Monthly Heterogeneous Markup Cyclicality and Inflation Dynamics

Mario Giarda* Will Lu† Antonio Martner‡

March 10, 2024

Abstract

We study the monthly cyclicality of the markup distribution using tax records for Chile. We find that markups (i) are unconditionally countercyclical; (ii) the countercyclicality is heterogeneous along the markups' distribution; (iii) markups increase in response to a contractionary monetary policy shock, and (iv) firms with higher average markups have more responsive markups to monetary surprises. We calibrate a model of firm heterogeneity with Kimball demands and show that markup heterogeneity affects inflation dynamics. We show that strategic complementarities flatten the Phillips curve through the correlation between market shares and pass-through from marginal costs to prices.

*Central Bank of Chile. Email: mgiarda@bcentral.cl
†Central Bank of Chile. Email: wlu@bcentral.cl
‡UCLA and Central Bank of Chile. Email: amartner@ucla.edu

1 Introduction

The study of inflation is intricately linked to price adjustment processes, often represented in models with price rigidities through Calvo adjustment prices or menu costs. However, most of the existing models overlook firm heterogeneity on price adjustments, primarily due to limitations in data availability. By omitting firm heterogeneity, models potentially disregard how the firm distribution responds to monetary policy shocks and its influence on aggregate inflation dynamics.

This paper aims to address this gap by investigating how the heterogeneity of firms, specifically their product market power, shapes the response of aggregate inflation to monetary policy. We delve into the cyclical nature of the markup distribution and its implications for aggregate inflation dynamics. We assume firms set a markup over marginal cost. Leveraging on a granular database for Chile, including prices and quantities at the firm level, we estimate markups under the production approach.

We study markup fluctuations because they shape the New Keynesian Phillips curve, influencing inflation dynamics. Markups, within the canonical New Keynesian (NK, henceforth) framework, exhibit variability due to nominal rigidities and serve as a sufficient statistic for the response of prices to marginal cost fluctuations (see Woodford, 2003 and Galí, 2015). When a shock to marginal costs occurs, which cannot be fully absorbed through price adjustments, markup fluctuations follow. Typically, NK models suggest that markups are countercyclical in response to demand shocks. This is because, in the presence of demand shocks, all prices increase, leading to a rise in marginal costs. Consequently, firms unable to adjust prices must reduce their markup. Thus, studying the fluctuations of markups over the business cycle and in response to demand shocks is essential for understanding NK models and addressing the empirical fit of models with firm heterogeneity for the study of aggregate inflation. Moreover, it offers an alternative avenue for empirically investigating Phillips curves, both at an aggregate level and conditional on the heterogeneity of markup fluctuations.

While there is enough evidence on the presence of price rigidities at sectoral levels (see Nakamura and Steinsson, 2008 among others), there is little evidence on how empirical markups fluctuate over the business cycle and their responses to aggregate shocks at a firm-level. Exceptions to these are the papers that study the cyclicality of markups unconditionally (see Anderson et al., 2018 and Burstein et al., 2020, among others) and in response to monetary policy shocks with quarterly frequency data (see Chiavari et al., 2021).

In this paper, empirically, we revisit the cyclicality of firm-level markups using administrative microdata of the universe of formal firms in Chile. Since firms in Chile must fill out tax forms for their purchases and income every month, we are able to estimate a firm-level markup at a monthly frequency. To compute markups, we follow De Loecker and Warzynski (2012) using price variation-free measures for output, as we can observe firm-level prices. We estimate production functions of quantities instead of value added. The main advantage of using quantities produced is that materials markups can be recovered, which is a more flexible input than labor (the other input to retrieve markups). A more flexible production input will better identify markups as potential frictions that might generate a wedge between the marginal product and the input price are less likely to occur[1].

We estimate monthly production functions to recover monthly markups. While most of the literature relies on yearly–or quarterly– markups, the granularity and frequency of our dataset allow us to estimate high-frequency markups. To ensure identification, we assume that the firm's intermediate input decisions are monthly (rather than yearly).

Using our production functions estimations, we study the cyclicality of the firm-level markup and their monthly responses to monetary policy shocks, using monetary policy surprises in Chile computed by Aruoba et al. (2021). Also, since we have the universe of firms in Chile, we study the cyclicality of the markup distribution, which results in a granular view of markup fluctuations. We find that markups are unconditionally countercyclical; the countercyclicality is heterogeneous along the markups' distribution; they increase in response to a contractionary monetary policy shock, and these responses are highly heterogeneous in the distribution of markups. In particular, we find that in response to a contractionary monetary policy shock, firms with large markups increase them further relative to lower markup firms. The latter implies that production contractions increase markup dispersion, meaning not only that there is a more inefficient allocation of resources in recessions but that the burden of the recessions is heterogeneous across firms.

To rationalize our findings, we develop a theoretical framework relying on Champion et al. (2023). This model consists of a continuum of heterogeneous firms that face Calvo price rigidities and Kimball demand for their goods. As shown by Baqaee et al. (2021), Kimball demand generates heterogeneity of markups and pass-through from marginal costs to prices. Due to strategic complementarities, firms with higher market share can charge a higher markup and, under some

[1] A shortcoming of papers that use COMPUSTAT, for example, is not only that they rely on big firms and do not observe prices, but the fact that they rely on "cost of goods sold, COGS" markups, which combine materials an labor costs, potentially biasing markups estimations.

unrestrictive assumptions, have a lower pass-through. Combining the latter with price rigidities, the model generates heterogeneous markup responses to monetary policy shocks, where firms with higher markups and lower pass-through have more responsive markups.

In line with our empirical facts, we analytically show that inflation dynamics are affected by firm's heterogeneous market power. With heterogeneous markups, the slope of the aggregate Phillips curve depends on the covariance between the market share and the passthrough from marginal costs to prices. When this covariance is negative, the slope of the Phillips curve falls, meaning the markup heterogeneity generates additional price stickiness at the aggregate level, and hence, there are larger real effects from monetary policy. A key result is that the slope of the Phillips curve depends on the demand superelasticity (the relationship between markups and market power) that drives the degree of strategic complementarities in the economy. We show that the larger this elasticity, the larger the covariance, leading to a further flattening of the Phillips curve.

Finally, we study the implications of these facts for inflation and show the conditions in which more market power leads to lower inflation volatility and its consequences for the business cycles and macroeconomic policies. We highlight the role of how the demand elasticity varies with firm size (the so-called superelasticity), governing the strength of the variable markup mechanism for transmitting shocks.

Related literature. Previous studies have focused on the evolution of aggregate wedges, such as labor wedges. Some examples are Galí et al. (2007); and most recently Bils et al. (2018), and Nekarda and Ramey (2020). On the other hand, Hong (2017), Burstein et al. (2020), Afrouzi and Caloi (2023), and Anderson et al. (2018), among others, study the cyclicality of markups unconditionally. Few papers study the response of firm-level markups to monetary policy shocks. Meier and Reinelt (2022) study the response of markup dispersion to monetary policy and motivate these differences with heterogeneity in nominal rigidities, focusing on the real effects of this misallocation mechanism. Chiavari et al. (2021) study the effect of monetary policy on heterogeneous markups focusing on firm age and propose a model similar to ours. Finally, Kouvavas et al. (2021) and Höynck et al. (2023) address a question similar to us, which is studying the response of GDP to monetary policy shocks conditional on high and low aggregate market power. The former does it at a sectoral level (studying sectors with high versus low markups), while the latter does it at an aggregate level over time.

Our contribution to this literature is, first, to show that the cyclicality of markups *conditional* to shocks can be different from the unconditional cyclicality, finding that the latter is negative conditional on monetary policy shocks even in the groups of firms (or sectors) that have zero or positive correlation with GDP. And second, the cyclicality is heterogeneous and increases on the markup level of the firm, which is consistent with models of strategic complementarities applied to business cycles (e.g., Kimball (1995)).

Organization of the Paper. The structure of the paper is as follows. Section 2 discusses the computation of our monthly firm-level markups. Section 3 studies the cyclical behavior of markups conditional on monetary policy shocks and unconditionally. Section 4 shows the model and studies the role of strategic complementarities for the New Keynesian Phillips curve. Section 5 studies quantitatively how heterogeneity in markups contributes to inflation and output fluctuations. Finally, section 6

2 Markup Computation

To correctly describe markup effects on inflation dynamics, its computation becomes crucial. Since De Loecker and Warzynski (2012) the market power literature has been active in estimating markups using the production approach, where there seems to be a consensus; markups have been growing in the last decades around the world (De Loecker and Eeckhout, 2018) driven by the firms at the top of markup distribution. The production approach markup estimation has some challenges, mainly in the production function estimation using revenue ($P \cdot Q$) instead of output (Q) to recover the output elasticity of a variable input. While Bond et al. (2021) argues that the latter might invalidate production approach markup estimations using revenue, De Loecker (2021) explains the seminal strategy in De Loecker and Warzynski (2012) accounts for prices on the production estimation procedure by treating them as a relevant omitted variable. While De Ridder et al. (2022) argues in favor that the production approach using revenue recovers unbiased markup changes estimates, the markup level is often biased. Our focus is on inflation dynamics, making both markup levels and its changes in time, relevant parameters to describe markup's effects on inflation correctly. In the model section, we will be explicit about why the markup level is needed to describe inflation dynamics.

This work does not intend to go deeper into the markup estimation discussion but to use the best possible estimation, following the production approach, with the available data. As prices are available in the data this work employs, the markup estimation performed is the (almost) ideal following De Loecker and Warzynski (2012) in data requirements. This work follows their methodology then to estimate markups.

Markup (μ) is defined as the price over the marginal cost. The cost minimization problem that firm i faces given its technology (Q_{it}), capital stock (K_{it}) and variable inputs (V_{it}) can be represented by:

$$\mathcal{L}(V_{it}, K_{it}, \lambda_{it}) = \sum_{V} P_{it}^{V} V_{it} + r_{it} K_{it} + \lambda_{it}(\bar{Q}_{it} - Q_{it}(V_{it}, K_{it}))$$

Where the first order condition with respect to variable input V is:

$$\frac{\partial \mathcal{L}}{\partial V_{it}} = P_{it}^{V} - \lambda_{it} \frac{\partial Q(.)}{\partial V_{it}} = 0$$

Rearranging and multiplying by $\frac{V_{it}}{Q_{it}}$:

$$\frac{\partial Q(.)}{\partial V_{it}} \frac{V_{it}}{Q_{it}} = \frac{1}{\lambda_{it}} \frac{P_{it}^{V} V_{it}}{Q_{it}}$$

In a cost minimization environment, the Lagrangian multiplier is equivalent to the marginal cost of a firm ($MC_{it} = \lambda_{it}$), and hence $\mu_{it} = \frac{P_{it}}{\lambda_{it}}$. Therefore, by multiplying by $\frac{P_{it}}{P_{it}}$ and rearranging the above equation, the expression for markup can be generated:

$$\frac{P_{it}}{\lambda_{it}} = \underbrace{\frac{\partial Q(.)}{\partial V_{it}} \frac{V_{it}}{Q_{it}}}_{\theta_{it}^{V}} \underbrace{\frac{P_{it} Q_{it}}{P_{it}^{V} V_{it}}}_{1/s_{it}^{V}}$$

$$\mu_{it} = \frac{\theta_{it}^{V}}{s_{it}^{V}} \tag{1}$$

Markup measure relies then on two objects: the variable input share (s_{it}^{V}), which is usually available in the data, and the output elasticity of a variable input (θ_{it}^{V}), which needs to be estimated and represents the critical challenge of this approach. This methodology does not require any assumption on the production function functional form, nor on the structure of demand, or how firms compete.

Measures of output and inputs are available from the data, but all parameters must be estimated using a production function estimation methodology. This work will rely on Ackerberg et al. (2015) to guarantee the identification of parameters and perform the estimation.

Any firm i at time t will choose its production level as a function of inputs (fixed K_{it}, and variables V_{it}), its correspondent coefficients, and a Hicksian neutral productivity (A_{it}). As the observed data is subject to imperfections, the estimation procedure allows for a measurement error of the observed output. In addition, firms are subject to unanticipated shocks to production. The measurement error and unanticipated shocks to production are grouped in an error term; ϵ_{it}. Hence, the observed output from the data is denoted by Y_{it} and is defined, using lowercase variables as logarithms, as:

$$y_{it} = q_{it} + \epsilon_{it}$$

ϵ_{it} are assumed to be independent and identically distributed. A fundamental assumption to reach identification is that firms do not observe ϵ_{it} when making their optimal input hiring decisions.

To form the output variable-input elasticities (θ_{it}^V), consistent estimators of the technology parameters (β) are required. This work will rely on linear production function functional forms assumptions[2] (eg. Cobb-Douglas, Translog production functions) to recover them. Hence, the production function is described by (lower case variables denote logarithms):

$$q_{it} = f(v_{it}, k_{it}; \beta) + A_{it} + \epsilon_{it}$$

We will rely on translog production functions as our benchmark. With the explained methodology above[3] is possible to estimate the output variable-input elasticity (θ_{it}^V) by assuming a production function functional. Cobb-Douglas is the natural candidate for the production function to simplify the estimation procedure. Nevertheless, when using Cobb-Douglas, the output elasticities are

[2]Following Ackerberg et al. (2015), Olley and Pakes (1992) and Levinsohn and Petrin (2003) proxy methods

[3]Also, note that the markup estimation is (at least) sensitive to the production function parametric assumption, production function output variable, variable input to estimate markup, production function estimation methodology, output and materials price controlling, time-variant versus time-invariant output variable input elasticity, and disaggregation level of the production function estimation.

constant within a firm in time and independent of input use intensity, which can omit relevant information. A second-order translog production function is assumed to allow output-variable input elasticities to be time-variant. Departing from Cobb-Douglas production functions does not restrict the output elasticities to be independent of input use intensity. Cobb-Douglas production functions implicitly assume that variation in technology is explained by variation in markups, which might generate biased results.

There are two candidates to use as output variables: value added and production. Valued added production functions are broadly used in the literature[4] where its central assumption is that materials are used in a fixed proportion (as a Leontief production function) to produce. We will depart from value-added production functions and use quantities as the output variable.

The main advantage of using production quantities is that the markup from materials can be recovered, which is a more flexible input than labor (the other input to retrieve markups). A more flexible production input will better identify the markup as potential frictions that might generate a wedge between the marginal product and the input price are less likely to occur. For example, materials have no hiring or firing costs (or are certainly smaller than labor ones). Therefore, the markup recovered from materials is more likely to be better identified than the one identified from a labor variable and, hence, is the one this work will use as a benchmark.

Assuming a translog production function means that $f()$ is supposed to be better approximated by a second-order polynomial -all in logarithms terms - where inputs, inputs squared, and interaction terms between every input are included. The logarithms assumption is critical to maintaining a linea production function.

This methodology allows to assume (at least) value added and gross output production functions, which is an advantage when data is restricted. Also, by being flexible, the methodology allows for comparing both production functions, conditional on data availability. This work can estimate both; a discussion on which is preferred is done in the estimation section.

Another critical assumption is that input prices vary and are serially correlated across firms.

[4] The most cited paper on markups estimation for the US, De Loecker et al. (2020), use valued added as its benchmark production function.

The latter allows lagged firms' input choices to identify production function inputs (and their interactions) coefficients (β).

The Hicksian-neutral productivity term from the production function (no matter its functional form assumption) A_{it} follows a first-order Markov process with innovation (or shock) to productivity in time t (φ_{it}):

$$A_{it} = h(A_{it-1}) + \varphi_{it} \tag{2}$$

The shock is potentially correlated with the production inputs choices at period t, which generates an endogeneity problem as noted by Olley and Pakes (1992) and Levinsohn and Petrin (2003). Using a control function for intermediate inputs following Levinsohn and Petrin (2003), where the material inputs demand decision is a function of the fixed input, the variable input, and productivity, it is possible to back up the productivity term. The material demand decision can be represented by:

$$m_{it} = g_t(k_{it}, v_{it}, A_{it})$$

Assuming that g_t is a strictly increasing monotonic function in A_{it}, meaning that as productivity increases, intermediate input demand also increases, the intermediate input demand can be inverted to recover productivity.

$$A_{it} = g_{it}^{-1}(k_{it}, v_{it}, m_{it}) \tag{3}$$

This work will assume as a benchmark an output-based (not valued-added) second-order translog production function with two variable inputs (materials and labor) and a fixed input (capital) and follow the Ackerberg et al. (2015) two-stage estimation procedure. The production function can be described then as:

$$\begin{aligned} y_{it} = A_{it} &+ \beta_l\, l_{it} + \beta_k\, k_{it} + \beta_m\, m_{it} + \beta_{lk}\, l_{it}\, k_{it} + \beta_{lm}\, l_{it}\, m_{it} + \beta_{mk}\, m_{it}\, k_{it} \\ &+ \beta_{ll}\, l_{it}^2 + \beta_{kk}\, k_{it}^2 + \beta_{mm}\, m_{it}^2 + \beta_{lkm}\, l_{it}\, k_{it}\, m_{it} + \epsilon_{it} \end{aligned} \tag{4}$$

The first stage consists of regressing the observed output on observed input and input interactions to recover a measure of expected output ($\hat{\phi}_{it}$) and an estimation for

ϵ_{it}:

$$y_{it} = \phi(v_{it}, k_{it}, m_{it}) + \epsilon_{it}$$

Where the expected output expression, using equation equation 3, is:

$$\begin{aligned}\hat{\phi}_{it} = g_{it}^{-1}(k_{it}, v_{it}, m_{it}) + \beta_l\, l_{it} + \beta_k\, k_{it} + \beta_m\, m_{it} + \beta_{lk}\, l_{it}\, k_{it} + \beta_{lm}\, l_{it}\, m_{it} + \beta_{mk}\, m_{it}\, k_{it} \\ + \beta_{ll}\, l_{it}^2 + \beta_{kk}\, k_{it}^2 + \beta_{mm}\, m_{it}^2 + \beta_{lkm}\, l_{it}\, k_{it}\, m_{it}\end{aligned} \tag{5}$$

Relying on the law of motion of the productivity term (equation 2), the second stage of the Ackerberg et al. (2015) implementation will recover production function coefficients (β). With the assumed production function, the vector β is described by:

$$\beta = (\beta_l, \beta_k, \beta_m, \beta_{lk}, \beta_{lm}, \beta_{mk}, \beta_{ll}, \beta_{kk}, \beta_{mm}, \beta_{lkm}) \tag{6}$$

The first step of the second stage is to compute productivity for any values of vector β. So that productivity for each firm can be computed as:

$$\begin{aligned}A_{it}(\beta) = \hat{\phi}_{it} - (\beta_l\, l_{it} + \beta_k\, k_{it} + \beta_m\, m_{it} + \beta_{lk}\, l_{it}\, k_{it} + \beta_{lm}\, l_{it}\, m_{it} + \beta_{mk}\, m_{it}\, k_{it} \\ + \beta_{ll}\, l_{it}^2 + \beta_{kk}\, k_{it}^2 + \beta_{mm}\, m_{it}^2 + \beta_{lkm}\, l_{it}\, k_{it}\, m_{it})\end{aligned}$$

With computed productivities for each firm in every period, the second step (of the second stage of Ackerberg et al. (2015) procedure) is to recover the innovation to productivity term, φ_{it}, by non-parametrically regressing $A_{it}(\beta)$ on its lag, $A_{it-1}(\beta)$.

The third step involves building moments to get the point estimates of production function parameters. Assuming that the innovation to productivity (φ_{it}) is orthogonal to all production function coefficients[5] is possible to form the following moments:

[5]The firm's decision of capital at period t, the firm determination of variable inputs at period $t-1$ and all the possible factor interaction terms.

$$\mathbb{E}\left[\varphi_{it}(\beta)\begin{pmatrix} l_{it-1} \\ m_{it-1} \\ k_{it} \\ l_{it-1}^2 \\ m_{it-1}^2 \\ k_{it}^2 \\ l_{it-1}\cdot m_{it-1} \\ l_{it-1}\cdot k_{it} \\ m_{it-1}\cdot k_{it} \\ l_{it-1}\cdot m_{it-1}\cdot k_{it} \end{pmatrix}\right] = 0 \tag{7}$$

Therefore, all production function parameters are recovered using the Generalized Method of Moments, given that there will be at least the same number of moments as parameters.

Specifically, this work uses materials output variable-input elasticity to form the benchmark markup. Hence, it is possible to consistently identify the five coefficients related to materials in equation 4 using lagged material optimal firm choices as instruments. The latter is done by assuming that material input prices are serially correlated over time, an assumption supported by the data used in this work (and by the original paper that developed this methodology, De Loecker and Warzynski, 2012).

2.1 Data used for estimation

The production function is estimated monthly (t = month). Performing the estimation at a low frequency may be irrelevant if the firm's technology does not change monthly. But the fact that the data used in this work allows for monthly production function estimations at the firm level highlights the importance of yearly input usage intensity variation (why is omitted when estimating yearly production functions). Input usage monthly variation will generate input shares variation in low frequency given place to monthly markup heterogeneity. Capital is almost unchanged within a given year, while labor does change month-to-month, but it is harder to hire or lay off workers than just stop buying and giving intermediate input; this is why this work will rely on markups from intermediate inputs (materials).

The firm-level monthly data is gathered from five different Chilean IRS (SII) sources. Information on firm sales, materials, and investment comes from the VAT monthly form (Form 29) from the SII. Data on monthly labor (headcounts) and wages come from a different SII form (DJ1887) while data on the stock of capital comes from the income IRS form 22, which is reported yearly. Data on the yearly stock of capital and monthly investment is used to perform a perpetual inventory method to recover firm-level monthly capital stock. The fifth source is data from electronic tax documents (invoices universe) that provide information on each product, including its price and quantity, traded domestically or internationally with at least one Chilean firm participant from 2014 on. This data is used to recover price variation-free variables for output and materials.

To achieve this, we construct firm-level output and intermediate goods input price indices, leveraging the richness of invoice-level price information available. For every formal firm in Chile, we have records of all the goods a firm sells and all the goods it purchases as intermediate inputs. This comprehensive data allows us to generate quantity indices for both aggregate production and aggregate intermediate goods inputs. Price indices for output and intermediate inputs are established using standard Tornqvist indices. We selected the year 2014 as the base year for constructing our price indices because it was the first year in which we observed prices for firm-to-firm transactions. This method is widely recognized for estimating aggregate production functions at the firm or plant level when price data is accessible (Dhyne et al. (2022) and De Roux et al. (2021)).

To maintain consistency in our approach, we compute firm-specific annual weighted average prices (P_{igt}) for each product (g) sold by firm i during year t. Subsequently, we construct firm-specific price indices (ΔP_{it}) for products observed in consecutive years using the product-level weighted average price and the share of the product present in both year $t-1$ and year t:

$$\Delta \log P_{it} = \sum_{g} \frac{s_{igt} + s_{igt-1}}{2} \Delta \log(P_{igt}) \tag{8}$$

s_{igt} represents the revenue share of product g for firm i at time t.

Consequently, we utilize the following output value for estimating the production function:

$$q_{it} = \frac{\text{Revenue}_{it}}{P_{it}} \tag{9}$$

A similar procedure is applied to materials, ensuring that the measure for materials used in the production function estimation is also free from price variation [6]:

$$m_{it} \approx \frac{\text{Material expenditure}it}{P_{it}^{M}} \tag{10}$$

There is an industry identifier for each firm at the 6-digit ISIC (rev. 4) level, allowing estimations from 21 (letter ISIC rev.4 levels) productive sectors up to more than 800 (when using six-digit sectors). The benchmark case uses only six ISIC rev.4 letter sectors.

3 Cyclical Behavior of Markups

In this section, we show some stylized facts about monthly markups. First, we show how we classify firms into the markups distibution. Second, we show evidence on several moments of the markups' distribution. And finally, we study the cyclicality of markups depending on several firms' quintile in the markup distribution and sector.

3.1 Classification of firms into quintiles

Since we are interested on the heterogeneity in the markup cyclicality, we first classify firms in the distribution of markups. We follow the ideas by AKM in which they in the context of labor income, classify firms and workers into quintiles by estimating a fixed effect regression. To do so, we estimate the following reggression:

$$\log \mathcal{M}_{it} = \boldsymbol{\alpha_i} + \psi_j + \delta_t + c + \epsilon_{it}$$

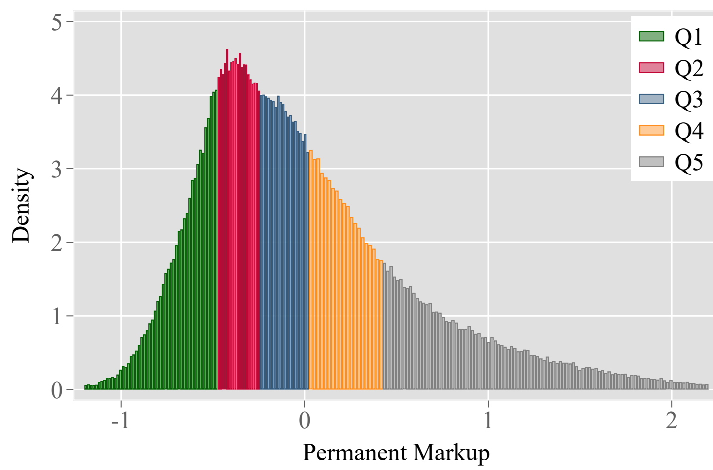
where $\boldsymbol{\alpha_i}$ is a firm fixed-effect, ψ_j is a sector fixed effect and δ_t is a month fixed effect. We are interested in $\boldsymbol{\alpha_i}$ which is a measure of *permanent markup*, it is the markup around which the firm fluctuates over the period we analyze. Figure (1) shows the distribution of markups fixed-effects.

3.2 Unconditional Cyclicality

Now we study the cyclicality of our monthly markups and the moments of the distribution. Figure 2 shows the detrended and deseasonalized median markup along with industrial production. At

[6] We have set the base year as 2014, when price data will be available for both input and output prices.

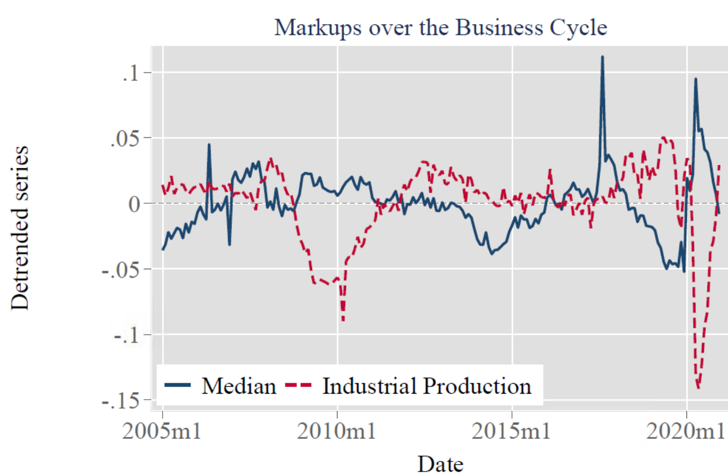
FIGURE 1: Distribution of markup fixed effects



Notes: . Source: Own elaboration, Internal Revenue Service, Chile.

first glance, there is a negative correlation between the detrended median markup and the cycle, with some periods where this decouples. 2020 is striking with a significant increase at the time of the start of the COVID pandemic associated with a strong recession. The coefficient of correlation between these two variables is -0.52.

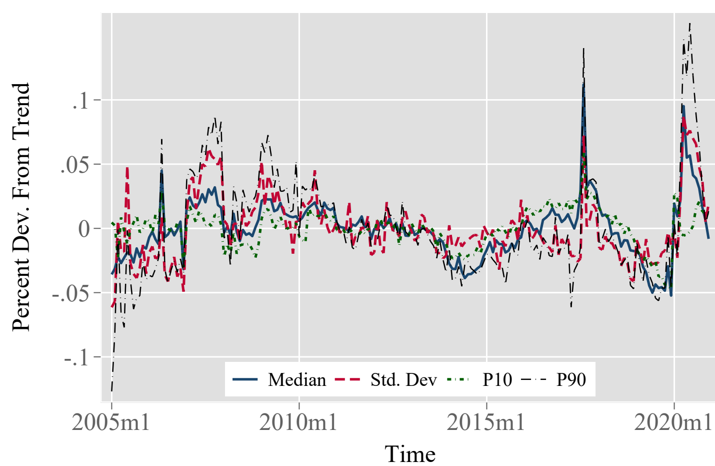
FIGURE 2: Median markup in the business cycle



Notes: . Source: Own elaboration, Internal Revenue Service, Chile.

In Figure 3 we show the evolution of the markups' distribution over time. We show the median, the standard deviation, and the tenth and 90th percentiles. Since we are interested in the response of the distribution of markups, it is worth showing the evolution of some aggregates. Apart from having mild countercyclicality in the median, we observe that not only the median fluctuates but also the standard deviation. We observe that the standard deviation is countercyclical and represented by the stronger cyclicality of the 90th percentile than the tenth. In fact, the coefficient of correlation between industrial production and the standard deviation is -0.53; for the tenth percentile, it is -0.12, and for the 90th percentile, it is -0.57.

FIGURE 3: Moments of the markups' distribution



Notes: . Source: Own elaboration, Internal Revenue Service, Chile.

Hence, at an aggregate level, markups are countercyclical, and the distribution becomes more dispersed in recessions. This is similar to the findings by Meier and Reinelt (2020), who show that in response to monetary policy shocks, there is a rise in the dispersion of markups, leading to the misallocation of resources over the business cycle.

Panel Regressions. Although the previous results are interesting, they could be affected by aggregation bias and other problems that arise from aggregation. To tackle that, we take advantage of the panel structure of our data and run panel regressions of industrial production growth on firm-level markup growth to account for individual level and time effects and exploit the heterogeneity more explicitly. Hence, we run the following regression.

$$\Delta \log \mathcal{M}_{it} = \alpha_i + \alpha_m + \beta \Delta \log Y_t + \Gamma' X_t + \epsilon_t, \tag{11}$$

where α_i and α_m are firm and month fixed effects; $\Delta \log Y_t$ is monthly industrial production growth (deseasonalized); and X_t are macroeconomic controls, in particular, inflation and exchange rate changes (to account for other sources of variability of the markup). Hence, we are interested in the parameter β. Table 1 shows the results of two specifications, with and without firm fixed effects. Thus, we find that *unconditionally*, our monthly markups are statistically significantly countercyclical; this means that in downturns, the markup increases. There are two main interpretations of this result. First, that in downturns, there is a cleansing effect that destroys firms that have, on average, a low markup, thus increasing markups on average (as in Baqaee and Farhi (2020)). A second one is that in the presence of nominal rigidities, some shocks (like demand shocks) that generate a contraction in the economy increase markups (as in the simple New Keynesian model). The reason is that with price rigidities, the adjustment does not go through prices, but through the margin the firm gets.

TABLE 1: Cyclicality of firm-level markups

	Dep. var: $\Delta \log \mathcal{M}_{it}$	
$\Delta \log Y_t$	-0.058***	-0.058***
	(0.013)	(0.013)
π_t	-0.01	-0.01
	(0.08)	(0.08)
$\Delta \log e_t$	0.07***	0.07***
	(0.01)	(0.01)
Month FE	Yes	Yes
Firm FE	Yes	No
N	7116398	7116398
Adjusted R^2	0.001	0.001

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The first interpretation we exposed above can be tackled by separating the sample into *markup sizes*. To do so, we take advantage of the classification we described above and run the following regression.

$$\Delta \log \mathcal{M}_{it} = \alpha_i + \alpha_m + \Gamma' X_t + \sum_{q=1}^{5} \beta_q D_q \Delta \log Y_t + \epsilon_t \tag{12}$$

where we allow the cyclicality β_q to vary depending on the quintile of the permanent markup. Therefore, if there is a somewhat cleansing effect, the markup of the lowest quintiles must not be higly countercyclical.

Table 2 shows the results of estimating Equation (12). We find several results that are interesting for us. First, the markups in all quintiles of the markups distribution are countercyclical. The markups in all the quintiles are significantly countercyclical, except for those in the third quintile. With this result, we claim that the cyclicality may not be due to the cleansing effect since the countercyclicality is present in all quintiles. the second, and most important result for us is the heterogeneity in the cyclicality. We find that the cyclicality is highly heterogeneous, and for out unconditional cyclicality looks like an inverted "U". Most importantly, we find that the fight quintile (firms with the highest markups) have the most countercyclical markups.

TABLE 2: Cyclicality of markups by quintile

	Dep. var: $\Delta \log \mathcal{M}_{it}$				
	Q_1	Q_2	Q_3	Q_4	Q_5
$\Delta \log Y_t$	-0.09***	-0.07***	-0.02	-0.04**	-0.13***
	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)
N	7103390				
Adjusted R^2	-0.014				

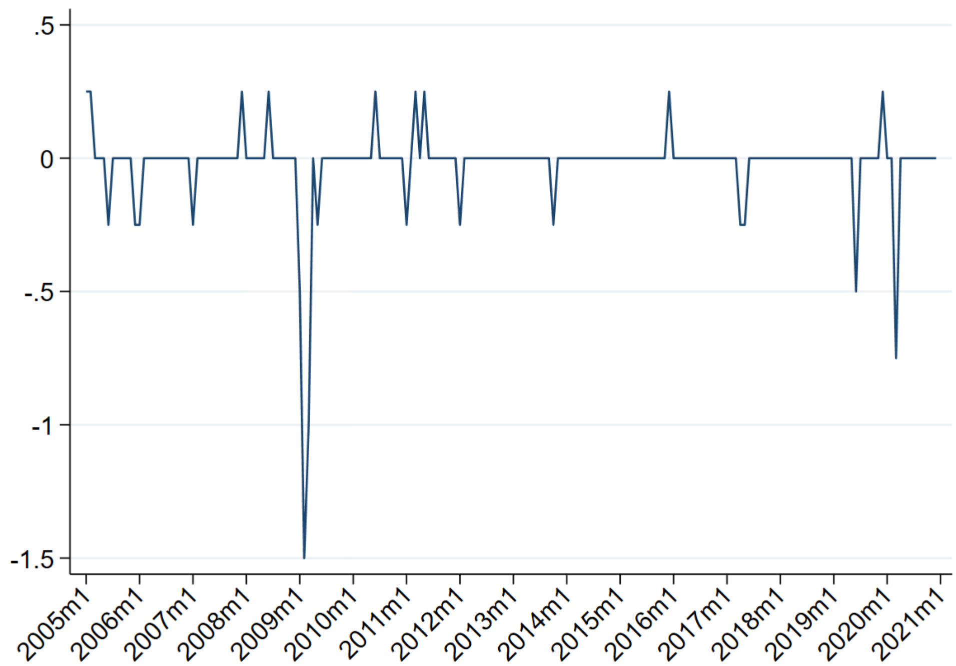
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

All these results point towards the existence of significant price rigidities and very dynamic adjustments the firms face. However, in order to confirm our hypothesis that markups are countercyclical due to nominal rigidities and that they are highly heterogeneous, we must evaluate the responses of these markups to identified demand shocks, which we do in the next subsection.

3.3 Markups' Response to Monetary Policy Shocks

This subsection studies the conditional cyclicality of our monthly markups based on the monetary policy surprises (MPS). Aruoba et al., 2021 conducted a comparative analysis of monetary policy surprises in Chile obtained from surveys as well as swaps on monetary policy rates and found that Bloomberg survey is the most suitable source for deriving MPS. Figure 4 shows the 24 MPS observed between 2005 and 2020. In their paper, they show that these surprises behave like a theoretical monetary policy shock: output and the CPI drop.

FIGURE 4: Monetary Surprises ε_t



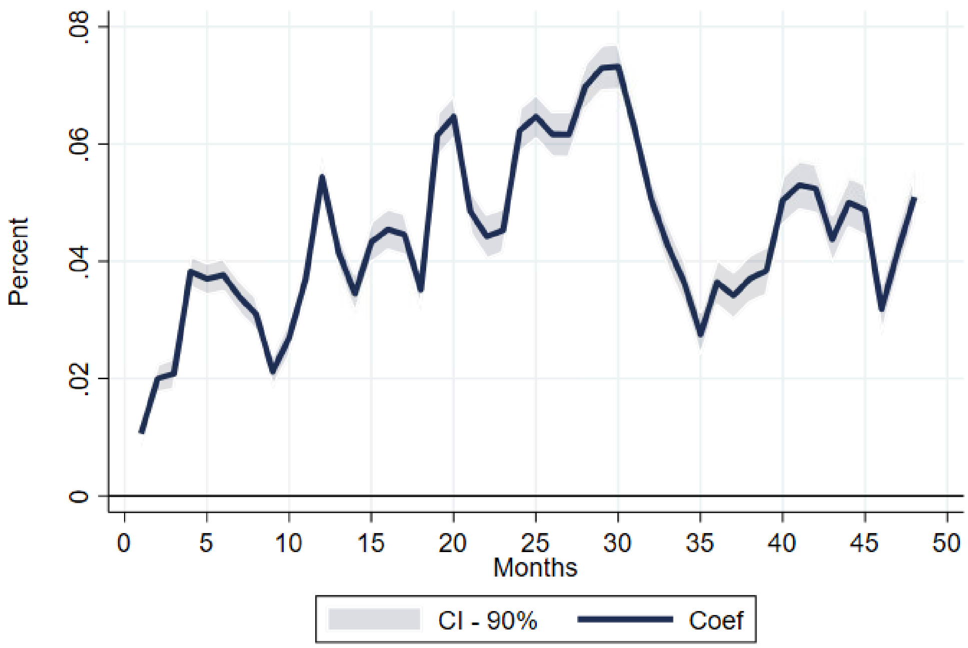
We investigate the dynamic effects of monetary policy surprises on changes in firm-level markups. We utilize a panel local projection (Jorda, 2005) methodology to do this. Doing so requires regressing a firm-level markup variable on the monetary policy surprise and controls. In particular, for a horizon of $h \geq 1$ months after the surprise, we estimate the following empirical model

$$\log \mathcal{M}_{it+h} - \log \mathcal{M}_{it} = \alpha_j + \alpha_{sq} + \beta_h \mathcal{MPS}_t + \beta_x X_{jt} + \epsilon_{jt+h}, \tag{13}$$

where t evolves in months and $\mathcal{MPS}_t$ is the monetary policy surprise in month t – in all but 24 months this variable is zero. We include firm fixed effects denoted by α_j. We also include firm level controls X_{jt} which includes size (employment) and sales. The main coefficient of interest is β_h as it shows, for every horizon h, the effect of the monetary surprise $\mathcal{MPS}_t$ on the dependent variable.

Figure 5 plots the coefficients of interest $\{\beta_h\}_{h=0}^{48}$ at different horizons with 90% confidence bands for the changes in firm-level markup regressions. Consistent with the new Keynesian approach, Figure 5 shows that markup responds positively when monetary policy tightens, with the response being significant starting from the first month after an MPS and increasing over time. The magnitude of the coefficient can be interpreted as 100 basis points decrease in a MPS leading to an accumulated increase of markup of between 1% and a maximum of 7%. All coefficients have the expected sign and are statistically significant. Therefore, we find evidence that firm-level markup is conditionally countercyclical on average. This means that firms facing some degree of price rigidities can not be ruled out.

FIGURE 5: Effects on Firm-level Markup β_h



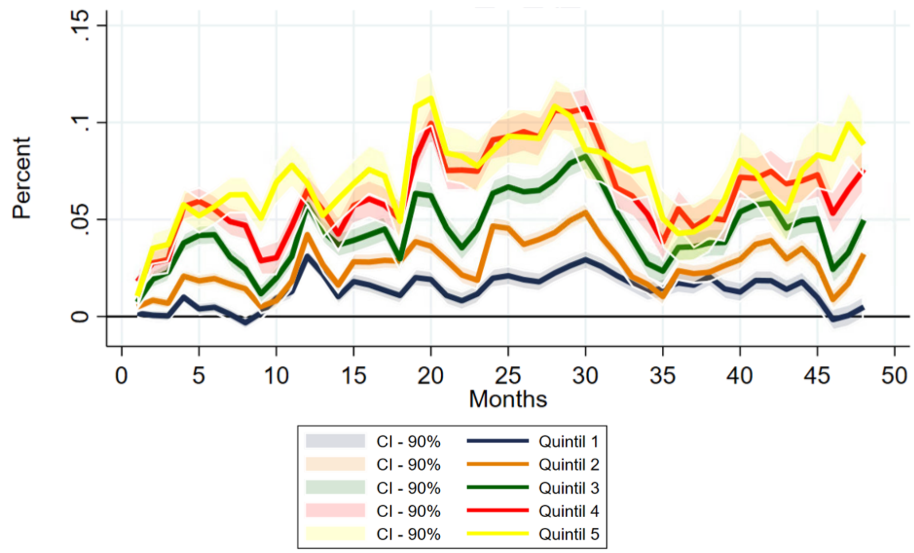
Firm heterogeneity. We are interested not only in the aggregate (or average) response to markups but also in the heterogeneity of the responses. In this paper, we study the responses of markups along *the distribution of markups*, i.e., how markups respond depending on their size. To do so, we use the quintiles we obtained with the fixed-effects procedure and run Equation (13) with quintile dummies as follows

$$\log \mathcal{M}_{it+h} - \log \mathcal{M}_{it} = \alpha_j + \alpha_{sq} + \sum_{q=1}^{5} \beta_h^q \mathcal{MPS}_t + \beta_x X_{jt} + \epsilon_{jt+h}, \tag{14}$$

where again, we are interested in the sequence of responses $\{\beta_h^q\}_{h=0}^{48}\ \forall\ q \in [1,5]$.

Figure 6 shows the results of estimating Equation (14). The first to note is the great heterogeneity between firms in the different quintiles; however, unlike the evidence on Table 2 for the unconditional cyclicality, we find that markups for all these groups respond negatively. Most importantly, we find that there is a monotone and positive relationship between the quintile of markup and its responses, in which firms with higher markups switched their markups further in response to a monetary policy shock than firms with lower markups. The differences are significant, with markups at the bottom with a maximum change of about 3%, while markups at the top change by about 10%.

FIGURE 6: Heterogenous Results: Quintiles of α_i



Discussion. The evidence provided in this section is important for several reasons. First, we show that on average markups are unconditionally countercyclical; however, by group they can not be strictly countercyclical. These results motivate estimating the responses to shocks, that theoretically, we have a clear prediction for the cyclicality of the markup (a monetary policy shock). In response to a monetary policy shock, we find that our markups are countercyclical for all the groups we study regardless of their unconditional cyclicality. Which implies that is important to go beyond the unconditional cyclicality. Second, we find a clear pattern between the response of markups and the "competitiveness position" (the markup size), which is that firms with higher market power have fewer incentives to switch prices in the short run. This means that the high markup dispersion we find may contribute to aggregate price stability. Therefore, the prescriptions of NK models that generate markup countercyclicality can not be ruled out in light of these empirical facts. Next, we study in a model following Baqaee et al. (2021) the consequences for aggregate inflation of the heterogeneity in markup response that we find.

4 Model

Our model is an extension of Nakamura and Steinsson (2008), which is a heterogeneous firm model with price rigidities. The model has a continuum of firms that demand labor to produce units of a differentiated variety and are subject to idiosyncratic productivity shocks. We assume consumers aggregate consumption with an implicit non-homothetic Kimball aggregator, following Baqaee and Farhi (2020). Due to the Kimball assumption there is a distribution for markups, demand elasticities, and marginal costs to price pass-through. Hence, at a microeconomic level, there is heterogeneity in these variables. We then show aggregate implications of such heterogeneities for prices and output. Then, we solve the model with aggregate shocks, that generate inflation beyond the steady state inflation rate, and evaluate the role of heterogeneity in the transmission of shocks.

Representative Consumer. Time is discrete $t = \{0, 1, 2, \dots\}$ The representative consumer maximizes

$$\mathbb{E}_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left(\log C_t - \frac{L_t^{1+\varphi}}{1+\varphi} \right) \right\} \tag{15}$$

subject to

$$P_t C_t + Q_t B_{t+1} = W_t L_t + R_t X_t + \Pi_t + B_t$$

where C_t denotes real consumption of the final good, L_t denotes labor supplied, P_t aggregate prices, Q_t the nominal price of a bond B_{t+1}, W_t is the nominal wage, and Π_t is nominal profits. There is a factor of production X_t with nominal rental rate R_t that is like an endogenous endowment households own but do not decide how much to accumulate.

The first-order conditions of the household problem are given by the Euler equation

$$Q_t = \mathbb{E}_t \left\{ \beta \frac{C_t}{C_{t+1}} \frac{P_t}{P_{t+1}} \right\}$$

and the labor supply

$$C_t L_t^{\varphi} = \frac{W_t}{P_t} \tag{16}$$

Final good producers. The final good is produced by perfectly competitive firms using a bundle of intermediate inputs. The technology for transforming the bundle of intermediate inputs y_{it} for $i \in [0,1]$ into the final good is represented by a Kimball aggregator of the form

$$\int_0^1 \Upsilon\left(\frac{y_{it}}{Y_t}\right) di = 1 \tag{17}$$

where the function $\Upsilon : \mathbb{R}_+ \to \mathbb{R}_+$ is strictly increasing and strictly concave. The CES is recovered if Υ is a power function.

The final goods producer takes input prices p_{it} as given and chooses input demand y_{it} for $i \in [0,1]$ to maximize profits

$$P_t Y_t - \int_0^1 p_{it} y_{it} di$$

subject to the Kimball aggregator (17) above.

Kimball Demand System. The implied inverse demand curve facing intermediate producer $i \in [0,1]$ is then given by

$$\frac{p_{it}}{P_t} = \Upsilon'(q_{it}) D_t, \quad q_{it} := \frac{y_{it}}{Y_t} \tag{18}$$

where P_t is the ideal price index, D_t is the Kimball demand index, and q_{it} is a measure of the relative size of firm i. The price index is the size-weighted average price

$$P_t = \int_0^1 p_{it} q_{it} di. \tag{19}$$

The Kimball demand index D_t is the inverse of the size-weighted average of marginal productivities

$$D_t = \left(\int_0^1 \Upsilon'(q_{it}) q_{it}\right)^{-1} \tag{20}$$

This demand system implies that a firm's sales share ω_{it} is pinned down by its relative size

$$\omega_{it} := \frac{p_{it} q_{it}}{P_t Y_t} = \Upsilon'(q_{it}) q_{it} D_t. \tag{21}$$

Firms We follow closely Nakamura and Steinsson (2008), which is a generalization of Golosov and Lucas Jr (2007). There is a continuum of firms in the economy indexed by z. Each firm belongs to one of the sectors J and specializes in the production of a differentiated product. The production function of firm j is given by

$$y_t(z) = A_t(z)L_t(z)^{1-\alpha} \tag{22}$$

where $y_t(z)$ denotes the output of firm z in period t, $L_t(z)$ is the quantity of labor employed by firm z, $1-\alpha$ denotes the labor share in production, and $A_t(z)$ is the idiosyncratic productivity of firm z at time t.

Firm z in sector j maximized the value of its expected discounted profits,

$$\mathbb{E}_t \sum_{s=0}^{\infty} D_{t,t+s}\Pi_{t+1}(z), \tag{23}$$

where profits in period t are given by

$$\Pi_t(z) = p_t(z)y_t(z) - W_tL_t(z) - \chi_j W_t I_t(z) - P_t U. \tag{24}$$

With $C_t = Y_t$ aggregate consumer demand, the demand for good z is given by

$$y_t(z) = \left(\frac{p_t(z)}{P_t}\right)^{-\theta} Y_t \tag{25}$$

with CES, while the inverse demand with Kimball is given by

$$y_t(z) = [\Psi_t']^{-1}\left(\frac{p_t(z)}{P_t}\right)\frac{Y_t}{D_t} \tag{26}$$

The firm maximizes profits, Equations (23) and (24) subject to the demand Eqn. (26) and the production function Eqn. (22). Finally, idiosyncratic productivity follows

$$\log A_t(z) = \rho \log A_{t-1}(z) + \epsilon_t(z).$$

4.1 Strategic Complementarities, Markup Cyclicality and the Phillips Curve

4.2 Flexible Prices

With flexible prices, firms maximize profits by setting nominal prices $p_t(z)$ that are a markup over their marginal cost:

$$p_t(z) = \mathcal{M}_t(z)\Psi_t(z)$$

As Baqaee and Farhi (2020) shows, the flexible price markup can be explicitly written as a function of the firm's market share, $q_t(z)$:

$$\mathcal{M}_t(z) = \mathcal{M}_t(q_t(z)) \tag{27}$$

with $q_t(z) = \frac{y_t(z)}{Y_t}$. The markup function depends on the demand elasticity $\sigma(q_t(z))$ that the firm of market share $q_t(z)$ faces, this is

$$\mathcal{M}_t(q_t(z)) = \frac{\sigma(q_t(z))}{\sigma(q_t(z)) - 1}, \quad \sigma(q_t(z)) := -\frac{\Psi'(q)}{\Psi''(q)q} \tag{28}$$

Then, denote with $\rho_t(z)$ the *pass-through* from individual marginal costs to prices, which is the elasticity of prices to individual marginal costs, absent pricing frictions,

$$\rho_t(z) := -\frac{\partial \log p_t(z)}{\partial \log \Psi_t(z)}$$

The pass-through can be written as a function of the firm's market share

$$\rho_t(z) = \frac{1}{1 + \sigma(q)\frac{\mathcal{M}'(q)q}{\mathcal{M}(q)}} = \frac{1}{1 - \mathcal{M}(q)\frac{\sigma'(q)q}{\sigma(q)}}$$

Result 1. *The passthrough is decreasing on markups.* The derivative of the passthrough to the markup is given by

$$\frac{\partial \rho_t(z)}{\partial q} = \frac{-\epsilon \mathcal{M}'(q)}{1 + \epsilon \mathcal{M}(q)} < 0$$

given that $\mathcal{M}'(q) > 0$.

This means that firms with higher markups have lower pass-through from marginal costs to prices.

4.3 Sticky Prices

Lets, for ease of the explanation, loglinearize the firm's problem. The firm's optimal reset price is given by

$$\ln p^{\star}_{it} = (1-\theta\beta)\left[\rho(\overline{q})\Psi_{it} + (1-\rho(\overline{q}))\left(\ln P_t + \ln D_t\right)\right] + \theta\beta\mathbb{E}_t\{\ln p^{\star}_{it+1}\} \tag{29}$$

Then, noting that $\Psi_{it} = \left(\gamma + \frac{\varphi+\alpha}{1-\alpha}\right)\ln Y_t = \psi \ln Y_t$, the markup follows Result 2, as follows.

Result 2. *Firms with higher markups (and lower pass-through have a more countercyclical markup.* Lets write the effective optimal markup ($\mu^{\star}_{it}$) as follows using $\ln p^{\star}_{it} = \mu^{\star}_{it} + \ln \Psi_{it} = \mu^{\star}_{it} + \psi y_t$:

$$\mu^{\star}_{it} = \underbrace{-(1-(1-\theta\beta)\rho(\overline{q}))\psi}_{\text{Cyclicality of }\mathcal{M}}\ln Y_t + (1-\theta\beta)(1-\rho(\overline{q}))\left(\ln P_t + \ln D_t\right) + \theta\beta\mathbb{E}_t\{\ln p^{\star}_{it+1}\} \tag{30}$$

This implies that the relationship between the markup and aggregate GDP, i.e. the cyclicality of the markup, depends on the coefficient of pass-through and the shape of the cost function. Thus, the cyclicality is given by the term $-(1-(1-\theta\beta)\rho(\overline{q}))\psi$ which implies that due to strategic complementarities in this model, firms with higher markups in steady state, and hence a lower passthrough, have a more cyclical markup. This result depends on the shock, but for a typical demand shock, this applies. This is consistent with our empirical fact that firms with higher markups have more responsive markups after a monetary policy shock.

Another interesting result is the fact that due to the strategic complementarities (and incomplete pass-through), aggregate prices (and inflation) play a significant role beyond their effects on marginal costs. This is because aggregate price level affects markups procyclically. Thus, we would observe a positive relationship between prices and firm-level markups in a high-inflation environment.

Thus, how markups respond to shocks helps us understand the mechanisms that determine the Phillips curve. Next, we study the consequences of these mechanisms for inflation.

Result 3. *Strategic and heterogeneity complementarities affect the Phillips curve.* Using the expression $\pi_{it} = (p^{\star}_{it} - p_{it-1})$, and noting that $\pi_t = (1-\theta)\int q_{it}\pi_{it}di$, individual inflation can be written as

$$\pi_t = \kappa\left[\mathbb{E}_i\{\rho_i\}\psi \ln Y_t + (1-\mathbb{E}_i\{\rho_i\})\ln D_t\right] + \beta\mathbb{E}_t\{\pi_{t+1}\},$$

with $\mathbb{E}_q[\rho_i] = \int_i q_{it}\rho_{it}di$, with q_{it} the market share. Thus, $\mathbb{E}_q\{\rho_i\} = \overline{q}_t\ \overline{\rho}_t + COV_i(q_{it}, \rho_{it})$, and

$\kappa = \frac{(1-\beta\theta)(1-\theta)}{\theta}$. We denote with $\overline{x}$ the average of the firms that change prices. Therefore, the Phillips curve reads

$$\pi_t = \kappa(\overline{q}_t \ \overline{\rho}_t + COV_i(q_{it}, \rho_{it}))\psi \ln Y_t + \kappa(1 - \mathbb{E}_i\{\rho_i\}) \ln D_t + \beta \mathbb{E}_t\{\pi_{t+1}\}. \tag{31}$$

The slope of the Phillips curve depends on the distribution of firms. In particular, it depends on the relationship between market shares and the pass-through to prices. This relationship is tightly related to the superelasticity. Additionally, we have the usual result that is given by the average pass-through, which is that the lower the pass-through, the lower the slope of the Phillips curve.

Discussion. Equation (31) shows the role of heterogeneity in markups and competitiveness position for the slope of the Phillips curve with these classes of demands. Notice that heterogeneity in markups is not enough to generate these results, but we need heterogeneity in firm sizes. Most of the calibrations in the literature show that the covariance is negative, which contributes to the flattening of the Phillips curve beyond the fact that under strategic complementarities ($\overline{\rho} < 1$), the slope of the Phillips curve is lower than in the monopolistic competition case.

5 Quantitative Exploration

5.1 The Firm´s Problem

We first describe the firm's problem

$$V\left(A_t(z), \frac{p_t(z)}{P_{t-1}},\right) = \max_{p_t(z)} \left\{\Pi_t^R + \mathbb{E}_t D_{t,t+1}^R V\left(A_{t+1}(z), \frac{p_{t+1}(z)}{P_t}\right)\right\}$$

where in the case of the Calvo pricing model we write the problem more explicitly as

$$V\left(A_t(z), \frac{p_t(z)}{P_{t-1}},\right) = \max_{p_t(z)} \left\{\Pi_t^R + (1-\theta)\mathbb{E}_t D_{t,t+1}^R V\left(A_{t+1}(z), \frac{p_{t+1}(z)}{P_t}\right) + \theta \mathbb{E}_t D_{t,t+1}^R V\left(A_{t+1}(z), \frac{p_t(z)}{P_t}\right)\right\}$$

where again θ is the probability of not adjusting prices.

With idiosyncratic shocks, we solve this problem with value function iteration. We provide the details in the quantitative section.

5.2 Calibration and Solution Method

While we estimate the monthly markup distribution for the universe of formal firms in Chile, we need to rely on a point estimate for the economy's markup to calibrate the quantitative Exploration. To do so, we separately estimate yearly output translog production functions for each 6-digit industry (626 industries) with at least 100 observations during our sample to recover output elasticities. Following Foster et al. (2022), we aim to permit output elasticities to vary as much as possible within the same aggregate industry. We can estimate production functions for 97% firm-year observations at a 6-digit industry with at least 100 observations. For the remaining 3% of firms-year observations do not have enough data we complement the production function estimation at 160 sectors and 9 sectors.

We pool together our firm-level markups and winsorize both tales of the distribution at 5% levels. We compute the median markup value, 1.409, which we will use as the benchmark markup for our model. We apply the same process for super-elasticities and find that the median super-elasticity is 0.198, which we will also take as our benchmark for calibration purposes.

6 Conclusion

We study firm-level markup cyclicality using administrative microdata from Chilean formal firms and provide a theoretical framework to rationalize our findings. Using monthly tax data for Chile, we estimate firm-level markups using quantities measures for output and materials used in production. We find that markups are countercyclical and present heterogeneous responses to monetary policy shocks.

To rationalize our empirical findings, we develop a model grounded in the framework of Champion et al. (2023) that accounts for heterogeneous product market power (markups). In the model, markup responses are a function of firm market share and pass-through dynamics. Both determinants will define how markups will react to monetary policy shocks. We analytically show how markup heterogeneity affects inflation dynamics, affecting the Phillips curve slope and amplifying the real effects of monetary policy.

References

ACKERBERG, D. A., K. CAVES, AND G. FRAZER (2015): "Identification properties of recent production function estimators," *Econometrica*, 83, 2411–2451.

AFROUZI, H. AND L. CALOI (2023): "Endogenous firm competition and the cyclicality of markups," *Review of Economics and Statistics*, 1–45.

ANDERSON, E., S. REBELO, AND A. WONG (2018): "Markups Across Space and Time," Working Paper 24434, National Bureau of Economic Research.

ARUOBA, B., A. FERNÁNDEZ, D. GUZMÁN, E. PASTÉN, AND F. SAFFIE (2021): "Monetary Policy Surprises in Chile: Measurement & Real Effects," Working Papers Central Bank of Chile 921, Central Bank of Chile.

BAQAEE, D., E. FARHI, AND K. SANGANI (2021): "The supply-side effects of monetary policy," Tech. rep., National Bureau of Economic Research.

BAQAEE, D. R. AND E. FARHI (2020): "Productivity and misallocation in general equilibrium," *The Quarterly Journal of Economics*, 135, 105–163.

BILS, M., P. J. KLENOW, AND B. A. MALIN (2018): "Resurrecting the role of the product market wedge in recessions," *American Economic Review*, 108, 1118–1146.

BOND, S., A. HASHEMI, G. KAPLAN, AND P. ZOCH (2021): "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data," *Journal of Monetary Economics*.

BURSTEIN, A., V. M. CARVALHO, AND B. GRASSI (2020): "Bottom-up Markup Fluctuations," Working Paper 27958, National Bureau of Economic Research.

CHAMPION, M., C. EDMOND, AND J. HAMBUR (2023): "Competition, Markups, and Inflation: Evidence from Australian Firm-level Data," in *RBA Annual Conference Papers*, Reserve Bank of Australia, acp2023-05.

CHIAVARI, A., D. SMIRNOV, AND M. MORAZZONI (2021): "Heterogeneous markups cyclicality and monetary policy," Tech. rep., Working paper.

DE LOECKER, J. (2021): "Comment on Bond et al (2021)," *Journal of Monetary Economics*.

DE LOECKER, J. AND J. EECKHOUT (2018): "Global market power," Tech. rep., National Bureau of Economic Research.

DE LOECKER, J., J. EECKHOUT, AND G. UNGER (2020): "The rise of market power and the macroeconomic implications," *The Quarterly Journal of Economics*, 135, 561–644.

DE LOECKER, J. AND F. WARZYNSKI (2012): "Markups and firm-level export status," *American economic review*, 102, 2437–71.

DE RIDDER, M., B. GRASSI, G. MORZENTI, ET AL. (2022): "The hitchhiker's guide to markup estimation," Tech. rep.

DE ROUX, N., M. ESLAVA, S. FRANCO, AND E. VERHOOGEN (2021): "Estimating production functions in differentiated-product industries with quantity information and external instruments," Tech. rep., National Bureau of Economic Research.

DHYNE, E., A. PETRIN, V. SMEETS, AND F. WARZYNSKI (2022): "Theory for extending single-product production function estimation to multi-product settings," Tech. rep., National Bureau of Economic Research.

FOSTER, L. S., J. C. HALTIWANGER, AND C. TUTTLE (2022): "Rising Markups or Changing Technology?" Tech. rep., National Bureau of Economic Research.

GALÍ, J. (2015): *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*, Princeton University Press.

GALÍ, J., M. GERTLER, AND D. LOPEZ-SALIDO (2007): "Markups, Gaps, and the Welfare Costs of Business Fluctuations," *The Review of Economics and Statistics*, 89, 44–59.

GOLOSOV, M. AND R. E. LUCAS JR (2007): "Menu costs and Phillips curves," *Journal of Political Economy*, 115, 171–199.

HONG, S. (2017): "Customer Capital, Markup Cyclicality, and Amplification," Working Papers 2017-33, Federal Reserve Bank of St. Louis.

HÖYNCK, C., M. LI, AND D. ZHANG (2023): "The (Unequal) Rise of Market Power, Nominal Rigidities, and Monetary Non-Neutrality," .

JORDA, O. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95, 161–182.

KIMBALL, M. (1995): "The Quantitative Analytics of The Basic Neomonetarist Model," *Journal of Money, Credit Banking*, 27, 1241–1277.

KOUVAVAS, O., C. OSBAT, T. REINELT, AND I. VANSTEENKISTE (2021): "Markups and inflation cyclicality in the euro area," .

LEVINSOHN, J. AND A. PETRIN (2003): "Estimating production functions using inputs to control for unobservables," *The review of economic studies*, 70, 317–341.

MEIER, M. AND T. REINELT (2020): "Monetary policy, markup dispersion, and aggregate TFP," Working Paper 2427, European Central Bank.

——— (2022): "Monetary policy, markup dispersion, and aggregate tfp," *Review of Economics and Statistics*, 1–45.

NAKAMURA, E. AND J. STEINSSON (2008): "Five Facts about Prices: A Reevaluation of Menu Cost Models*," *The Quarterly Journal of Economics*, 123, 1415–1464.

NEKARDA, C. J. AND V. A. RAMEY (2020): "The cyclical behavior of the price-cost markup," *Journal of Money, Credit and Banking*, 52, 319–353.

OLLEY, S. AND A. PAKES (1992): "The dynamics of productivity in the telecommunications equipment industry," .

WOODFORD, M. (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press.

A Heterogeneity by Sector

TABLE 3: Cyclicality of markups by sector

Agriculture	Mining	Manufacturing
-0.175***	-0.0462	0.108***
(0.0186)	(0.0289)	(0.0152)
Electricity	Construction	Retail
-0.0163	0.179***	-0.0212
(0.0596)	(0.0201)	(0.0143)
Transport & ICT	Financial	Other Services
-0.0687***	-0.429***	-0.303***
(0.0163)	(0.0268)	(0.0152)
Observations	7103390	
Adjusted R^2	-0.014	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$