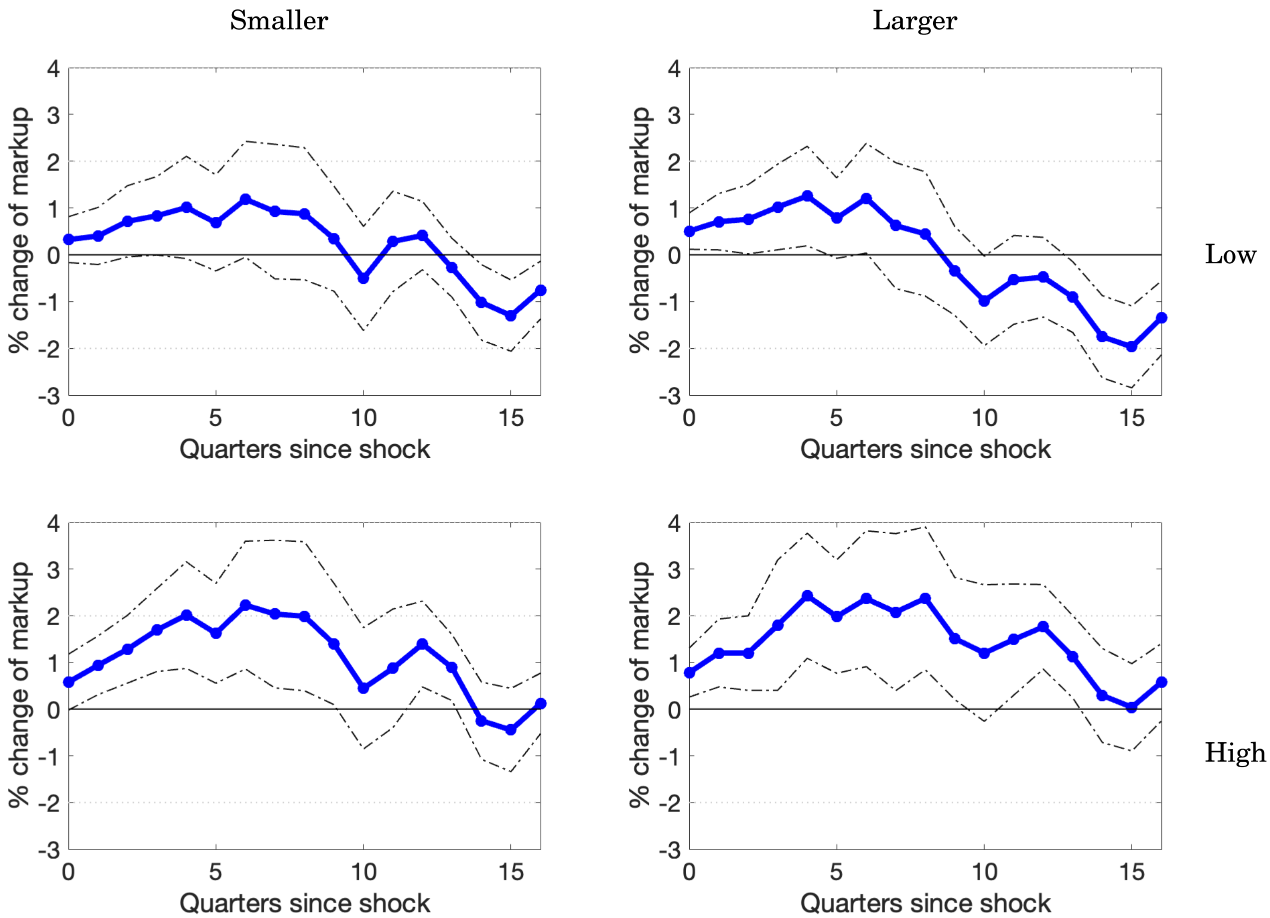


Figure 26: Heterogeneous Effects of Monetary Policy Shocks on Markups

Note: The figure shows the impulse responses of the baseline measure of markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with binary indicators based on the 70 percentiles of the distributions of intangible intensity, and the median of the distribution of market share. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

Alternative markups. In Figures 27 - 30, I show that the cross-sectional heterogeneity is largely robust to alternative measures of markups that include: (a) COGS-based markups in a 3-digit sectoral translog production function, (b) OPEX-based markups in a 3-digit sectoral Cobb-Douglas production function, (c) COGS-based markups in a 2-digit sectoral-time Cobb-Douglas production function, and (d) OPEX-based markup in a 2-digit sectoral-time Cobb-Douglas production function. For expository purpose, I report the IRFs for the two groups based on low market share/low intangible intensity and high market share/high intangible intensity as in the main text. In particular, In Figure 30, the group-specific IRFs for OPEX-based markups estimated in a series of rolling-window regressions are such that the smaller and low intensity firms display a statistically significant counter-cyclical responses, whereas the larger and high intensity firms show procyclical responses. These highly heterogeneous responses across groups of firms are the reason behind the volatile average response shown in Figure 14 (d). The overall picture is consistent with the baseline finding of heterogeneity in markup responses along the market share/intangible intensity dimensions.

(a) Smaller and Low Intensity (b) Larger and High Intensity

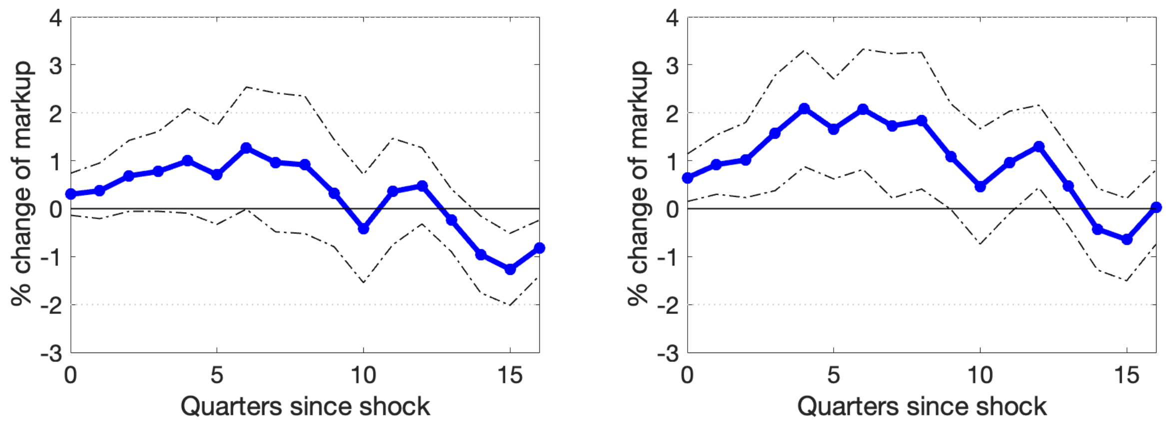


Figure 27: Heterogeneous Effects of COGS markups in translog production function

(a) Smaller and Low Intensity (b) Larger and High Intensity

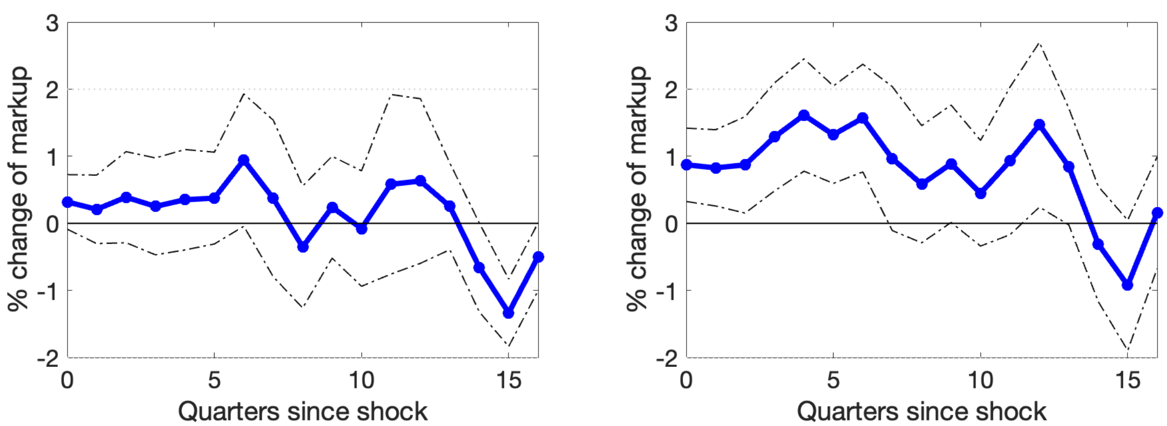


Figure 28: Heterogeneous Effects of OPEX markups in Cobb-Douglas production function

(a) Smaller and Low Intensity (b) Larger and High Intensity

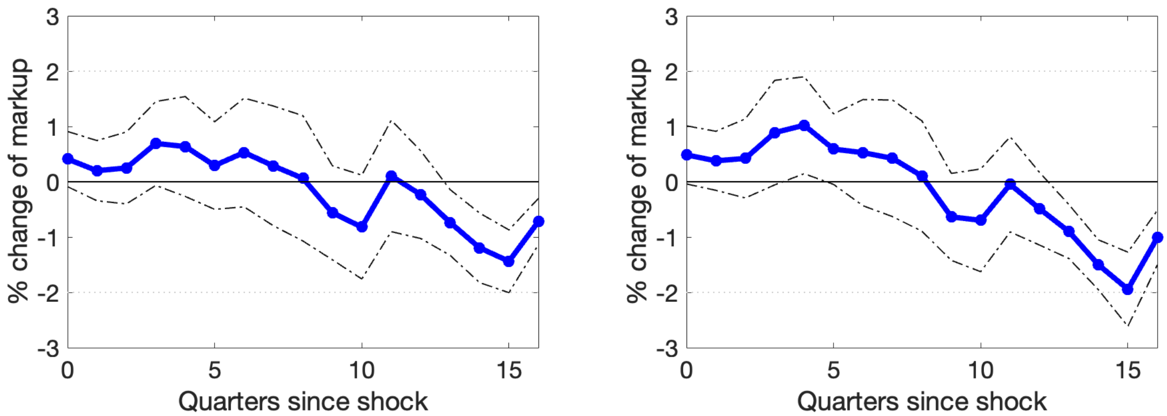


Figure 29: Heterogeneous Effects of COGS markups in Cobb-Douglas RW production function

(a) Smaller and Low Intensity (b) Larger and High Intensity

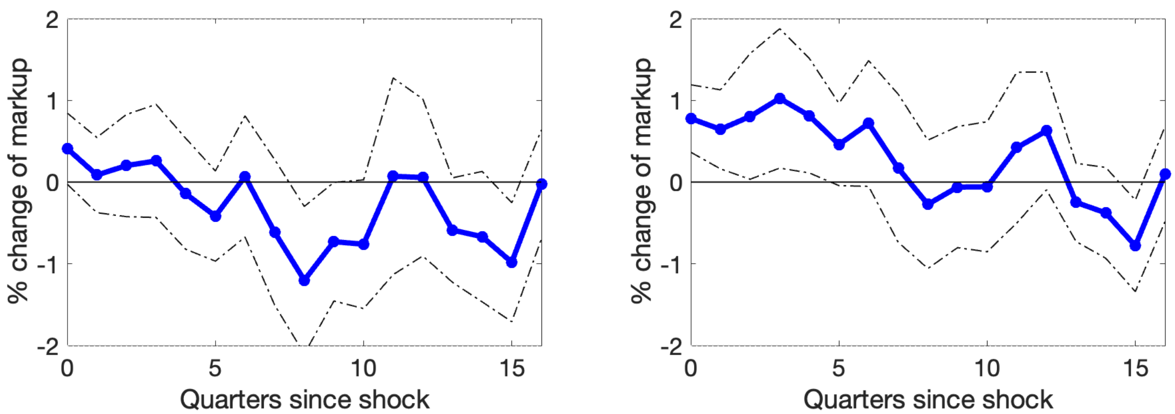


Figure 30: Heterogeneous Effects of OPEX markups in Cobb-Douglas RW production function

Note: These figure shows the impulse responses of different measure of markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with binary indicators based on distributions of firm market share and intangible intensity as in the main text. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

Alternative monetary policy shock. I likewise show that the cross-sectional heterogeneity is robust to alternative identification of monetary policy shocks. Specifically, I use the Romer-Romer shocks to replace GSS shocks as the instrument in the second stage regression and report the findings for the two baseline groups in Figure 31.

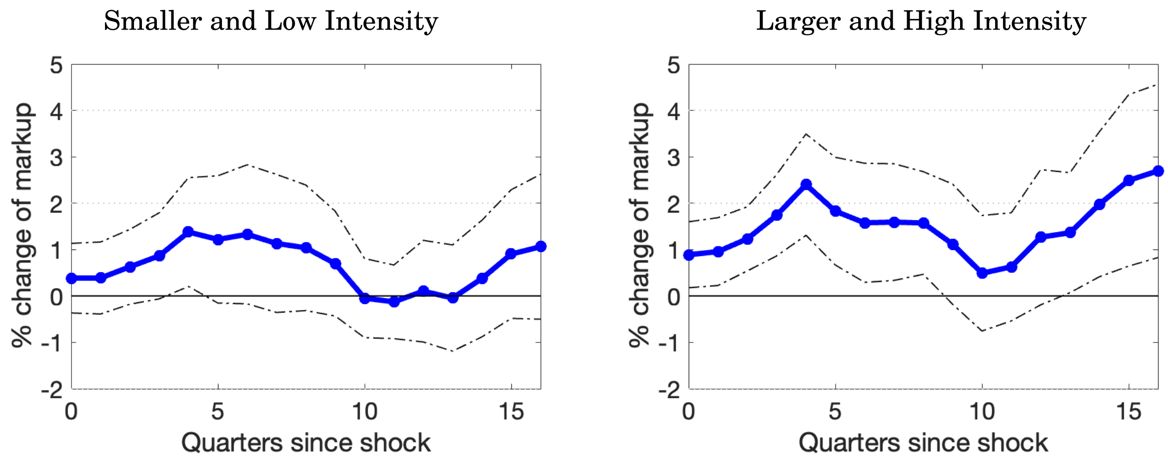
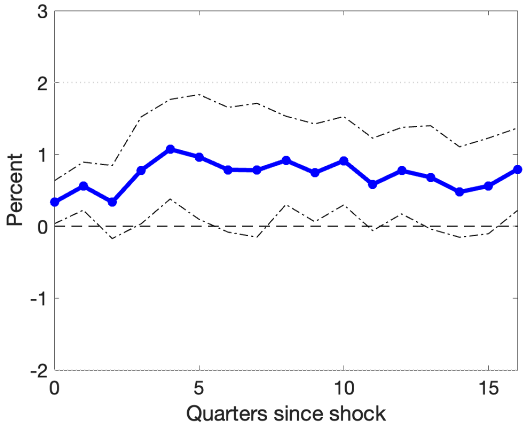


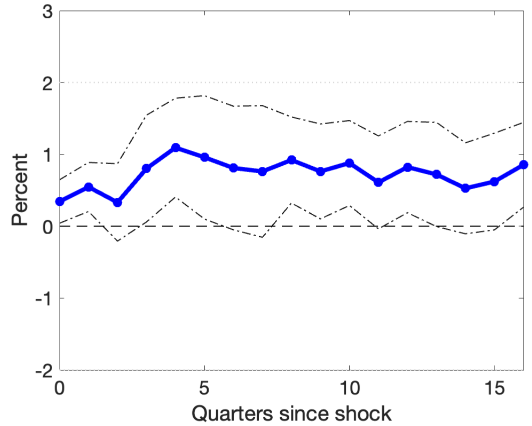
Figure 31: Heterogeneous Effects of COGS markups to Romer-Romer shocks

Note: These figure shows the impulse responses of baseline measure of markups in LP-IV regressions with Romer-Romer shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with binary indicators based on distributions of firm market share and intangible intensity as in the main text. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

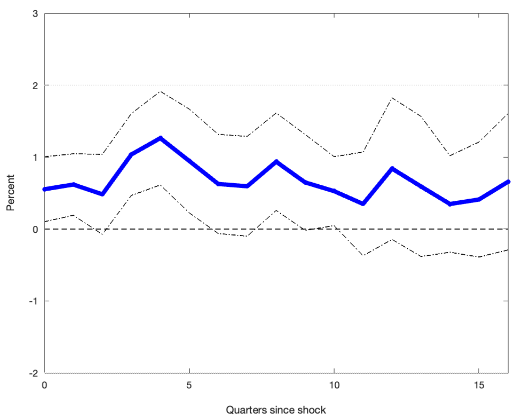
Relative effect. An alternative way to examine the cross-sectional heterogeneity is to assess the relative differences of markups' response with respect to the least responsive base group. This method allows for a formal assessment of the statistical differences between the impulse responses across groups at each horizon. Formally, I re-run the second stage baseline regression with the Fed Funds rate (instrumented by the GSS shocks) added as an additional regressor and designate the low market share/low intangible intensity firms as the base group. In Figure 32, I show that the markup response of high market share/high intangible intensity firms are significantly greater than that of the base group across a variety of markups; the relative effect remains robust when the Romer-Romer shocks are used in place of GSS shocks.



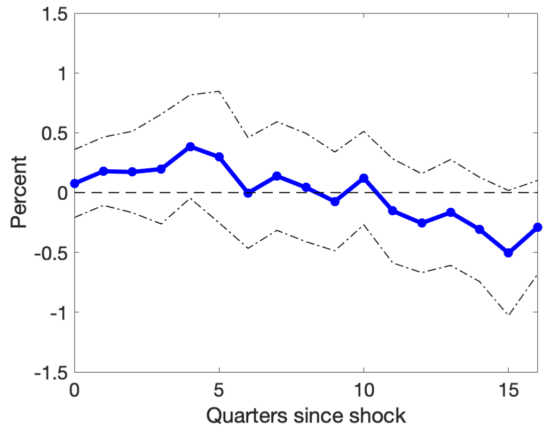
(a) COGS Cobb-Douglas



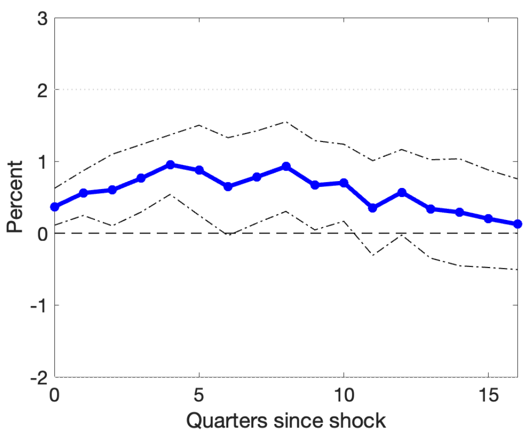
(b) COGS translog



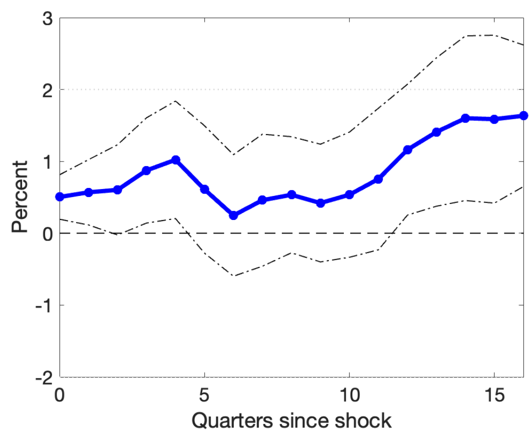
(c) OPEX Cobb-Douglas



(d) COGS Cobb-Douglas with RW



(e) OPEX Cobb-Douglas with RW



(f) COGS Cobb-Douglas and RR shocks

Figure 32: Relative Effects of Markups by High Market Share/High Intangible intensity

Note: The figure shows the impulse responses for different measure of markups in LP-IV regressions with different monetary shocks instrumenting for one percentage decrease in the Fed Funds rate. The focus is on the firms in the high market share/high intangible intensity group as defined in the main text. The specification includes the interest rate as a regressor in the second stage regression, and the IRFs are relative to the base group of low market share/low intangible intensity. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

Conditioning on R&D intensity. I first consider sectoral R&D intensity as an alternative firm characteristic and report the IRFs for all eight groups. In Figure 33, I plot the IRFs for the base groups conditioning on high R&D intensity, whereas in Figure 34, each group's responses conditioning on low R&D intensity is displayed. The heterogeneity between low market share/low intensity firms and high market share/high intensity firms remain robustness regardless of R&D intensity. At the same time, I find that all four groups of firms in high R&D sectors have more pro-cyclical markups than their counterparts in low R&D sectors, even though for some group, the heterogeneity is negligible. The high market share/high intensity firms in high R&D sectors are displaying the most pro-cyclical response across all eight groups. I interpret the finding as suggesting that pro-cyclical R&D may serve as a missing mechanism behind the pro-cyclical average response.

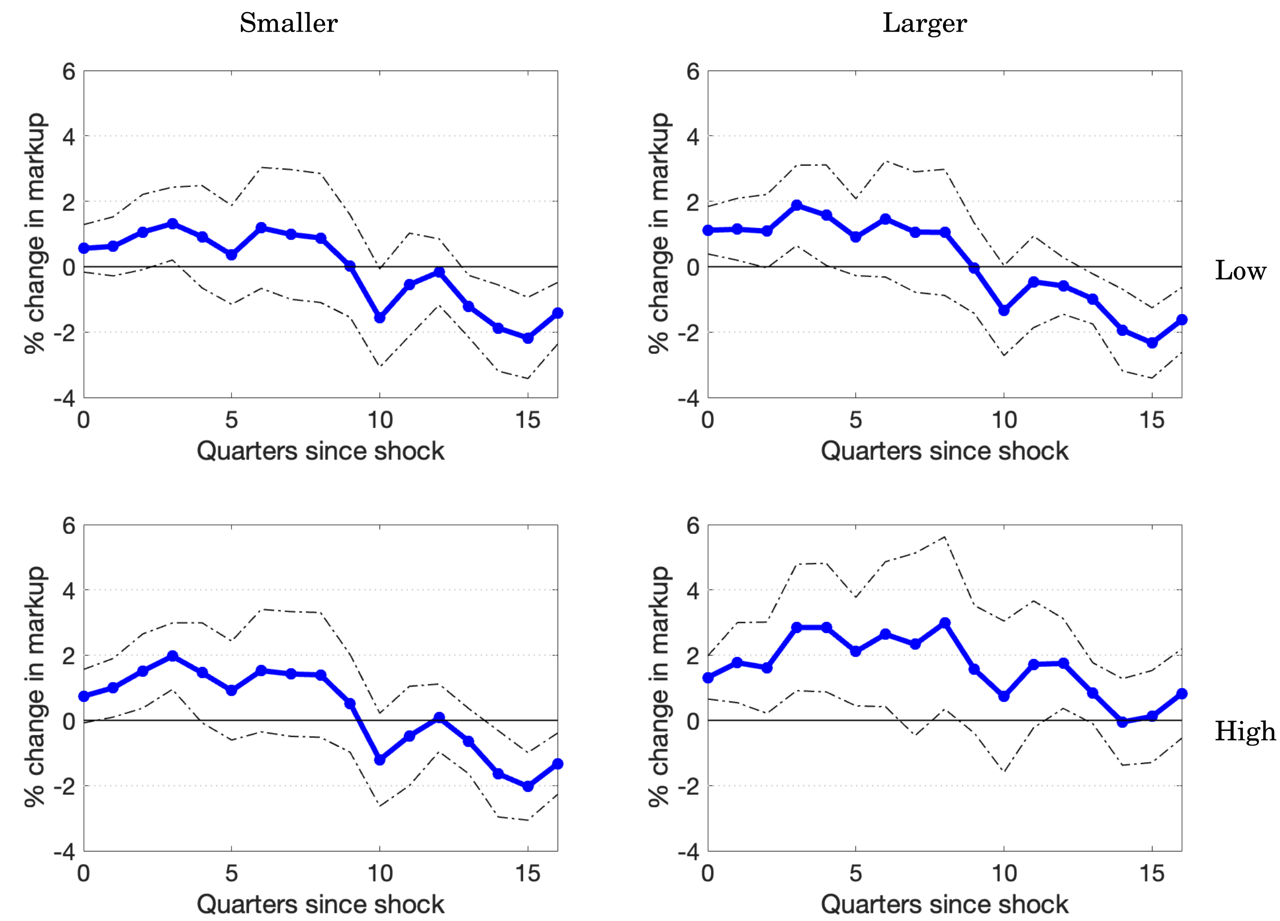


Figure 33: Heterogeneous Effects of Markups condition on High R&D Intensity

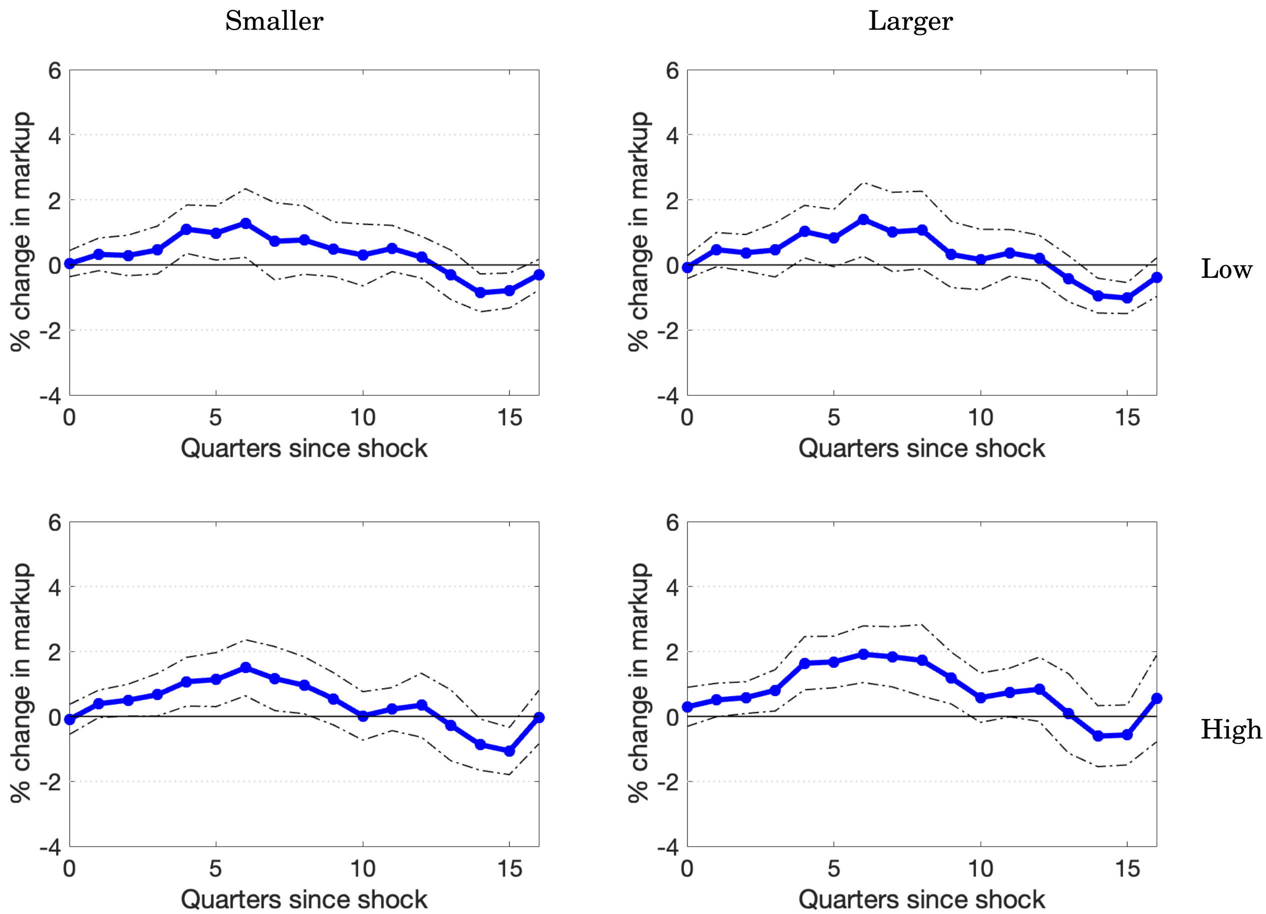


Figure 34: Heterogeneous Effects of Markups condition on Low R&D Intensity

Note: The figure shows the impulse responses of the baseline measure of markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with binary indicators based on distributions of firm market share, intangible intensity, and sectoral-level R&D intensity as described in the main text. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

Conditioning on financial frictions. For alternative conditioning variables that are well-populated, I use the firm-level distribution from the previous quarter. It is plausible that financial frictions may interact with SG&A spending such that the responses of markups are driven by those firms that are financially constrained. Moreover, firm size (measured by total asset) is a common proxy for financial friction that could be highly correlated with firm market share. I consider various common proxies for financial frictions in the literature that include size proxied by total asset (Gertler and Gilchrist, 1994; Crouzet and Mehrotra, 2020), leverage (Ottonello and Winberry, 2017), and liquidity (Jeenas, 2019), and interact the medians of respective distribution with the baseline groups. Taking size as an example, in Figure 35, I compare the markup responses of low market share/low intangible intensity/small firms to those of high market share/high intangible intensity/small firms. That is, by holding the additional conditioning variable fixed, it allows me to infer the marginal contribution of market share and intangible intensity for a given size of firms. Among small firms that are likely to be financially constrained, the high market share/high intangible intensity firms are exhibiting the most pro-cyclical markups, whereas the markups of the low market share/low intangible intensity barely respond at all. When conditioning on being large in the absolute sense (i.e., those with high total assets), Figure 36 displays largely similar results. I also show that similar results are obtained for alternative proxies for financial friction in Figure 37 for high leverage and in Figure 38 for low liquidity. These additional empirical results collectively suggest that once market share and intangible intensity are taken into account, there is little differences in markup responses across firms based on measures of financial frictions.

(a) Smaller and Low Intensity (b) Larger and High Intensity

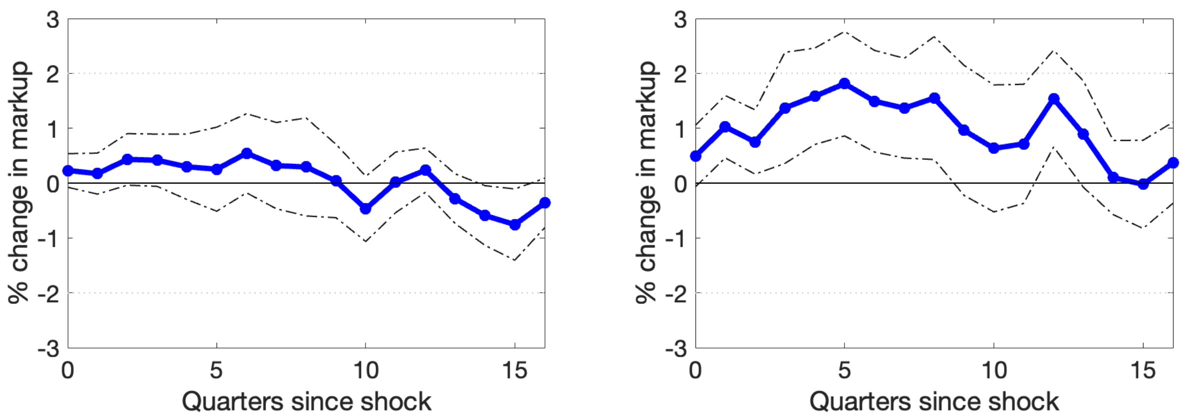


Figure 35: Heterogeneous Effects of Monetary Policy Shocks on Markups condition on Low Total Asset

Note: The figure shows the impulse responses of the baseline markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with three binary indicators based on distributions of firm market share, intangible intensity, and total assets as described in the main text. Low total asset refers to firms with total asset below the median of its distribution in the previous quarter. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

(a) Smaller and Low Intensity (b) Larger and High Intensity

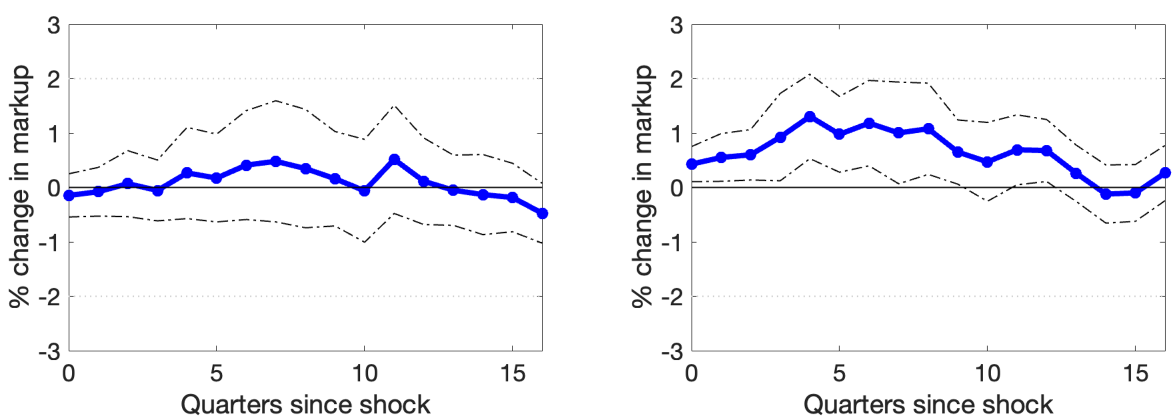


Figure 36: Heterogeneous Effects of Monetary Policy Shocks on Markups condition on High Total Asset

Note: The figure shows the impulse responses of the baseline markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with three binary indicators based on distributions of firm market share, intangible intensity, and total assets as described in the main text. High total asset refers to firms with total asset above the median of its distribution in the previous quarter. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

(a) Smaller and Low Intensity (b) Larger and High Intensity

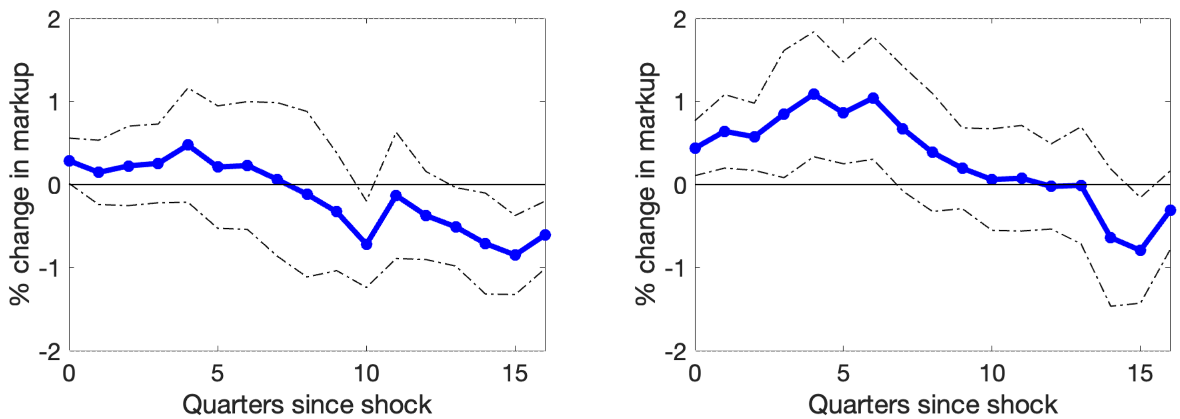


Figure 37: Heterogeneous Effects of Monetary Policy Shocks on Markups condition on High Leverage Ratio

Note: The figure shows the impulse responses of the baseline markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with three binary indicators based on distributions of firm market share, intangible intensity, and total assets as described in the main text. High leverage ratio refers to firms with leverage above the median of its distribution in the previous quarter. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

(a) Smaller and Low Intensity (b) Larger and High Intensity

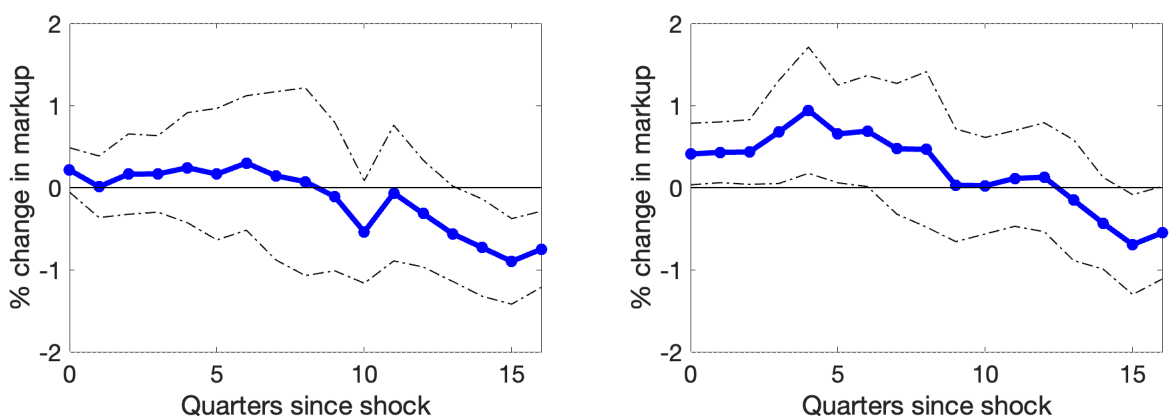


Figure 38: Heterogeneous Effects of Monetary Policy Shocks on Markups condition on Low Liquidity

Note: The figure shows the impulse responses of the baseline markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with three binary indicators based on distributions of firm market share, intangible intensity, and total assets as described in the main text. Low liquidity refers to firms with liquidity the median of its distribution in the previous quarter. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

Conditioning on exporter status. The third characteristic is a firm's exporter status. Figure 39 shows that market share/intangible intensity remains a robust predictor of cross-sectional heterogeneity for markups among multinational firms in Compustat. Similar results are obtained for firms that are designated as domestic in Figure 40.

(a) Smaller and Low Intensity (b) Larger and High Intensity

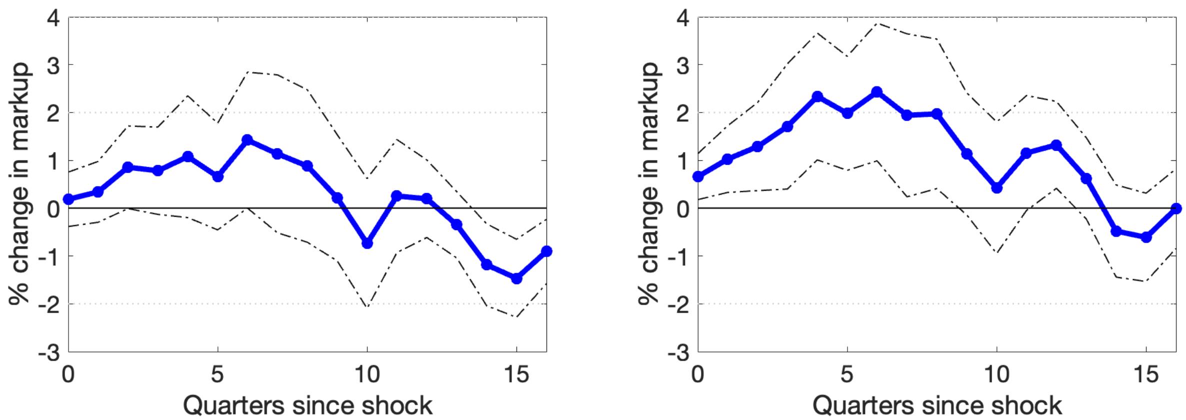


Figure 39: Heterogeneous Effects of Markups condition on Exporter status

Note: The figure shows the impulse responses of the baseline markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with three binary indicators based on distributions of firm market share, intangible intensity, and exporter status as described in the main text. Exporter refers to firms that are designated as exporter. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.

(a) Smaller and Low Intensity (b) Larger and High Intensity

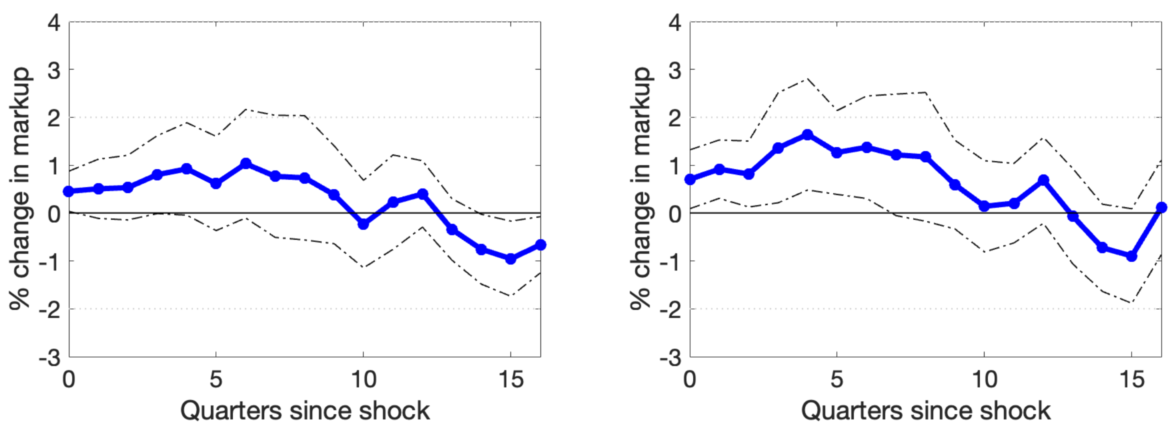


Figure 40: Heterogeneous Effects of Markups condition on non-Exporter status

Note: The figure shows the impulse responses of the baseline markups in LP-IV regressions with GSS shocks instrumenting for one percentage decrease in the Fed Funds rate. The interest rate is interacted with three binary indicators based on distributions of firm market share, intangible intensity, and exporter status as described in the main text. Non-Exporter refers to firms that are not designated as exporter. Dotted lines plot the 90% standard error bands, with standard errors clustered at firm-time level and adjusted using the Driscoll-Kraay method.