ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

Customer Capital, Markup Cyclicality, and Amplification

Authors	Sungki Hong
Working Paper Number	2017-033B
Revision Date	October 2017
Citable Link	https://doi.org/10.20955/wp.2017.033
Suggested Citation	Hong, S., 2017; Customer Capital, Markup Cyclicality, and Amplification, Federal Reserve Bank of St. Louis Working Paper 2017-033. URL https://doi.org/10.20955/wp.2017.033

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

Customer Capital, Markup Cyclicality, and Amplification

Sungki Hong*

February 25, 2019

Abstract

This paper studies the importance of firm-level price-markup dynamics for business cycle fluctuations. Using state-of-the-art IO techniques to measure the behavior of markups over the business cycle at the firm level, I find that markups are counter-cyclical with an average elasticity of -1.1 with respect to real GDP. Importantly, I find substantial heterogeneity in markup cyclicality across firms, with small firms having significantly more counter-cyclical markups than large firms. Then, I develop a general equilibrium model by embedding *customer capital* (due to deep habits as in Ravn, Schmitt-Grohé, and Uribe, 2006) into a standard Hopenhayn (1992) model of firm dynamics with entry and exit. The calibrated model replicates these empirical facts and produces counter-cyclical firm sales dispersions consistent with the data. The resulting input misallocation amplifies both the volatility and persistence of the aggregate productivity shocks driving the business cycle. (JEL Codes: E31, E32, L11, L13.)

Keywords: Business Cycle, Customer Capital, Entry and Exit, Markup Cyclicality.

*Federal Reserve Bank of St. Louis, Email: sungki.hong@stls.frb.org. I am deeply grateful to my advisers Oleg Itskhoki, Richard Rogerson, Jakub Kastl, and Mark Aguiar for their continual guidance and support. For useful suggestions and comments, I am also indebted to Nobu Kiyotaki, Ben Moll, Dmitry Mukhin, Ezra Oberfield, Esteban Rossi-Hansberg, Chris Sims, and all seminar participants at Princeton University, Federal Reserve Bank of St Louis, Queen's University, USC, HKU, CUHK-Shenzhen, Hanyang University, and Sogang University. The following views are those of the author and do not necessarily reflect the views of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

1 Introduction

Understanding dynamics of price-cost markups is an important issue in business cycle analysis. Chari, Kehoe, and McGrattan (2007) find that movements of the *labor wedge*[1] account for much of the aggregate fluctuations, and Bils, Klenow, and Malin (2016) find that the variation in markups explains at least half of the variation in the labor wedge. Also, the endogenous markup variation is an important propagation mechanism for business cycle models to generate a more volatile employment response. Moreover, markup cyclicality has important implications for policymakers. The standard New Keynesian Philips Curve (NKPC) implies that the dual main objectives of the central bank do not conflict: Stabilizing inflation also stabilizes output at the desired level. But this "divine coincidence" breaks down with cyclical movements in desired markups, as stabilizing inflation no longer ensures zero output gap. Blanchard and Gali (2007) stress that understanding the interactions between the NKPC and variations in desired markups "should be high on macroeconomists' research agendas."

The existing literature has focused on measuring and modeling the behavior of the aggregate (or sectoral) markup. In contrast, in this paper, I measure the markup behaviors at the firm level, and study the role of markups by explicitly modeling firm heterogeneity, and by deriving its implications for aggregate economic outcomes when subjected to business cycle shocks. A key advantage of this approach is that it allows one to assess whether the firm-level choices of markups in the model are consistent with the data. Aggregate measures abstract from heterogeneity in markups and potentially confound changes in markups at the level of the individual firm with changes in the composition of firms.

My paper makes three contributions. Key to this approach is the availability of reliable estimates of markup variability at the firm level. Because such estimates do not exist, the first contribution of my paper is to use the recent IO technique to measure markup behavior at the firm level. Having documented these patterns, I then develop a general equilibrium model with heterogeneous firms that can account for the salient features of markups in the micro data. Third, I explore the implications of this model when subjected to aggregate shocks.

For empirical analysis, I use firm-level data of four large European countries (France, Germany, Italy, and Spain) from Amadeus that covers manufacturing sectors.[2] The dataset contains both production data and financial balance sheet data. However, the dataset pro-

[1] Labor wedge is defined as the ratio between the marginal rate of substitution of consumption for leisure and the marginal product of labor.

[2] I use France as the benchmark country since it has a longer time period (2003 – 2013) available for online download from WRDS than the other three countries.

vides neither the price nor the marginal cost necessary for directly measuring markup. Instead, I follow insights from Hall (1986) and De Loecker and Warzynski (2012), relying on a firm's optimality condition of cost minimization with respect to a static input. Therefore, I use material input for markup estimation because material does not suffer much from adjustment costs and other dynamic considerations (see Levinsohn and Petrin, 2003). Previous studies measure markup with labor inputs, but it is well known that the measurement of true costs of labor inputs suffers from several frictions, including but not limited to hiring and firing costs, overhead costs, etc. (as discussed by Rotemberg and Woodford, 1999). Also, I check robustness for functional forms of production function when estimating the marginal cost. My result suggests strong countercyclicality of markups. I estimate that markups are countercyclical with an average elasticity of -1.1, with respect to real GDP with a Cobb-Douglas functional form, and -0.7 with Translog. In other words, when real GDP decreases by 1%, firms increase their markups by 0.7% - 1.1%. This is in line with the recent findings by Bils, Klenow, and Malin (2018) using US industry-level data. Furthermore, taking full advantage of firm-level data, I find substantial heterogeneity of markup cyclicality among firms. In particular, if I split firms based on their revenue share within own industries, I find that small firms' markups are 60% more countercyclical than large firms'.

Furthermore, exploiting the granularity of my firm-level data, I find that the empirical results above are robust along several dimensions. First, I check if the same results hold in the other three countries, namely, Germany, Italy, and Spain. Although these countries cover shorter time periods than France, the main results stay the same. Next, I check whether the firm size actually approximates other firms' characteristics that matter for heterogeneity. It is possible that the firm size actually reflects firm age since young firms tend to be small and old firms tend to be large. I find that the size effect on markup cyclicality remains significant, controlling for the age effect. The other possibility is that firm size represents a firm's financial health, because large firms probably have less collateral constraints and better relationships with lending institutes. A recent study by Gilchrist, Schoenle, Sim, and Zakrajsek (2015), using samples from Compustat matched with PPI data by BLS, documents that low-liquidity firms raised prices, while high-liquidity firms decreased prices during the recent financial crisis. However, the firm size still matters after firms' financial balance sheet variables are included in the regression. Lastly, notice that the heterogeneity is not driven by sector, since I define small and large firms based on their revenue share within 4-digit sectors.

Next, I explore the economic mechanism that gives rise to these empirical patterns by building a general equilibrium model. In particular, I embed *customer capital* (due to deep habits as in Ravn, Schmitt-Grohé, and Uribe, 2006) into a standard Hopenhayn (1992) model of firm dynamics with entry and exit. The driving force of a business cycle in the economy is the aggregate productivity shock. A key feature of the model with customer capital is that a firm's decision about markups becomes dynamic – higher markups today imply higher profits per unit sold today, but by decreasing the quantity sold lead to lower customer capital in the future. Firms calculate the present value of all future profits with procyclical stochastic discount factors. Thus, the incentive for firms to *invest* in customer capital is high in booms and low in recessions, resulting in countercyclical markups. Also, the endogenous firm's entry and exit decision also affects the firm's pricing. During recessions, the endogenous probability of an exit increases, and this implies that firms place lower weight on future profits, leading them to charge higher markups. Since the exit risk is larger for small firms during downturns, the extensive margin of an exit affects them more, resulting in more countercyclical markups.

Taken together, the model is able to replicate the two main empirical findings of markup cyclicality both qualitatively and quantitatively: 1) Firms have countercyclical markups, and 2) small firms have more countercyclical markups than large firms. The model can explain around one-third to two-thirds of the magnitude of markup elasticity with respect to aggregate output, and it closely captures the heterogeneity of markup cyclicality between small and large firms as estimated in the data. Furthermore, the model could explain 60% of the cross-sectional standard deviation of markup level in the data. This is due to firms' dynamic pricing decisions through different stages of a life cycle.

Interestingly, although the model was only disciplined to explain firm-level markup behaviors, it also captures other features over business cycles. First, the model is able to generate endogenous fluctuations of firm size distribution over business cycles caused by aggregate productivity fluctuations. Heterogenous markup dynamics of firms imply that, during recessions, small firms produce less and shrink relative to large firms by raising their markups, which results in an increase in the dispersion of firm size. This implication has support in the data. In the model, deflated sales of firms in the lowest-size quintile are 17 percentage points more responsive to aggregate output than the highest-size quintile, for which the difference is 22 percentage points in the data. Notice that the model generates the countercyclical dispersion of firm size without relying on an exogenous second-moment shock in productivity growth as in Bloom (2009) and Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2014). Instead, the exogenous process that governs firms' idiosyncratic productivities is time-invariant in my model. What drives the heterogeneous responses of firms is the endogenous probability of staying in the market, which depends on not only the productivity, but also the customer capital that firms have accumulated.

Finally, the other business cycle implication of the model is that the resulting input misallocation amplifies the aggregate productivity. I define the measured Total Factor Productivity (TFP) as the ratio of total output to total input, and it has a non-trivial dynamic in the model. There are two opposing effects that drive the movements of measured TFP relative to the aggregate productivity. The first effect comes from the selection of firms' productivities at the extensive margin (*cleansing effect*). During recessions, the productivity threshold for entrants and exiters increases, hence the mean for continuing firms' productivities increases. This cleansing effect tends to make the measured TFP procyclical. The second effect comes from the misallocation of inputs due to dispersion of firm size, which contributes to the countercyclicality of the measured TFP. In the simulation, upon a negative aggregate productivity shock, the misallocation effect dominates in the beginning. Later on, the selection effect dominates. Hence, the measured TFP initially drops, rebounds, and then overshoots compared to the aggregate productivity. I find that the misallocation effect, which is the main focus of this paper, amplified the standard deviation of the measured TFP by 25%.

Related Literature My work is related to several strands in the literature. First, I add to the extensive literature estimating the cyclicality of markup. Bils (1987), Rotemberg and Woodford (1999), and Nekarda and Ramey (2013) use labor input to test the markup cyclicality at the aggregate level. Other studies infer the cyclicality using inventories (Kryvtsov and Midrigan, 2012), retail prices (Stroebel and Vavra, 2014), and advertising expenses (Hall, 2014). I follow Bils, Klenow, and Malin (2018) to use material inputs, but I estimate the markup at the firm level using IO techniques by De Loecker and Warzynski (2012). To the best of my knowledge, this paper is the first one to use this technique to address the price-cost cyclicality at the firm level. This paper is also related to the literature estimating heterogeneity in markup cyclicality. Bils, Klenow, and Malin (2012) study how markup cyclicality changes based on durability of goods. Chevalier and Scharfstein (1996) examine the pricing behavior of supermarkets with leveraged buyout (LBO). Gilchrist, Schoenle, Sim, and Zakrajsek (2015) document that more financially constrained firms raised prices more during the recent financial crisis. My work exclusively explores the heterogeneity of markup cyclicality due to firm size, which endogenously affects a firm's exit decision over business cycles. I show that the result is robust against alternative economic mechanisms in these earlier works.

Second, the paper relates to models with variable markups. Some business cycle models assume an exogenous markup process, sometimes called "cosh-push" shock, without giving explicit explanations for the underlying economic mechanisms (e.g., Clarida, Gali, and Gertler, 1999; Smets and Wouters, 2003, 2007). Other models try to ground the source of markup variation. In the New Keynesian model, a sticky price with procyclical marginal cost generates a countercyclical markup. In Rotemberg and Saloner (1986), implicit collusion of oligopolistic firms gives rise to variation in markup. In Jaimovich and Floetotto (2008) and Bilbiie, Ghironi, and Melitz (2012), due to firm entry and exit, variation in product variety affects the goods elasticity of substitution, and thus markup. Edmond and Veldkamp (2009) analyze a model where the countercyclical earnings dispersion leads to countercyclical markup. These models explain the movements of aggregate markups, but are either silent to the heterogeneity of cyclicality or have opposite predictions at the firm level. A closely related model to this paper is by Gilchrist, Schoenle, Sim, and Zakrajsek (2015). In their model, firms also have dynamic pricing problems in the customer market. During downturns, more financially constrained firms tend to finance their fixed payments (e.g., operating cost, debt payment) through internal financing by raising markups, but at the cost of future market shares. However, my model shows that, even without financial frictions, if firms face entry and exit decisions at the extensive margin (as in Hopenhayn, 1992; Clementi and Palazzo, 2013; and Clementi, Khan, Palazzo, and Thomas, 2014), heterogeneity of markup cyclicality could still emerge. Hence, my model serves as a complement to their financial friction model.[3] Another related literature on variable markups is the large body of research in international macroeconomics regarding exchange rate pass-through. These studies focus on responses of prices to exchange rate shocks, instead of shocks over business cycles. They find that, due to strategic complementarities in pricing, firms' prices respond sluggishly and heterogeneously to a common exchange rate shock (see, for instance, Atkeson and Burstein, 2008; Amiti, Itskhoki, and Konings , 2014, 2016). In contrast to these studies, a firm's pricing in my model is driven by an intertemporal decision of customer capital accumulation.

Third, the paper speaks to literature on dispersion of firm size and input misallocation. Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2014) document countercyclical dispersion of innovations in firm productivity. Similarly, Kehrig (2015) finds that the *level* of firm productivity is countercyclical. The model in this paper generates countercyclical firm size dispersion without any underlying exogenous second-moment shock in firm productivity. The dispersion occurs endogenously since small firms close to the entry-exit margin respond differently from large incumbents at the other end of the distribution. Previous studies (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009) also stress that the misal-

[3] Moreira (2015) builds a similar model to this paper to study effects of initial business conditions on entrants' sizes. However, there are several differences between her work and mine. First, the focus of my paper is markup cyclicality, while her paper focuses on size of newborn businesses. Second, in terms of methodology, her model is a partial equilibrium model, and the business cycle is driven by an exogenous demand shock. In contrast, I build a general equilibrium model where the aggregate shock comes from aggregate productivity. This key difference allows for a stochastic discount factor, which is the main driver of the markup cyclicality.

location of inputs can lead to a decrease in measured TFP. Instead of imperfection in the input markets, several papers emphasize how the distortion in the output market through markup variation could give rise to inputs misallocation. Peters (2013) studies the effect of endogenous misallocation due to market power on economic growth. Opp, Parlour, and Walden (2014) extend the collusion model of Rotemberg and Saloner (1986) under the general equilibrium framework and analyze the resource misallocation across industries. Edmond, Midrigan, and Xu (2015) explore the pro-competitive gains from trade liberalization quantitatively. In contrast, my paper focuses on the inputs misallocation due to heterogeneity of markup cyclicality along the business cycle, and I find that its effect on measured TFP is substantial in the quantitative exercises.

The remainder of the paper is organized as follows. In Section 2, I document the markup behavior at the firm level. In Section 3, I build a general equilibrium model of firm dynamics and describe its key mechanism. Then, I calibrate the model to match standard moments in the data in Section 4. In Section 5, I show that the model can match the empirical findings and its business cycle implications. I then conclude in Section 6.

2 Empirical Study of Markup Cyclicality

In this section, I describe the estimation strategy for firm-level markup and the data I use for empirical analysis. Then, I show the empirical findings about markup dynamics and several robustness checks.

2.1 Estimation of Markup

For firm i at time t, I define the markup as price over marginal cost:

$$\mu_{it} = \frac{P_{it}}{MC_{it}}. \tag{1}$$

In practice, we seldom observe price and marginal cost together in data. To overcome this problem, I follow the insights from Hall (1986) and De Loecker and Warzynski (2012), that rely on a firm's cost minimization condition with respect to static inputs not subject to adjustment frictions. I closely follow De Loecker and Warzynski's (2012) approach, which involves estimations of both production function and markup at the firm level. However, this methodology still suffers from the lack of information in the data, especially the firm-level price data. I will discuss the potential issues for the estimation caused by this missing information.

In particular, consider a firm i at time t that produces gross output Y_{it} with production function:

$$Y_{it} = F_{it}(M_{it}, L_{it}, K_{it}, A_{it}),$$

where M_{it} is material input, L_{it} is labor input, K_{it} is capital input, and A_{it} is firm productivity. For a given level of output Y_{it}, its associated Lagrange function with cost minimization is

$$\mathcal{L}(M_{it}, L_{it}, K_{it}, \lambda_{it}) = P_{it}^{M} M_{it} + W_{it} L_{it} + R_{it} K_{it} + \lambda_{it}(Y_{it} - F_{it}(M_{it}, L_{it}, K_{it}, A_{it})),$$

where P_t^M, W_{it}, and K_{it} are input prices for material, labor, and capital, respectively, and λ_{it} is the Lagrange Multiplier associated with the output constraint. Economically, it stands for marginal cost of production for a given level of output Y_{it}. Largely, material is categorized as static input, while labor and capital are categorized as dynamic inputs. A static input is free of any adjustment costs and a firm's dynamic decisions. The firm's optimality condition for the material input M_{it} is

$$P_{it}^{M} = \lambda_{it} \frac{\partial F_{it}}{\partial M_{it}}.$$

Since markup is equal to $\mu_{it} = \frac{P_{it}}{\lambda_{it}}$ as defined above, rearranging the optimality condition yields the following:

$$\mu_{it} = \theta_{it}^{M} \left(\alpha_{it}^{M}\right)^{-1}, \tag{2}$$

where $\theta_{it}^{M} = \frac{\partial \log F_{it}}{\partial \log M_{it}}$ is output elasticity with respect to material input, and $\alpha_{it}^{M} = \frac{P_{it}^{M} M_{it}}{P_{it} Y_{it}}$ is material cost share of revenue. This states that variation in markup is driven by the output elasticity and the input share. In data, we can easily observe the input share, but we need to estimate the production function to recover the output elasticity. And the output elasticity depends on the assumption of production function. For empirical analysis, I use two functional forms popularly used in the literature and discuss their pros and cons below.

In contrast, a firm's decision about dynamic inputs involves solving the full dynamic problem. For example, if there are adjustment costs for labor, such as hiring and firing costs, a firm has to consider how a labor decision affects not only today's outcome, but also the future's expected profits. Hence, the wedge between the output elasticity and labor share captures an additional wedge caused by the dynamic decisions. In previous macroeconomics literature, researchers rely on labor input for markup estimation and conclude that markup is procyclical. However, the results might be biased due to the wedge caused by the dynamic nature of labor inputs. Rotemberg and Woodford (1999) show that, controlling for adjustment costs, the markup is actually countercyclical.

Due to the reasons stated above, I use material input to examine markup cyclicality. Furthermore, I run my analysis at the firm level, which is different from previous analyses conducted at the industry level. Industry-level markup is a cost-weighted average markup, and thus the results include the compositional effect of firm sizes. However, as I discuss below, it is important to examine markup cyclicality at the firm-level because it helps us understand and find the underlying economic mechanisms that generate the empirical patterns. Also, finding the correct mechanism is important to understand its welfare implications.

Production Function

In general, the output elasticity θ_{it} depends on the form of production function. For the estimation, I restrict to two functional forms of production function commonly used in the literature: Cobb-Douglas and Translog. For the case of the Cobb-Douglas form, the production function has the following form:

$$y_{it} = \theta_m m_{it} + \theta_l l_{it} + \theta_k k_{it} + a_{it} + \epsilon_{it}, \tag{3}$$

where lower case letters denote the natural logarithm of variables, and ϵ_{it} is a production shock to the firm or classical measurement error. Note that all θ output elasticity terms are constants. If there is no production shock ϵ_{it} to the firm, then the markup is:

$$\mu_{it} = \theta_m \left(\alpha_{it}^M\right)^{-1}. \tag{4}$$

In this case, the change in markup is equivalent to the change in inverse of material share, hence there is no need to θ_m. But if a ϵ_{it} shock hits the firm or is a classical measurement error, the markup expression needs to be adjusted to the following:

$$\mu_{it} = \theta_m \left(\alpha_{it}^M \exp\left(\epsilon_{it}\right)\right)^{-1}. \tag{5}$$

This assumption requires the estimates of production elasticities to calculate the residual ϵ_{it}. This calculation also recovers the levels of the markups.

However, this choice of functional form is restrictive. It does not allow any nonlinearity of or interactions among inputs. Hence, I consider the case of the Translog form, which allows for more flexibility in terms of the production function estimation. Specifically, the Translog production function is:

$$\begin{aligned} y_{it} =& \theta_m m_{it} + \theta_l l_{it} + \theta_k k_{it} + \theta_{mm} m_{it}^2 + \theta_{ll} l_{it}^2 + \theta_{kk} k_{it}^2 \\ &+ \theta_{ml} m_{it} l_{it} + \theta_{mk} m_{it} k_{it} + \theta_{lk} l_{it} k_{it} + a_{it} + \epsilon_{it}. \end{aligned} \tag{6}$$

In this case, the output elasticity is a function of inputs, and markup becomes

$$\mu_{it} = (\theta_m + 2\theta_{mm} m_{it} + \theta_{ml} l_{it} + \theta_{mk} k_{it}) \left(\alpha_{it}^M \exp(\epsilon_{it})\right)^{-1}, \tag{7}$$

hence ignoring the nature of time-varying elasticity might bias the result of markup cyclicality. However, a simple OLS regression estimate of production elasticities is biased, due to the unobserved productivities a_{it} in the error term. To overcome this issue, I closely follow the IO techniques developed by Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves, and Frazer (2015). For details of implementation, please see Appendix A. Notice that the flexibility of Translog comes with a price. In my data set, like most other data sets, I do not have firm-specific price deflator, rather I deflate the firm revenue with an industry-level price index. Hence, the firm-specific price deflator enters the error term and potentially causes bias in the estimation. Nevertheless, the Translog function provides a robustness check of my result.

2.2 Data

The firm-level data in manufacturing sectors for four large European countries comes from Bureau Van Dijck's (BvD) Amadeus dataset.[4] This dataset has recently been used in many studies. I focus on the manufacturing sector for it suits the estimation of production function. I only use the firm labeled the "unconsolidated" to avoid double-counting its sales with sales in its parent company. The dataset contains 12,392 firm observations on average each year from 2003 to 2013.

I now define the variables used in the empirical analysis. An *industry* is defined as a 4-digit NACE Rev. 2 code. I measure a firm's *revenue* $P_{it}Y_{it}$ with operating revenue (gross output). Since the dataset does not have information about firm-level price P_{it}, I instead use the 2-digit industry-level price deflator P_t to measure the *output* Y_{it}. I measure the *material cost* $P_{it}^M M_{it}$ with total material costs and deflate it with aggregate material price P_t^M to get *material input* M_{it}. I use the number of employees as measure of *labor input* L_{it}. I measure the *capital input* K_{it} with tangible fixed assets. To control for the input quality across firms in a production function estimation, I have included the firm-specific average wage, defined as the ratio of total labor costs and total number of employees. Finally, I estimate the production function coefficients at a 2-digit industry level.

I take business cycle measures from Eurostat, and Federal Reserve website. We use four measures of a business cycle: (1) real GDP, (2) total employment, (3) total working hours,

[4] The four countries include France, Germany, Italy, and Spain. I use France as my baseline case since other countries have a shorter time-period available for online download.

and (4) recession period defined by NBER. All business cycle variables are quadratically detrended.

2.3 Empirical Modeling Strategy

To empirically test the markup cyclicality, I bring the following specification into the data:

$$\log \mu_{it} = \alpha_i + \phi_0 \log Y_{it} + \log Y_{it} \times \Omega'_{it}\phi_1 + \chi'_{it}\beta + \epsilon_{it}, \tag{8}$$

where μ_{it} is firm-level markup, α_i is a firm's fixed effect, Y_{it} is a business cycle indicator, Ω'_{it} is a vector of a firm's characteristic variables, and χ_{it} is a vector of control variables. ϕ_0 captures the average markup cyclicality for all firms. ϕ_1 captures the heterogeneity in markup cyclicality across firms.

However, the fixed-effect (FE) regression above is not efficient if ϵ_{it} has serial auto-correlation, which is highly likely for time-series data. Hence, to control for the potential auto-correlation in ϵ_{it}, I run the following first-difference (FD) regression:

$$\Delta \log \mu_{it} = \phi_0 \Delta \log Y_{it} + \log \Delta Y_{it} \times \Omega'_{it}\phi_1 + \Delta\chi'_{it}\beta + \Delta\epsilon_{it}. \tag{9}$$

Note that I take first difference in all terms except for the term that captures the heterogeneity in markup cyclicality Ω'_{it}.

2.4 Empirical Results

Inverse of Material Share

I begin my empirical analysis by measuring the markup with the inverse of material share. The underlying assumption is that the specification of the production function is Cobb-Douglas and that there is no production shock. I evaluate the markup cyclicality with fixed-effect regression (8). I do not estimate the production function with this specification. Instead, I control the potential differences in output elasticity with firm fixed effects. Table 1 reports the results for this specification. Column (1) shows that markup is countercyclical and statistically significant for all firms. The elasticity of markup with respect to aggregate output is -1.1. This means that a one-percent increase of output from its non-linear trend causes the markup to decline by 1.1 percent.[5]

Moreover, with micro-level data, I find that there is substantial heterogeneity in markup cyclicality. One interesting and important aspect to look at is whether the markup cyclicality

[5] Regression results with other business cycle measures are qualitatively similar.

Table 1: Heterogeneity in Markup Cyclicality ϕ: Fixed Effects Regression

Dep. Variable	(1)	(2)	(3)	(4)	(5)
$\log \mu_{it}$	All	Large	Small	All	All
$\log Y_t$	-1.13	-0.76	-1.21	-1.21	-1.22
	(0.17)	(0.21)	(0.16)	(0.16)	(0.16)
$\log Y_t \times \text{Large}_i$				0.45	
				(0.12)	
$\log Y_t \times s_{it}$					3.52
					(1.61)
Firm FE	Yes	Yes	Yes	Yes	Yes
N	96507	16048	80459	96507	96507
adj. R^2	0.89	0.87	0.89	0.89	0.89

Note: μ_{it} is the inverse of material share of revenue. $\log Y_t$ is log aggregate output quadratically detrended. $Large_i$ is an indicator for a firm with more than 1% average market share within a 4-digit industry. s_{it} stands for market share in a 4-digit industry. All specifications include firm fixed effects (FEs). The standard errors are clustered at the time level and are reported in parentheses.

depends on firm size. Markups of small and large firms potentially move differently for several reasons: market power, pricing frictions, etc. To explore the heterogeneity, I split firms into small and large firms based on their sizes. I define a large firm as one with more than 1% of market share within a 4-digit industry. Columns (2) and (3) report the results for a subsample of large and small firms, respectively. I find that large firms' markup elasticity is small, equal to -0.8, and small firms have a larger elasticity of -1.2. This implies that small firms' markups fluctuate 60% more than large firms' along the business cycle. Recently, Bils, Klenow, and Malin (2018) used intermediate input to estimate markup cyclicality at the industry level and find that elasticity with respect to real GDP is equal to -0.9. My finding with the large firms is largely comparable to this number, since large firms drive most of the industry-level markup, by its definition. Another way to identify differential effects between small and large firms is by pooling all firms and interacting aggregate output Y_t with a dummy variable Large_i, as in column (4). The result shows that the difference is significant. Alternatively, instead of using a dummy variable, I interact aggregate output with market share $s_{it} = \frac{p_{it}y_{it}}{\sum_i p_{it}y_{it}}$. The results in column (5) indicate the same findings.

In a panel regression, the error term ϵ_{it} in the fixed-effect specification is likely to follow an AR(1) process, and the resulting serial auto-correlation could bias the results above. Hence, I use first-difference regression for a robustness check. Another advantage of using the first-difference regression is that the constant output elasticity term under Cobb-Douglas specification drops out by first-differencing the markup. The results are reported in the

Table 2: De Loecker-Warzynski Markup Estimates

Moments	(1) Cobb-Douglas	(2) Translog
Mean	1.58	1.54
Std. Log	0.52	0.60
p5	0.58	0.10
p25	0.83	0.86
p50	1.13	1.15
p75	1.68	1.65
p95	4.01	4.88
p75/p25	2.01	1.92

Note: This table shows the summary statistics of firm-level markup estimations based on De Loecker and Warzynski (2012). Column 1 shows the results with Cobb-Douglas specification, while column 2 shows the results with Translog specification.

Appendix. I find the same results as with fixed-effect regression: Markup is countercyclical, and small firms' markups fluctuate more.

De Loecker - Warzynski Estimates

I use the De Loecker - Warzynski (DLW) method to calculate the markups under both Cobb-Douglas and Translog specifications. The results are reported in Table 2. The results are very similar under both speifications. The mean markups are 1.58 and 1.54, and the standard deviations of log are 0.52 and 0.60 for Cobb-Douglas and Translog, respectively. The distribution by percentiles is also very similar.

I report the fixed-effect regression results in Table 3 for DLW estimates. The results of the Cobb-Douglas specification in the first two columns are very close to the results with the inverse of material share. The average markup cyclicality is -1.1, and large firms have less countercyclical markups. This suggests that the potential biases caused by production shock might not be very large.

Next, although Cobb-Douglas is a convenient assumption for an estimation, it has a strong restriction in terms of specification that output elasticities are independent of input usages. And one could wrongly attribute the change in production technology to the change in markup. To depart from Cobb-Douglas, I assume the Translog production function, which allows for higher order and interaction terms of inputs. I use this alternative specification for a robustness check.

I report the results for Translog in columns 3 and 4. The results are consistent with

Table 3: Heterogeneity in Markup Cyclicality ϕ: De Loecker-Warzynski Estimates

	Cobb-Douglas		Translog	
Dep. Variable	(1)	(2)	(3)	(4)
$\log \mu_{it}$	All	All	All	All
$\log Y_t$	-1.06	-1.17	-0.65	-0.82
	(0.21)	(0.19)	(0.13)	(0.18)
$\log Y_t \times s_{it}$		4.97		9.87
		(1.65)		(3.41)
Firm FE	Yes	Yes	Yes	Yes
N	63899	63899	61054	61054
adj. R^2	0.89	0.89	0.85	0.85

Note: μ_{it} is the firm-level markup estimated with the De Loecker-Warzynski (2012) method. Columns (1) and (2) show results with a Cobb-Douglas production specification, and columns (3) and (4) show results with a Translog production specification. $\log Y_t$ is log aggregate output quadratically detrended. s_{it} stands for market share in a 4-digit industry. The standard errors are clustered at the time level and are reported in parentheses.

main findings under the Cobb-Douglas specification, but there are two differences in terms of magnitude. First, the markup is less countercyclical under Translog. It implies that firms actually substitute among the inputs along the business cycle, possibly due to different factor prices, adjustment costs, etc. Second, the cyclicality differences between small and large firms are larger. This means that firms adjust their input for production differently. For example, a firm will be more reluctant to change its labor inputs if it faces greater adjustment frictions in labor than other firms, which would cause the differences in the output elasticity.

Robustness Check

Here, I investigate whether the results above are robust to alternative samples. First, I check whether the results still hold in other countries. In the Amadeus dataset, there are other countries available for download, but they have shorter time periods available for online download from WRDS. Nevertheless, I choose Germany, Italy, and Spain for a robustness check and find that results are similar. Results are reported in the Appendix.

Next, I investigate if the firm size actually captures other characteristics of firms. Young firms tend to be small, and old firms tend to be large. Hence, the size effect on markup cyclicality could be due to age effect. Also, a recent study by Gilchrist, Schoenle, Sim, and Zakrajsek (2015) finds that financially constrained firms raised markups more to avoid using costly external financing to pay a fixed payment of operation during the Great Recession.[6] And firm size might be proxying the financial conditions of the firms. I check these by adding additional interaction terms between aggregate output and age and financial variables, respectively. Results in the Appendix suggest that the firm size still plays a prominent role in determining the markup cyclicality after controlling for these factors.

3 Model

In this section, I develop a general equilibrium model that is able to quantitatively match the empirical findings of markup dynamics in the previous section.[7] As in other standard RBC models, the business cycles are driven by exogenous shocks to aggregate productivities.

Specifically, I embed *customer capital* into a standard Hopenhayn's (1992) endogenous firm entry and exit model. In the model, the customer capital determines the level of demands of firms' outputs. To enlarge its customer capital, a firm can sell more of its products today to gain more market shares in the future. In other words, a firm sees the product sales as a form of investment into customer capital. Some earlier works in the literature capture this idea in the model. In Rotemberg and Woodford (1991), firms compete to gain market shares by lowering markups. In Klemperer (1995), customers purchasing from one firm have switching costs to a competitor firm. On one hand, firms have incentives to lock in customers by lowering its price. On the other hand, they have some degree of market power over their current customers. To capture this investment in the customer capital idea in a tractable way, I adopt the framework of *deep habits* model developed by Ravn, Schmitt-Grohé, and Uribe (2006), in which households have good-specific habit formations. However, even though the deep habits model has the advantage of tractability, it possibly has different welfare implications from other models and needs modifications when compared to the data. In the quantitative analysis, I will discuss this welfare measurement bias caused by the good-specific habit formation. Moreover, firms face entry and exit problems as in Hopenhayn (1992).

This simple setup is able to generate heterogenous markup cyclicality. In the model, a firm always faces a trade-off between *invest* and *harvest* motives. On one hand, a firm wants to lower its price to attract customers to invest in its own customer capital. On the other hand, since a firm has a certain degree of market power over the locked-in customers, it wants to raise its price to harvest the profit. During recessions, since small firms are more likely

[6] An earlier study by Chevalier and Scharfstein (1996) finds that more financially constrained supermarkets have more countercyclical markups.

[7] There are alternative ways of modeling the firm dynamics. In the Appendix, I discuss two prominent models in the literature: (i) an oligopolistic competition model, and (ii) a New Keynesian model. I show that they are not consistent witht the micro-level evidence.

to exit the market, they put less weight on the future benefit of the customer capital and increase the price to exploit the customers as much as possible before exiting the market.

3.1 Household

Time is discrete and is indexed by $t = 1, 2, 3, ...$ and is of infinite horizon. A representative household consumes a variety of products produced by a continuum of firms indexed by $i \in [0, 1]$. The household maximizes the discounted expected utility

$$\mathrm{E}_t \sum_{\tau=0}^{\infty} \beta^{\tau} \left[\frac{\tilde{C}_{t+\tau}{}^{1-\sigma}}{1-\sigma} - \omega \frac{L_{t+\tau}{}^{1+\nu}}{1+\nu} \right], \tag{10}$$

where $\tilde{C}_t$ is habit-adjusted consumption bundle, and L_t is labor. The household has constant relative risk aversion σ, disutility of labor ω, and inverse Frisch elasticity of labor supply ν.

The household consumes out of a continuum of firms, each producing a specific variety $i \in I$. For any given time, only a subset of varieties $I_t \subset I$ are available for consumption. The habit-adjusted Dixit-Stiglitz consumption aggregator is defined as

$$\tilde{C}_t = N_t^{\xi - \frac{1}{\rho-1}} \left[\int_{i \in I_t} \left(c_{it} b_{it}^{\theta} \right)^{\frac{\rho-1}{\rho}} di \right]^{\frac{\rho}{\rho-1}}, \tag{11}$$

where c_{it} denotes a household's consumption of variety i. N_t is the aggregate measure of available varieties at time t, and ξ measures the degree of love-for-variety.[8] b_{it} is variety i's habit stock. The household takes variety-specific habit stocks as given. Namely, the household follows the "Keeping up with the Joneses" behavior in consumption.[9] External habit stock b_{it} evolves according to

$$b_{it+1} = (1-\delta) b_{it} + \delta c_{it}, \tag{12}$$

where $\delta \in [0, 1]$ governs both depreciation rate of past habit stock and conversion rate of consumption into habit stock. Finally, $\rho > 1$ governs the intratemporal elasticity of substitution across habit-adjusted consumption, and θ governs the degree of habit formation.

A household's demand for variety i is given by:

$$c_{it} = N_t^{\xi(\rho-1)-1} \left(\frac{p_{it}}{\tilde{p}_t} \right)^{-\rho} b_{it}^{\theta(\rho-1)} \tilde{C}_t, \tag{13}$$

[8] $\xi = 0$ corresponds to the case of consumers not having any preference over variety, while $\xi = \frac{1}{\rho-1}$ to the case of the usual Dixit-Stiglitz preference.

[9] Nakamura and Steinsson (2011) study the case in which the household takes habit formation as internal and find that firms face a time-inconsistency problem.

where p_{it} is product i's price, and $\tilde{p}_t$ is habit-adjusted price index:

$$\tilde{p}_t = N_t^{-\xi+\frac{1}{\rho-1}} \left[\int_{i \in I_t} \left(\frac{p_{it}}{b_{it}^{\theta}} \right)^{1-\rho} di \right]^{\frac{1}{1-\rho}}. \tag{14}$$

The household is subject to the following budget constraint:

$$\tilde{p}_t \tilde{C}_t + \mathrm{E}_t(q_{t,t+1} D_{t+1}) = D_t + w_t L_t + \Phi_t, \tag{15}$$

where D_{t+1} is Arrow-Debreu securities, w_t is nominal wage, and Φ_t is a household's share of firms' profits and operating costs.

With all the specifications and constraints above, a household's optimization problem yields the following two FOCs:

$$\frac{w_t}{\tilde{p}_t} = \omega L_t^{\nu} \tilde{C}_t^{\sigma}, \tag{16}$$

$$q_{t,t+\tau} = \beta^{\tau} \left(\frac{\tilde{C}_{t+\tau}}{\tilde{C}_t} \right)^{-\sigma} \frac{\tilde{p}_t}{\tilde{p}_{t+\tau}}, \tag{17}$$

which are a household's intratemporal trade-off between consumption and leisure, and the intertemporal deicision, respectively.

3.2 Firm

Incumbent

Incumbent i produces output y_{it} with the following production technology:

$$y_{it} = A_t a_{it} l_{it}^{\alpha}, \tag{18}$$

where A_t is aggregate productivity to all firms, a_{it} is firm-specific productivity, and l_{it} is labor input for production. α controls the degree of returns-to-scale in production technology.[10]

[10] Aggregate productivity A_t is the source of aggregate fluctuation in my model. But note that the result of my model does not depend on whether it is a supply-side shock or demand-side shock. I have constructed a partial equilibrium model with an exogenous aggregate demand shock, and all the results in this paper hold. This is in contrast to a New Keynesian model, in which the markup is counter-cyclical to a demand-side shock, while pro-cyclical to a supply-side shock.

A_t and a_{it} follow the AR(1) processes respectively:

$$\log A_t = \rho_A \log A_{t-1} + \varepsilon_t^A, \tag{19}$$

$$\log a_{it} = \rho_a \log a_{it-1} + \varepsilon_{it}^a, \tag{20}$$

where both innovations ε_t^A and ε_{it}^a are drawn from normal distribution with standard deviation σ_A and σ_a, respectively. Denote $\Pi_A(A_{t+1}|A_t)$ and $\Pi_a(a_{t+1}|a_t)$ as conditional distribution of aggregate and idiosyncratic productivity, respectively.

At all $t \geq 0$, the distribution of incumbents over the two dimensions of idiosyncratic productivities and customer capitals is denoted by $\Gamma_t(b, a)$. Note that the mass of incumbents N_t is the integral of the distribution, $N_t = \int\int d\Gamma_t(b, a)$. And let Λ_t denote the vector of aggregate state variables, and its transition operator is $B(\Lambda_{t+1}|\Lambda_t)$. In the later section, I show that $\Lambda_t = \{A_t, \Gamma_t\}$.

In each period t, incumbent i faces demand (13) and takes into account that the habit stock b_{it} evolves as (12). After production, a firm draws an operating cost ψ from distribution G and decides whether to pay the cost to continue operating. If the firm decides to exit the market, it scraps the exit value and can not reenter the market in the future. For ease of calculation, ψ is measured in terms of habit-adjusted aggregate consumption. Firms discount each period's profit with a one-period stochastic discount factor $q(\Lambda, \Lambda')$, which depends on the current aggregate state Λ and the future aggregate state Λ'. Also, firms take the habit-adjuted price index $\tilde{p}(\Lambda)$ and the wage $w(\Lambda)$ as given, where both variables depend on the current aggregate state Λ.

Given customer capital b, idiosyncratic productivity a, and aggregate state Λ, an incumbent's dynamic problem is

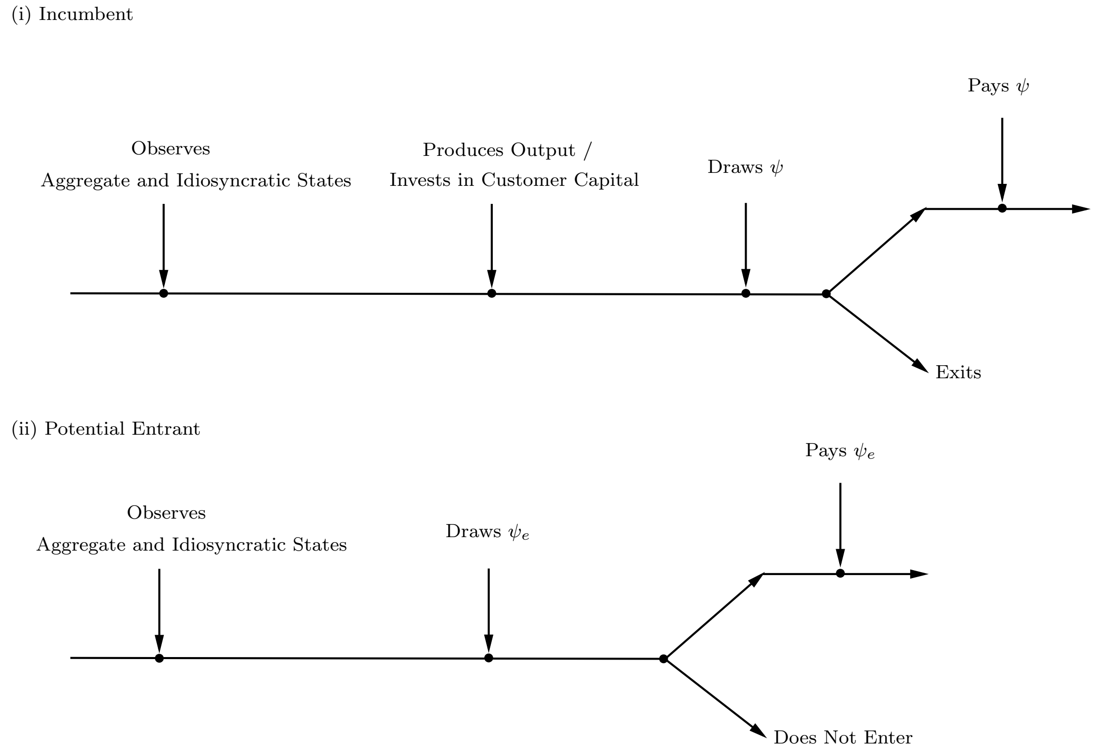
$$\begin{aligned} V(b, a, \Lambda) = \max_{y,l,p,b'} \; & py - w(\Lambda)l + \int_{\psi} \max\left\{0, -\tilde{p}(\Lambda)\psi + \mathrm{E}_{a,\Lambda}\left[q(\Lambda, \Lambda')V(b', a', \Lambda')\right]\right\} dG(\psi) \\ & \text{s.t. constraints (12), (13), and (18) satisfy.} \end{aligned} \tag{21}$$

Without loss of generality, I normalize the exit value to be zero. An incumbent decides to stay in the economy if and only if $\psi \leq \psi^*$, where ψ^* is a firm-specific threshold value of the operating cost implicitly defined by

$$\tilde{p}(\Lambda)\psi^*(b', a, \Lambda) = \mathrm{E}_{a,\Lambda}\left[q(\Lambda, \Lambda')V(b', a', \Lambda')\right]. \tag{22}$$

Since optimal choice of the next period's customer capital b' is based on state variables $\{b, a, \Lambda\}$, the threshold value can also be written as $\psi^*(b, a, \Lambda)$.

Figure 1: Timing of a Firm's Decision in Period t



The mass of exiting firms at time t is

$$N_t^x = \int \int \Pr\left[\psi \geq \psi^*(b, a, \Lambda)\right] d\Gamma_t(b, a), \tag{23}$$

where x denotes *exit*. And the mass of incumbents that continue to the next period is

$$N_t^c = N_t - N_t^x, \tag{24}$$

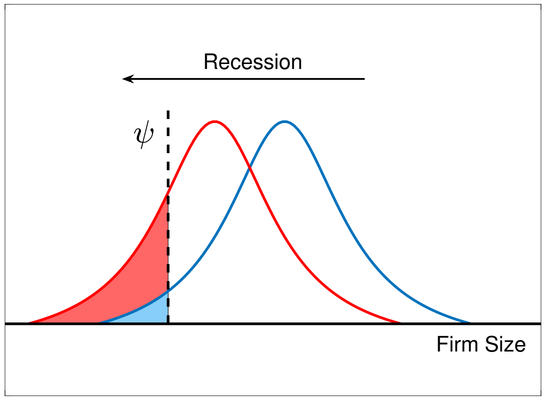
where c denotes *continuation*.

Markup Dynamics

To help understand the intuition of the model, I write down the equation that summarizes the markup dynamics below (for full derivation of FOCs for incumbent's problem, please see Appendix B):

$$\mu_{it}^{-1} - \bar{\mu}^{-1} = G(\psi_{it}^*)\mathrm{E}_t \left\{ q_{t,t+1} \left[(1-\delta)\frac{p_{it+1}}{p_{it}} \left(\mu_{it+1}^{-1} - \bar{\mu}^{-1}\right) + \delta \frac{\theta(\rho-1)}{\rho} \frac{p_{it+1} y_{it+1}}{p_{it} b_{it+1}} \right] \right\}, \tag{25}$$

Figure 2: Shift of Firm Size Distribution in a Recession



Note: Blue and red bell curves plot the firm size distribution in a steady state and a recession, respectively. ψ is the operating cost. Blue and red shaded areas indicate the measure of firms exiting the market.

where μ_{it} is firm i's price markup over marginal cost, $\bar{\mu} \equiv \frac{\rho}{\rho-1}$ is a monopolistic firm's constant price markup, $G(\psi_{it}^*)$ is firm i's probability of staying in the market in the next period $t+1$, and $q_{t,t+1}$ is the stochastic discount factor. Notice that when customer capital stays at a constant ($\delta = 0$), or there is no good-specific habit formation ($\theta = 0$), then the firm always charges a monopolistic markup in a standard economy $\mu_{it} = \bar{\mu}$. In contrast, when a firm faces a customer market, its markup is always below the monopolistic markup $\mu_{it} < \bar{\mu}$. This is because a firm has an incentive to *invest* in its customer capital, hence it does not fully exploit its market power and sets a lower markup than $\bar{\mu}$. Hence, in the equilibrium, whenever a firm raises its markup, a firm always has an option to increase its markup close to the level of a monopolist to boost the short term profits.

To help pin down what factors affect the movements of markups, I consider a special case where the lagged customer capital fully depreciates $\delta = 1$. In this case, the customer capital in the next period is equal to the current output $b_{it+1} = y_{it}$ according to (12), and (25) simplifies to:

$$\mu_{it} = \bar{\mu} \left\{ 1 + \theta G(\psi_{it}^*) \mathrm{E}_t \left(q_{t,t+1} \frac{p_{it+1} y_{it+1}}{p_{it} y_{it}} \right) \right\}^{-1}. \tag{26}$$

Notice that there are three factors that come into play: (i) the probability of staying in the market in the next period $G(\psi_{it}^*)$; (ii) the stochastic discount factor $q_{t,t+1}$; and (iii) the revenue growth rate of the firm $\frac{p_{it+1} y_{it+1}}{p_{it} y_{it}}$. In the original *deep habits* model by Ravn, Schmitt-Grohe, and Uribe (2006), the second factor is the main driver of markup cyclicality. Given a procyclical stochastic discount factor, a firm values the benefit of extra future customer capital more in booms rather than in recessions, thus resulting in countercyclical markups.

Table 4: Heterogeneity in Markup Cyclicality: Exit Indicator

Dep. Variable $\log \mu_{it}$	(1)	(2)
$\log Y_t$	-1.27	-1.26
	(0.16)	(0.15)
$\log Y_t \times s_{it}$	3.33	3.32
	(1.03)	(1.04)
$\log Y_t \times \text{EXIT}_{it+1}$		-0.29
		(0.14)
Firm FE	Yes	Yes
N	56445	56445
adj. R^2	0.92	0.92

Note: μ_{it} is the firm-level inverse of material share of revenue. $\log Y_t$ is log aggregate output quadratically detrended. s_{it} stands for market share in a 4-digit industry. $EXIT_{it}$ is the dummy variable for a firm exit, which is equal to one if a firm exits in the next period. The standard errors are clustered at the time level and are reported in parentheses.

Since the stochastic discount factor is common across all firms, it affects all firms by the same magnitude.

However, in my model, I want to emphasize two other channels that bring heterogeneity into markup dynamics. First, the movement of $G(\psi^*_{it})$ is heterogeneous among firms. Figure 2 illustrates the difference in exit probability for small and large firms in a recession. In downturns, small firms are less likely to stay in the market than are large firms. Hence, even discounting the future profits with the same stochastic discount factors, small firms have greater incentives to exploit their current customers with high markups. Maintaining a large customer capital is an attractive option only if firms continue operations in the market. Benefit of maintaining their own customer base is low if they do not stay in the market any more. To test this channel empirically in the data, I include an interaction term of output and a dummy variable for exit in the fixed-effect regression (8). Exit is defined as the year of the firm's final appearance in the data. For this empirical analysis, I restrict my sample to years 2005 – 2010 for France. The reasons for not using the whole sample period are twofold. First, the data tend to drop a firm that does not report for a certain period, which creates potential survival bias. Second, any online download of the data (WRDS) caps the number of firms to be downloaded, which makes an exit dummy a noisy measure. I choose years 2005 to 2010 because this time period has an average exit rate close to the estimates in Bellone, Musso, Nesta, and Quere (2006). The results are reported in Table 4. Column (1) shows the baseline results from Section 2. Column (2) shows the results with interaction of output and exit dummy, and we see that the coefficient is negative and significant at the 5% level.

Second, the revenue growth rate $\frac{p_{it+1}y_{it+1}}{p_{it}y_{it}}$ is different for firms in general. Cross-sectionally, firms with high expected growth rates charge low markups to quickly attain optimal size of customer capital. Over business cycles, small firms tend to grow faster after recessions, and *vice versa*, contributing to more procyclical markups for them. However, the effect of this channel is quantitatively small. I have calibrated a model that features heterogenous firms but without an entry and exit, and I find that the difference of markup cyclicality among firms is small.

Entrant

Every period, there is a mass of prospective entrants M_t:

$$M_t = 1 - N_t. \tag{27}$$

The total number of varieties in the economy is fixed and is normalized to one here. Only one firm can produce a good of variety i in the economy. A potential entrant can only enter into the production line of variety i absent of an incumbent. [11]

Before entering the market, a potential entrant draws a signal $\tilde{a}$ about its idiosyncratic productivity if it becomes an incumbent. $\tilde{a}$ is drawn from the distribution $h(\tilde{a})$. Foster, Haltiwanger, and Syverson (2008) find that the mean of entrants' physical productivities are not different from incumbents. Hence, I assume that $h(\cdot)$ is the stationary distribution of an incumbent's idiosyncratic productivity AR(1) process, $\log \tilde{a} \sim N(0, \sigma_a/(1-\rho_a^2))$. And its actual idiosyncratic productivity upon entry a' follows the same process as an incumbent's AR(1) process: $\log a' = \rho_a \log \tilde{a} + \varepsilon^a$. To enter the economy in the next period, it has to pay an entry cost ψ_e, which is drawn from a distribution denoted as $G^e(\psi_e)$. All entrants are endowed with the same level of initial customer capital b_0.

Given a draw of entry cost ψ_e from $G^e(\cdot)$, initial customer capital b_0, signal $\tilde{a}$, and aggregate state Λ, a potential entrant's problem can be written as

$$V_e(\psi_e, b_0, \tilde{a}, \Lambda) = \max\left\{0, -\tilde{p}(\Lambda)\psi_e + \mathrm{E}_{\tilde{a},\Lambda}\left[q(\Lambda,\Lambda')V(b_0, a', \Lambda')\right]\right\}, \tag{28}$$

A potential entrant enters the economy if and only if the entry cost ψ_e is less than or equal

[11] There is a one-to-one mapping between variety and a firm. Clementi, Khan, Palazzo, and Thomas (2014) use the same assumption. They assume that there is a fixed stock of blueprints in the economy, and only a blueprint not used by an incumbent can be used by an entrant. This assumption allows depletion effects of incumbents.

to a threshold $\psi_e^*(b_0, \tilde{a}, \Lambda)$, which is implicitly given by the following:

$$\tilde{p}(\Lambda)\psi_e^*(b_0, \tilde{a}, \Lambda) = \mathrm{E}_{\tilde{a},\Lambda}\left[q(\Lambda, \Lambda')V(b_0, a', \Lambda')\right]. \tag{29}$$

And since all entrants have the same initial customer capital b_0, for any $t \geq 0$, denote an entrant's distribution as $\Gamma_t^e(a')$. And the mass of actual entrants into the next period is

$$N_{t+1}^e = M_t \int \Pr[\psi_e \leq \psi_e^*(b_0, \tilde{a}, \Lambda)] dG^e(\psi_e). \tag{30}$$

3.3 Recursive Competitive Equilibrium

A Recursive Competitive Equilibrium of the economy is a list of functions: (i) household policy functions $\mathbf{c}(\Lambda)$, $L_s(\Lambda)$, and $\mathbf{D}(\Lambda)$; (ii) firm value functions and aggregate profits $V(b, a, \Lambda)$, $V_e(b_0, \tilde{a}, \Lambda)$, and $\Phi(\Lambda)$; (iii) firm policy functions $y(b, a, \Lambda)$, $l(b, a, \Lambda)$, $p(b, a, \Lambda)$, and $b'(b, a, \Lambda)$; (iv) wage and prices $w(\Lambda)$, $\tilde{p}(\Lambda)$ and $q(\Lambda, \Lambda')$; and (v) an entrant's distribution Γ^e and incumbent's distribution Γ such that

1. Taking $\mathbf{p}$, $w(\Lambda)$, $q(\Lambda, \Lambda')$, and $\Phi(\Lambda)$ as given, $\mathbf{c}(\Lambda)$, $L_s(\Lambda)$, and $\mathbf{D}(\Lambda)$ solve the household's problem, where $\mathbf{p}$ is a vector of available goods' prices.

2. Taking $\tilde{p}(\Lambda)$, $w(\Lambda)$, and $q(\Lambda, \Lambda')$ as given, $V(b, a, \Lambda)$, $V_e(b_0, \tilde{a}, \Lambda)$, $y(b, a, \Lambda)$, $l(b, a, \Lambda)$, $p(b, a, \Lambda)$, and $b'(b, a, \Lambda)$ solve the firm's problem.

3. Aggregate profit is given by

$$\Phi(\Lambda) = \int \left[p(b, a, \Lambda)y(b, a, \Lambda) - w(\Lambda)l(b, a, \Lambda)\right] d\Gamma(b, a).$$

4. The budgets constraint (15) satisfies.

5. $w(\Lambda)$ is given by (16).

6. $\tilde{p}$ is given by (14).

7. $q(\Lambda, \Lambda')$ is given by (17).

8. For all measurable sets of entrants' idiosyncratic productivities $\mathcal{A}$,

$$\Gamma^e(\mathcal{A}) = M \int\int\int \mathbb{1}\{a' \in \mathcal{A}\} \times \mathbb{1}\{\psi_e \leq \psi_e^*(b_0, \tilde{a}, \Lambda)\} \times d\Pi_a(a'|\tilde{a}) \times dh(\tilde{a}) \times dG^e(\psi_e),$$

where M and ψ_e^* are given by (27) and (29), respectively.

Table 5: Fixed Parameter

Parameter	Symbol	Value
Annual Discount Factor	β	0.96
Coefficient of Relative Risk Aversion	σ	1
Inverse of Frisch Elasticity of Labor Supply	ν	0
Degree of Production Returns-to-Scale	α	0.7
Steady-state Labor Supply	L_{ss}	1/3
Elasticity of Substitution	ρ	1.6
Degree of Habit Formation	θ	1.49
Depreciation Rate of Customer Capital	δ	0.19

Note: This table shows externally calibrated parameters.

9. For all measurable sets of customer capital and idiosyncratic productivities $\mathcal{B} \times \mathcal{A}$,

$$\Gamma'(\mathcal{B} \times \mathcal{A}) = \int\int\int \mathbb{1}\{b'(b, a, \Lambda) \in \mathcal{B}\} \times \mathbb{1}\{a' \in \mathcal{A}\} \times \mathbb{1}\{\psi \leq \psi^*(b, a, \Lambda)\}$$
$$\times\, d\Gamma(b, a) \times d\Pi_a(a'|a) \times dG(\psi) + \mathbb{1}\{b_0 \in \mathcal{B}\} \times \Gamma^e(\mathcal{A}),$$

where ψ^* is given by (22).

4 Calibration

In this section, I choose common values used in literature for a subset of parameters and calibrate the rest to match standard moments in the data.

4.1 Fixed Parameters

The externally fixed parameters are listed in Table 5. To be comparable with the data, I set one period in the model equal to one year. The annual discount factor is $\beta = 0.96$. I assume log utility in consumption, $\sigma = 1$. For labor supply, I assume an inverse Frisch elasticity $\nu = 0$ as in Hansen (1985) and Rogerson (1988). And as explained above, this simplifies the equilibrium computation. I set the production elasticity of labor input as $\alpha = 0.7$. I choose a labor disutility parameter ω such that the steady-state labor supply $L_{ss} = 1/3$.

For the parameters that govern the product demand dynamics, I use the results of structural estimation from Foster, Haltiwanger, and Syverson (2016). The elasticity of substitution is $\rho = 1.6$, the degree of habit formation is $\theta = 1.49$, and the depreciation rate of customer capital is $\delta = 0.19$. $\rho = 1.6$ implies that a simple monopolistic firm would charge a markup $\mu = 2.7$. However, due to strong incentives to invest in customer capital, the mean

Table 6: Parameter Calibration

A. Targeted Moments in Data		
Target	Data	Model
Average Exit (Entry) Rate	7.5%	8.5%
Entrant's Relative Size	51%	54%
Exiter's Relative Size	41%	43%
$corr(s_{it}, s_{it-1})$	0.96	0.94
IQR of $std(\Delta \log s_{it})$	0.22	0.22
Autocorr. of Aggregate Output	0.61	0.61
Std. Dev. of Aggregate Output	0.027	0.027
B. Fitted Parameters		
Parameter	Symbol	Value
Initial Customer Capital	b_0	$0.58 \times b_{ss}$
Mean Parameter of Operating Cost	m_ψ	-2.73
Std Deviation Parameter of Operating Cost	σ_ψ	0.48
Persistence of Idiosyncratic Productivity	ρ_a	0.83
Standard Deviation of Idiosyncratic Productivity	σ_a	0.16
Persistence of Aggregate Productivity	ρ_A	0.61
Standard Deviation of Aggregate Productivity	σ_A	0.0138

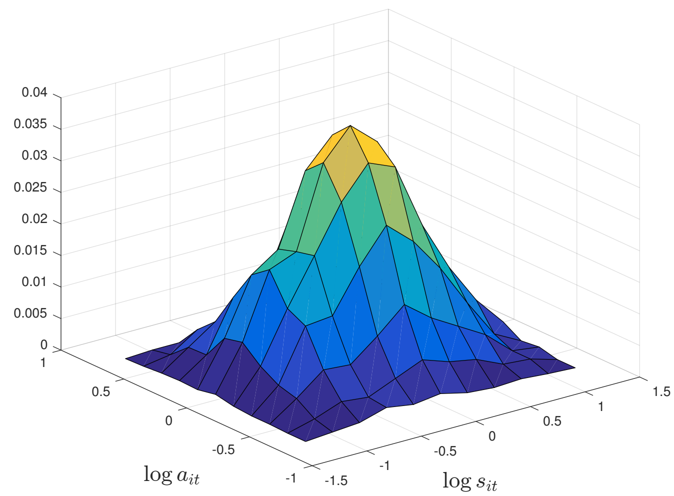
Note: Panel A shows targeted moments for model calibration and panel B shows parameters chosen to match empirical moments

markup in simulation is actually around $\mu = 1.5$, close to what most literature has found. This value is also close to the empirical estimate in the data.

4.2 Fitted Parameters

The remaining parameters that I choose to match the empirical moments are listed in Table 6. Since calibration of the full set of parameters in a general equilibrium requires a significant amount of time, I split the calibration procedure into two steps. In the first step, I calibrate the parameters that govern the cross-sectional moments of firms. They are the parameters regarding the distribution of entry costs, the relative size of entrants, the distribution of operating costs, and idiosyncratic productivities. I calibrate theses parameters in the steady state equilibrium ($A_t = 1$). I assume that the entry cost ψ_e is drawn from a uniform distribution with bounds $[0, \bar{\psi}_e]$. To be consistent with the finding in Foster et al. (2008), I set $\bar{\psi}_e$ such that entrants have the same average productivities as incumbents. I set the initial customer capital for entrants b_0 to match the relative size of entrants to incumbents. I assume that the operating cost ψ is drawn from a log normal distribution with mean m_ψ and standard deviation σ_ψ. I set $\{m_\psi, \sigma_\psi\}$ to match the average exit rate and the relative

Figure 3: Stationary Distribution of Firms



Note: a_{it} is idiosyncratic productivity, and s_{it} is market share. s_{it} is demeaned for illustration purposes.

size of exiters. And for idiosyncratic productivities, I set ρ_a and σ_a to match the persistence and the standard deviation of market share. Note that I only need $\rho_a = 0.83$ to match the high persistence of market share (0.96). This is because the customer capital b_{it} generates additional autocorrelations across periods. And this persistence estimate is very close to the one in Foster et al. (2008).

In the second step, I calibrate the persistence ρ_A and standard deviation σ_A of aggregate productivity in the general equilibrium. I set these two parameters such that the persistence and the standard deviation of X_t^D match the data.[12]

5 Numerical Results and Analysis

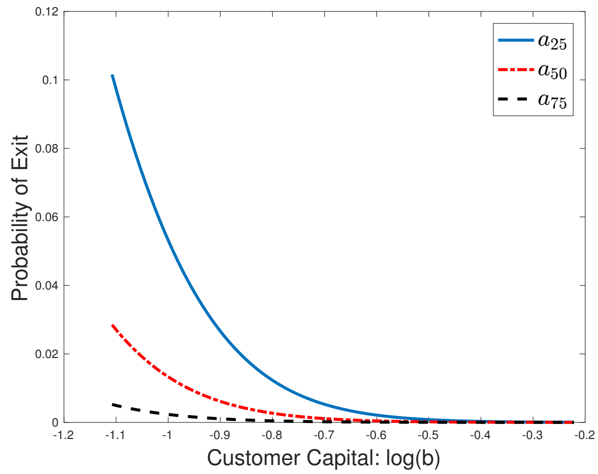
In this section, I first describe the model's implication for firm dynamics in the steady state. Then, I discuss the numerical results of the model with aggregate fluctuations. In particular, I do an impulse response analysis to analyze variables of interest.

5.1 Steady-State Results

Figure 3 displays the stationary distribution of firms over idiosyncratic productivity a_{it} and demaned market share in revenue $s_{it} = \frac{p_{it}y_{it}}{\sum_i p_{it}y_{it}}$. In contrast to Hopenhayn (1992), where there is a one-to-one mapping between idiosyncratic productivity and size (unless stated

[12]Note that, to be consistent with the empirical work, I use the data-consistent variable X_t^D as the regressor, instead of the welfare-consistent variable X_t. See the Appendix for the definition.

Figure 4: Exit Probability of Firms: $1 - G(\psi^*)$



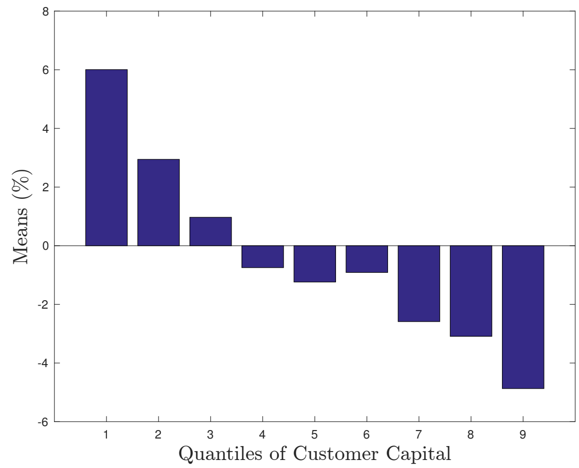
otherwise, size refers to market share in this paper,), customer capital b_{it} jointly affects firms' size. The correlation between idiosyncratic productivity and size is high, $\rho_{\log a_{it}, \log s_{it}} = 0.73$, but not perfect. Previous papers generate this pattern with firms facing adjustment costs for inputs (e.g., capital). However, there are no such adjustment costs in this model, and firms are simply constrained by the level of current demand (b_{it}). A firm with high productivity but low demand can only grow in size slowly by accumulating its customer capital.

Figure 4 displays the steady-state exit probability of firms. I plot the exit probability for firms with 25^{th} percentile, 50^{th} percentile, and 75^{th} percentile of idiosyncratic productivities in the steady state. First, we note that, conditional on customer capital, the exit rate drops rapidly as productivity increases. This is due to a high persistence of productivity as calibrated in the model ($\rho_a = 0.83$). A firm knows that it can keep its high productivity once drawn, hence its discounted expected profit increases substantially and lowers the exit probability. On the other hand, for a given productivity level, exit probability smoothly increases as its customer capital shrinks. Hence, a firm with high productivity can still possibly exit due to low demand.

Figure 5 displays the mean of revenue growth rate conditional on customer capital. Overall, firms with low customer capital are growing, while firms with high customer capital are shrinking. This is due to two forces. The first force is that low b firms are expanding output to catch up with optimal size implied by high a, while high b firms are shrinking to adjust to an optimal size implied by a new low a. The second force is that the productivity process is mean-reverting, and there is positive, though low, correlation between customer capital and productivity, $\rho_{\log a_{it}, \log b_{it}} = 0.35$.

Lastly, Table 7 compares the markup distributions between the data and the customer

Figure 5: Revenue Growth Rate $\frac{p_{it+1}y_{it+1}}{p_{it}y_{it}}$



capital model. Column 2 indicates that the customer capital model does not only match the mean, but also captures 60% of dispersion in the level of markup. This is because a markup decision becomes dynamic as a firm experiences different stages of growth as discussed above. And the incentive to attract new customers could get so strong that some firms are willing to charge a markup below one, consistent with the data.

5.2 Results with Aggregate Fluctuations

5.2.1 Heterogeneity in Markup Cyclicality

I now move to the case with aggregate fluctuations of A_t. First, I show that my model is able to match the heterogeneity of markup cyclicality, which is the key moment to target in this paper. To compare my model to the data, I split firms into five categories based on their market share s_{it}: $\mathcal{C} \in \{1, 2, 3, 4, 5\}$. And I separately run the following first difference regression:

$$\Delta \log \mu_{it} = \phi_{\mathcal{C}} \Delta \log X_t^D + \varepsilon_{it}. \tag{31}$$

My results are summarized in Figure 6. First, as discussed in the empirical section, small firms have more countercyclical markups than large firms in the data. And the result is robust against alternative specifications of the production function (Cobb-Douglas and Translog). My model generates the same magnitude of heterogeneity in markup cyclicality. The result is encouraging since the only moments I target in the calibration are static moments, but the model successfully captures the heterogenous dynamics of markups.

Second, the model captures one half of average cyclicality when compared to the Cobb-Douglas case (-0.5 vs -1.1) and 60% for the Translog (-0.5 vs -0.7). This possibly implies

Table 7: Markup Estimates: Data vs Model

Moments	(1) DLW	(2) Customer Capital
Mean	1.58	1.50
Std. Log	0.52	0.31
p5	0.58	0.88
p25	0.83	1.18
p50	1.13	1.40
p75	1.68	1.76
p95	4.01	2.36
p75/p25	2.01	1.50

Note: This table shows De Loecker-Warzynski markup estimates and model-simulated markups. Column 1 shows De Loecker-Warzynski markup estimates with a Cobb-Douglas specification. Column 2 shows the results of the baseline customer capital model.

that other models of markup cyclicality are still important. For example, one element the model lacks is sticky price. It will be an important extension to add this feature into the model for future research to understand the interactions between the sticky price and cyclical variations in desired markups.[13]

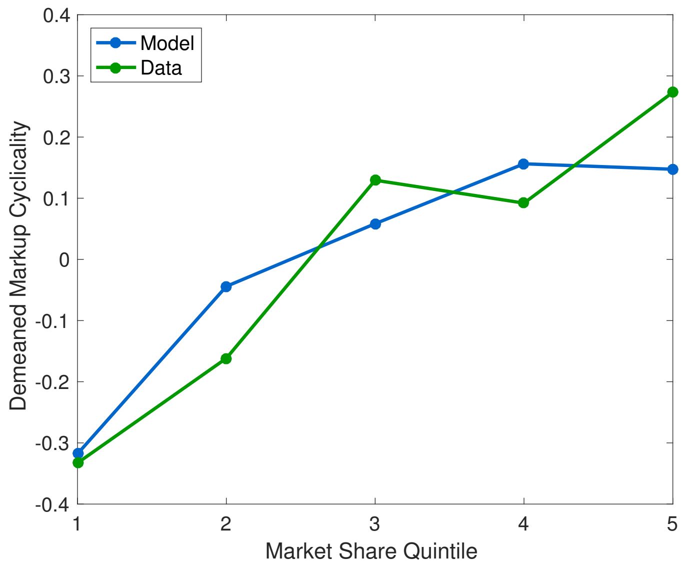
5.2.2 Countercyclical Firm Size Dispersion

To analyze the firm dynamics upon a negative aggregate productivity shock, I do the following Impulse Response analysis. Specifically, I run a simulation of 2,000 independent economies, each with T periods. At time $t = T_0 < T$, I impose a one-and-a-half standard deviation shock to the aggregate productivity A_t, and all economies evolve normally afterward. For each period t, I calculate the average value of interested variables and normalize to give the percentage deviation from their pre-shock values.

Figure 7 depicts the heterogeneous responses of firms. The middle panel depicts the markup responses of small and large firms, where small refers to the bottom market share tercile and large refers to the top market share tercile. Again, a small firm's markups rise more than a large firm's upon a negative aggregate productivity shock. An interesting and important implication of the markup dynamics is their transmission into a firm's output

[13] The volatility of the stochastic discount factor $q_{t,t+1}$ determines the magnitude of markup cyclicality. One way to increase its volatility is to increase the value of relative risk aversion. However, a model with a higher risk aversion parameter $\sigma = 2$, an upper bound suggested by Chetty (2006), only generates average markup cyclicality of -0.7, still lower than the Cobb-Douglas case in absolute size. Also, the aggregate labor becomes counter-cyclical due to a strong wealth effect from a highσ.

Figure 6: Markup Cyclicality by Firm Size: Model vs Data



Note: The figure plots the coefficient estimate of regression 31 for each quintile of market share. The solid blue line shows data estimates and the green line with circles shows the simulation of customer capital.

dynamics. The bottom panel shows the output responses of small and large firms. Since a small firm raises markups more than a large firm, its output declines relatively more. Hence, the dispersion of firm output increases endogenously. And note that the output gap between the two groups remains for a long period after the negative shock. The reason is that the law of motion for customer capital (12) is a persistent process, so it takes time for a small firm to catch up with a large firm.

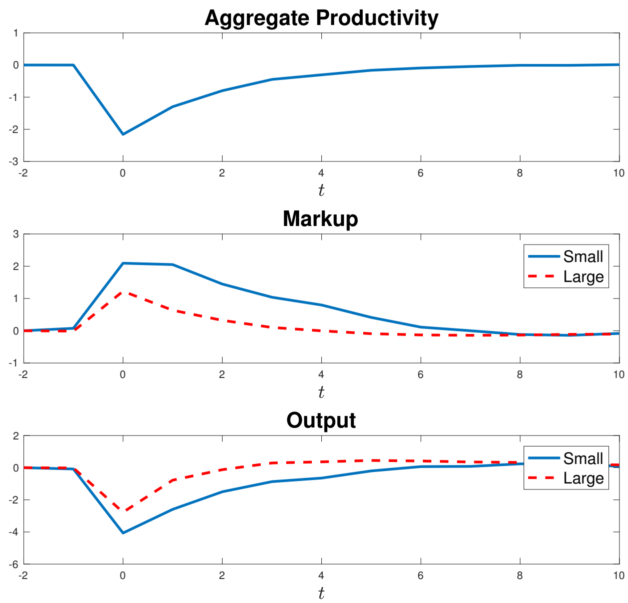
Unfortunately, we don't observe real output in the data. However, we can test the dispersion by looking at deflated value-added. In my simulation, to be consistent with the data, I deflate the sales with the CPI p_t^D. For both simulation and data, I run the following regressions:

$$\Delta \log \left(\frac{p_{it} y_{it}}{p_t^D} \right) = \phi_{\mathcal{C}}^{py} \Delta \log X_t^D + \varepsilon_{it}, \tag{32}$$

where $\mathcal{C} \in \{1, 2, 3, 4, 5\}$ as above. The results are in Figure 8. We can see that large firms tend to have less volatile sales, both in the model and data. The model explains 62% of heterogeneity in output cyclicality.

In Figure 9, I plot the time profile of labor, output, and idiosyncratic productivity dispersion. Again, similar to output, the dispersion of labor goes up upon the shock. And due to decreasing returns-to-scale for labor input, the dispersion of labor is greater than the dispersion of output. In the literature, much interest has been placed on the second moment of productivity. Using the US census data, Bloom, Floetotto, Jaimovich, Saporta-Eksten, and

Figure 7: Impulse Response to TFP A_t



Note: The shock is a negative $1.5 \times \sigma_A$ to the economy. The first panel plots the impulse response of the aggregate productivity A_t. The second panel plots the markup impulse response of small and large firms. The third panel plots the output impulse response of small and large firms.

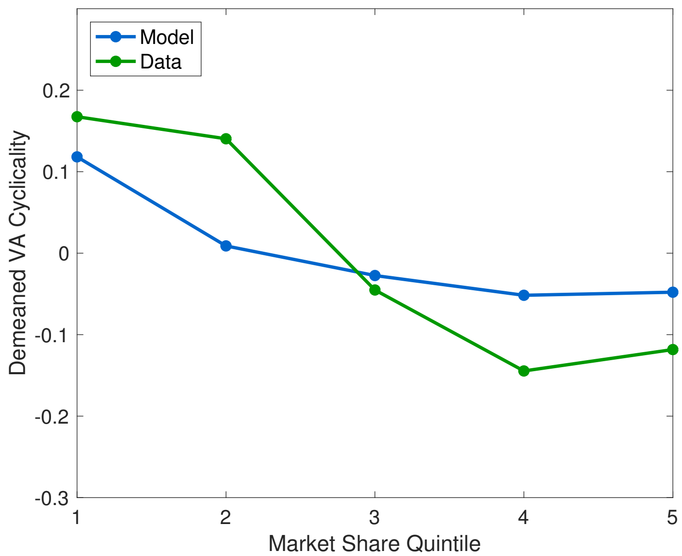
Terry (2014) find that the second moment of productivity growth is counter-cyclical, and Kehrig (2015) find that the second moment of level of productivity is also counter-cyclical. Hence, the time-varying distribution of productivity might be driving the result of size dispersion. But interestingly, the dispersion of idiosyncratic productivities actually goes down in the model. The reasoning is that there is a strong selection effect at the entry and exit. During recessions, only firms with higher than average productivity survive, so the distribution of productivity compresses. Overall, if anything, the cyclicality of the productivity dispersion dampens the cyclicality of the firm size dispersion.

5.2.3 Dynamics of Measured TFP

Next, I analyze the model's implications for the measured TFP. Following the standard business cycle accounting framework, I define the measured TFP Z_t as a ratio of total output divided by total inputse:

$$Z_t \equiv \frac{X_t}{L_t^{\alpha}}, \tag{33}$$

Figure 8: Value-Added Cyclicality by Firm Size: Model vs Data



Note: The figure plots the coefficient estimate of regression 32 for each quintile of market share. The figure compares the case of empirical estimates of markup and the simulation of customer capital. All values are demeaned within each case.

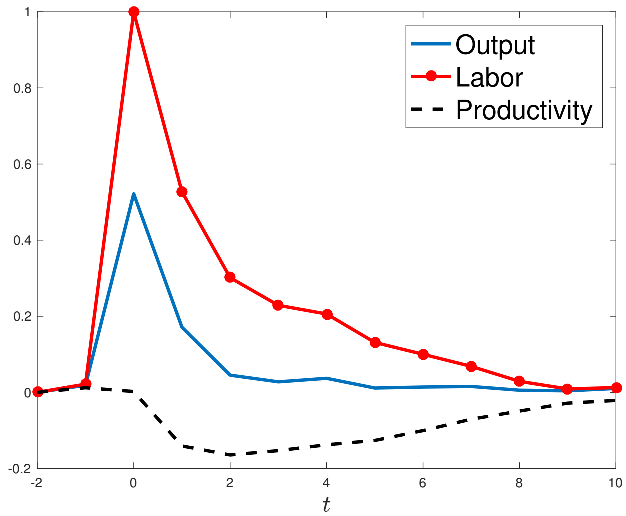
where L_t is total labor supply $L_t \equiv \int l(b, a, \Lambda_t) d\Gamma_t(b, a)$. With some algebra, it can be shown that

$$Z_t = A_t \times \underbrace{\mathrm{E}(a)_t}_{\textit{Compositional Effect}} \times \underbrace{\left[\int \left(\frac{a}{\mathrm{E}(a)_t} b^\theta \right)^{\frac{\rho-1}{\rho}} \left(\frac{l(b, a, \Lambda_t)}{L_t} \right)^{\alpha \frac{\rho-1}{\rho}} d\Gamma_t(b, a) \right]^{\frac{\rho}{\rho-1}}}_{\textit{Misallocation Effect}}.$$

As seen above, the difference between Z_t and A_t is due to two terms associated with compositional and misallocation effect. First, $\mathrm{E}(a)_t$ is the mean of incumbents' idiosyncratic productivity. This variable is time varying due to the composition of firms over business cycles. High-productivity firms survive during recessions, and low-productivity firms enter the economy during booms. So $\mathrm{E}(a)_t$ is counter-cyclical. The second term measures the change in TFP due to misallocation of labor input. The second term above $(\cdot)^{\alpha \frac{\rho}{\rho-1}}$ is a concave function, and if labor input l is more dispersed, the measured TFP Z_t is more downward biased relative to A_t due to Jensen's inequality. Finally, I denote a data-consistent counterpart for measured TFP as Z_t^D and define it as $Z_t^D \equiv X_t^D / L_t^\alpha$ accordingly.

To see the amplification of measured TFP due to misallocation, I depict the time profile of measured TFP and correlation of productivity and labor in Figure 10. Upon the negative shock, the measured TFP Z_t^D drops by 10% more than the true aggregate productivity A_t. However, Z_t^D surpasses A_t at $t = 3$ and remains above for 10 periods after the shock. This

Figure 9: Dispersion of Firm Size and Productivity



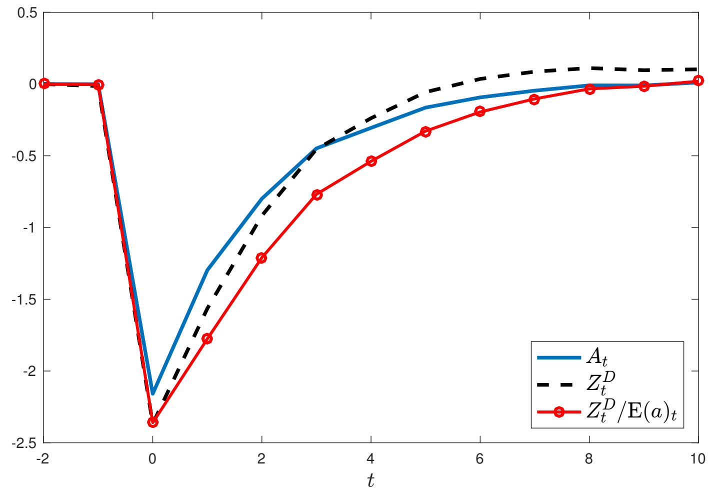
Note: The shock is a negative $1.5 \times \sigma_A$ to the economy. The solid blue line depicts the standard deviation of output y_{it}. The dotted-solid red line depicts the standard deviation of labor l_{it}. The dashed black line depicts the standard deviation of idiosyncratic productivity a_{it}.

rebound phenomenon is due to the strong selection effect at the extensive margin. During recessions, the surviving firms have higher idiosyncratic productivities, which contributes to the *rebound*. Hence, to examine the amplification effect due to misallocation, I also plot the measured TFP net of the compositional effect, $Z_t^D/\mathrm{E}(a)_t$. We see that the alternative measure is always below the true productivity A_t and converges more slowly. In fact, the persistence and the standard deviation of measured TFP (net of compositional effect) has increased by 14% and 25%, respectively. In the bottom panel, I plot the labor allocative efficiency measured by correlation between idiosyncratic productivity a_{it} and labor l_{it}. Clearly, we see that the allocation efficiency drops at the beginning and slowly recovers to pre-shock level.

6 Conclusion

This paper documented the firm-level markup behavior over the business cycle and established a general equilibrium model with heterogeneous firms consistent with the empirical findings. Among manufacturing firms from the Amadeus dataset, firms display strong countercyclical markups, and, interestingly, small firms' markups are more countercyclical than large firms'. I then build a quantitative model that features customer capital and an endogenous firm entry and exit decisions. In the model, a negative aggregate productivity shock increases small firms' exit probabilities more than large firms', and thus small firms increase

Figure 10: Comparison of Aggregate Productivity A_t and Measured TFP Z_t^D



Note: The shock is a negative $1.5 \times \sigma_A$ to the economy. The figure depicts the time profile of aggregate productivity A_t, the measured TFP Z_t^D, and the measured TFP net of compositional effect $Z_t^D/\mathrm{E}(a)_t$.

their short-term profits by raising markups and put less weight on maximizing long-term profits. Also, the model has two further implications. First, because small firms have more volatile exit risks over business cycles than large firms, their pricing, and thus output responses, are more volatile. Hence, during recessions, small firms' output declines more than large firms', which gives rise to a more spread-out firm size dispersion. In addition to that, the resulting input misallocation amplifies the standard deviation of the measured TFP.

In spite of the already rich structure of the model, I did not incorporate two other features that are important for business cycle and policy analysis. First, I abstracted from adjustment frictions for the labor input. Incorporating this element into the model is important for two reasons. The first reason is that the current model could not explain the seemingly contradictive evidence of markup cyclicality while using static inputs (material) and dynamic inputs (labor). The second reason is that the incorporation would allow for a decomposition analysis of labor wedge into a product market wedge (markup) and labor market wedge, but the model in this paper attributes all of the labor wedge to the markup. The second feature absent from the model is nominal rigidity in price setting. With price stickiness, we can study the dynamics between inflation rate and output fluctuations and provide guidance for policymakers regarding monetary policy. Adding these two features will be important extensions for future works.

References

[1] Ackerberg, D. A., Caves, K., and Frazer, G. Identification properties of recent production function estimators. *Econometrica 83*, 6 (2015), 2411–2451.

[2] Amiti, M., Itskhoki, O., and Konings, J. Importers, exporters, and exchange rate disconnect. *The American Economic Review 104*, 7 (2014), 1942–1978.

[3] Amiti, M., Itskhoki, O., and Konings, J. International shocks, variable markups and domestic prices. *The Review of Economic Studies* (2019).

[4] Atkeson, A., and Burstein, A. Pricing-to-market, trade costs, and international relative prices. *The American Economic Review 98*, 5 (2008), 1998–2031.

[5] Bellone, F., Musso, P., Quéré, M., and Nesta, L. Productivity and market selection of french manufacturing firms in the nineties. *Revue de l'OFCE*, 5 (2006), 319–349.

[6] Bilbiie, F., Ghironi, F., and Melitz, M. J. Endogenous entry, product variety, and business cycles. *Journal of Political Economy 120*, 2 (2012), 304–345.

[7] Bils, M. The cyclical behavior of marginal cost and price. *The American Economic Review* (1987), 838–855.

[8] Bils, M., Klenow, P. J., and Malin, B. A. Testing for keynesian labor demand. In *NBER Macroeconomics Annual 2012, Volume 27*. University of Chicago Press, 2012, pp. 311–349.

[9] Bils, M., Klenow, P. J., and Malin, B. A. Resurrecting the role of the product market wedge in recessions. *American Economic Review 108*, 4-5 (2018), 1118–46.

[10] Blanchard, O., and Galí, J. Real wage rigidities and the new keynesian model. *Journal of Money, credit and Banking 39*, s1 (2007), 35–65.

[11] Bloom, N. The impact of uncertainty shocks. *Econometrica 77*, 3 (2009), 623–685.

[12] Bloom, N., Floetotto, M., Jaimovich, N., Saporta Eksten, I., and Terry, S. Really uncertain business cycles. *US Census Bureau Center for Economic Studies Paper No. CES-WP-14-18* (2014).

[13] Chari, V. V., Kehoe, P. J., and McGrattan, E. R. Business cycle accounting. *Econometrica 75*, 3 (2007), 781–836.

[14] Chetty, R. A new method of estimating risk aversion. *The American Economic Review 96*, 5 (2006), 1821–1834.

[15] Chevalier, J. A., and Scharfstein, D. S. Capital-market imperfections and countercyclical markups: Theory and evidence. *The American Economic Review 86*, 4 (1996), 703–725.

[16] Clarida, R., Gali, J., and Gertler, M. The science of monetary policy: A new keynesian perspective. *Journal of Economic Literature 37* (1999), 1661–1707.

[17] Clementi, G. L., Khan, A., Palazzo, B., and Thomas, J. K. Entry, exit and the shape of aggregate fluctuations in a general equilibrium model with capital heterogeneity. *Unpublished Working Paper* (2014).

[18] Clementi, G. L., and Palazzo, B. Entry, exit, firm dynamics, and aggregate fluctuations. Tech. rep., National Bureau of Economic Research, 2013.

[19] De Loecker, J., and Warzynski, F. Markups and firm-level export status. *The American Economic Review 102*, 6 (2012), 2437–2471.

[20] Edmond, C., Midrigan, V., and Xu, D. Y. Competition, markups, and the gains from international trade. *The American Economic Review 105*, 10 (2015), 3183–3221.

[21] Edmond, C., and Veldkamp, L. Income dispersion and counter-cyclical markups. *Journal of Monetary Economics 56*, 6 (2009), 791–804.

[22] Foster, L., Haltiwanger, J., and Syverson, C. The slow growth of new plants: Learning about demand? *Economica 83*, 329 (2016), 91–129.

[23] Gilchrist, S., Schoenle, R., Sim, J., and Zakrajsek, E. Inflation dynamics during the financial crisis. *FEDS Working Paper* (2015).

[24] Hall, R. E. Market structure and macroeconomic fluctuations. *Brookings papers on economic activity*, 2 (1986), 285–322.

[25] Hall, R. E. What the cyclical response of advertising reveals about markups and other macroeconomic wedges. *Hoover Institution, Stanford University* (2014).

[26] Hansen, G. D. Indivisible labor and the business cycle. *Journal of monetary Economics 16*, 3 (1985), 309–327.

[27] Hopenhayn, H., and Rogerson, R. Job turnover and policy evaluation: A general equilibrium analysis. *Journal of political Economy* (1993), 915–938.

[28] Hopenhayn, H. A. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society* (1992), 1127–1150.

[29] Hsieh, C.-T., and Klenow, P. J. Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics 124*, 4 (2009), 1403–1448.

[30] Jaimovich, N., and Floetotto, M. Firm dynamics, markup variations, and the business cycle. *Journal of Monetary Economics 55*, 7 (2008), 1238–1252.

[31] Kehrig, M. The cyclical nature of the productivity distribution. *Earlier version: US Census Bureau Center for Economic Studies Paper No. CES-WP-11-15* (2015).

[32] Klemperer, P. Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *The review of economic studies 62*, 4 (1995), 515–539.

[33] Kryvtsov, O., and Midrigan, V. Inventories, markups, and real rigidities in menu cost models. *The Review of Economic Studies* (2012), rds028.

[34] Lee, Y., and Mukoyama, T. Entry, exit, and plant-level dynamics over the business cycle. *Working Paper* (2013).

[35] Levinsohn, J., and Petrin, A. Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70*, 2 (2003), 317–341.

[36] Moreira, S. Firm dynamics, persistent effects of entry conditions, and business cycles. *Working Paper* (2015).

[37] Nakamura, E., and Steinsson, J. Price setting in forward-looking customer markets. *Journal of Monetary Economics 58*, 3 (2011), 220–233.

[38] Nekarda, C. J., and Ramey, V. A. The cyclical behavior of the price-cost markup. Tech. rep., National Bureau of Economic Research, 2013.

[39] Olley, G. S., and Pakes, A. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica 64*, 6 (November 1996), 1263–97.

[40] Opp, M. M., Parlour, C. A., and Walden, J. Markup cycles, dynamic misallocation, and amplification. *Journal of Economic Theory 154* (2014), 126–161.

[41] Peters, M. Heterogeneous mark-ups, growth and endogenous misallocation. *Working Paper* (2013).

[42] Ravn, M., Schmitt-Grohé, S., and Uribe, M. Deep habits. *The Review of Economic Studies 73*, 1 (2006), 195–218.

[43] Restuccia, D., and Rogerson, R. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics 11*, 4 (2008), 707–720.

[44] Rogerson, R. Indivisible labor, lotteries and equilibrium. *Journal of monetary Economics 21*, 1 (1988), 3–16.

[45] Rotemberg, J. J., and Saloner, G. A supergame-theoretic model of price wars during booms. *The American Economic Review 76*, 3 (1986), 390–407.

[46] Rotemberg, J. J., and Woodford, M. Markups and the business cycle. In *NBER Macroeconomics Annual 1991, Volume 6*. MIT Press, 1991, pp. 63–140.

[47] Rotemberg, J. J., and Woodford, M. The cyclical behavior of prices and costs. *Handbook of macroeconomics 1* (1999), 1051–1135.

[48] Smets, F., and Wouters, R. An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European economic association 1*, 5 (2003), 1123–1175.

[49] Smets, F., and Wouters, R. Shocks and frictions in us business cycles: A bayesian dsge approach. *The American Economic Review 97*, 3 (2007), 586–606.

[50] Stroebel, J., and Vavra, J. House prices, local demand, and retail prices. Tech. rep., National Bureau of Economic Research, 2014.

A Production Function Estimation

In this appendix, I present the estimation procedure for a Translog production function. For a Cobb-Douglas production function, all the steps are the same except for setting higher-order terms for inputs to be zero. In particular, I estimate the following gross output production function:

$$
\begin{aligned}
y_{it} =& \theta_m m_{it} + \theta_l l_{it} + \theta_k k_{it} + \theta_{mm} m_{it}^2 + \theta_{ll} l_{it}^2 + \theta_{kk} k_{it}^2 \\
&+ \theta_{ml} m_{it} l_{it} + \theta_{mk} m_{it} k_{it} + \theta_{lk} l_{it} k_{it} + a_{it} + \epsilon_{it},
\end{aligned}
$$

where m_{it} denotes material, l_{it} denotes labor, k_{it} denotes capital, a_{it} is unobserved productivity, and ϵ_{it} is some unanticipated production shock to the firm or classical measurement error. The main challenge of the estimation is controlling for the unobserved productivity a_{it}, which is in generally correlated with the inputs. Olley and Pakes (1996) and Levinsohn and Petrin (2003) suggest a "control function" method to solve the simultaneity issue.

I closely follow Ackerberg, Caves, and Frazer's (2015; hereafter ACF) assumption and approach for estimation, and my procedure is the following:

1. Given the timing assumption in ACF,[14] a firm's demand function for materials is

$$
m_{it} = m(a_{it}, l_{it}, k_{it}, \mathbf{x}_{it}),
$$

where $\mathbf{x}_{it}$ is a set of control variables, which includes wage and time-fixed effects. Assume that the demand function for materials can be inverted for a_{it}, and substitute it into the production function to get

$$
\begin{aligned}
y_{it} =& \theta_m m_{it} + \theta_l l_{it} + \theta_k k_{it} + \theta_{mm} m_{it}^2 + \theta_{ll} l_{it}^2 + \theta_{kk} k_{it}^2 \\
&+ \theta_{ml} m_{it} l_{it} + \theta_{mk} m_{it} k_{it} + \theta_{lk} l_{it} k_{it} + m^{-1}(m_{it}, l_{it}, k_{it}, \mathbf{x}_{it}) + \epsilon_{it}.
\end{aligned}
$$

No parameter is identified at this first stage. The purpose of this stage is to get an estimate of the "anticipated" output $\hat{h}_{it}$ free of the error $\hat{\epsilon}_{it}$,

$$
y_{it} = \hat{h}(m_{it}, l_{it}, k_{it}, \mathbf{x}_{it}) + \hat{\epsilon}_{it}.
$$

Since $h(\cdot)$ is a highly nonlinear function in general, I estimate it with a high-order polynomial by running an OLS regression.

[14] ACF assume that labor choice is decided some time between $t-1$ and t, after k_{it} is chosen at $t-1$, but before m_{it} is chosen at t.

2. After the first stage, for any set of parameters $\boldsymbol{\theta} = (\theta_m, \theta_l, \theta_k, \theta_{mm}, \theta_{ll}, \theta_{kk}, \theta_{ml}, \theta_{mk}, \theta_{lk})$, we can compute the implied productivity:

$$\begin{aligned} a(\boldsymbol{\theta})_{it} =& \hat{h}(m_{it}, l_{it}, k_{it}, \mathbf{x}_{it}) - \theta_m m_{it} - \theta_l l_{it} + \theta_k k_{it} - \theta_{mm} m_{it}^2 - \theta_{ll} l_{it}^2 - \theta_{kk} k_{it}^2 \\ & - \theta_{ml} m_{it} l_{it} - \theta_{mk} m_{it} k_{it} - \theta_{lk} l_{it} k_{it}. \end{aligned}$$

   And by running an AR(1) regression on $a(\boldsymbol{\theta})_{it}$, we can recover the innovation shock to the productivity $\hat{\xi}_{it}(\boldsymbol{\theta})$:

$$a(\boldsymbol{\theta})_{it} = \rho_a a(\boldsymbol{\theta})_{it-1} + \xi_{it}(\boldsymbol{\theta}).$$

3. The main identification assumption is that the innovation to productivity $\xi_{it}(\boldsymbol{\theta})$ is independent of a set of lagged variables (m_{it-1}, l_{it-1}, k_{it}, m_{it-1}^2, l_{it-1}^2, k_{it}^2, $m_{it-1}l_{it-1}$,$m_{it-1}k_{it}$, $l_{it-1}k_{it}$). I use a standard GMM technique to obtain the estimation of $\boldsymbol{\theta}$. These moment conditions are similar to those in ACF.

After the estimation of the production function, I can easily recover the firm-specific markup. In particular, say for material input, the estimated markup $\hat{\mu}_{it}$ is equal to

$$\hat{\mu}_{it} = (\hat{\theta}_m + 2\hat{\theta}_{mm} m_{it} + \hat{\theta}_{ml} l_{it} + \hat{\theta}_{mk} k_{it}) \left(\frac{P_{it}^M M_{it}}{P_{it} \frac{Y_{it}}{\exp{(\hat{\epsilon}_{it})}}} \right)^{-1},$$

where $\hat{\epsilon}_{it}$ is the estimated error term from the first-stage regression. This is to capture the timing assumption that a firm makes an input decision before realizing the "unanticipated" shock ϵ_{it}.

B FOCs for the Incumbent's Problem

B.1 First-Order Conditions for Incumbents

The incumbent's problem can be written in terms of the habit-adjusted price index $\tilde{V} \equiv V/\tilde{p}$. And for simplicity of the derivation, I omit a firm's idiosyncratic productivity a from the state variables. Then, the incumbent's transformed value function becomes the following:

$$\begin{aligned}\tilde{V}(b; A, \Lambda) = \max_{y,l,p,b'} & \frac{py - wl}{\tilde{p}} + \kappa \left[Al^{\alpha} - y \right] \\ & + \lambda \left[(1-\delta)b + \delta y - b' \right] + \eta \left[N^{\xi(\rho-1)-1} \left(\frac{p}{\tilde{p}} \right)^{-\rho} b^{\theta(\rho-1)} \tilde{C} - y \right] \\ & + \int \max \left\{ 0, -\psi + \beta \mathrm{E} \left[\left(\frac{\tilde{C}'}{\tilde{C}} \right)^{-\sigma} \tilde{V}(b'; A', \Lambda') \right] \right\} dG(\psi),\end{aligned}$$

where κ is the Lagrange multiplier for the production of output constraint, λ for law of motion for customer capital constraint, and η for firm product demand constraint, respectively. The FOCs with respect to the choice of output y, labor l, price p, and next period's customer capital b' are the following:

$$\begin{aligned} y: & \quad \eta = \frac{p}{\tilde{p}} - \kappa + \delta \lambda \\ l: & \quad \kappa \alpha A l^{\alpha-1} = \frac{w}{\tilde{p}} \\ p: & \quad \frac{p}{\tilde{p}} = \rho \eta \\ b': & \quad \lambda = \beta G(\psi^*) \mathrm{E} \left\{ \left(\frac{\tilde{C}'}{\tilde{C}} \right)^{-\sigma} \left[(1-\delta)\lambda' + \theta(\rho-1)\eta' \frac{y'}{b'} \right] \right\}. \end{aligned}$$

Noting the normalization of the exit value as 0 helps simplify the last FOC with the Leibniz integral rule.

Combining the first three FOCs yields the following:

$$\frac{p}{\tilde{p}} \left(\mu^{-1} - \bar{\mu}^{-1} \right) = \delta \lambda,$$

where $\bar{\mu} \equiv \frac{\rho}{\rho-1}$ is the optimal markup in an economy without habit (δ or θ equals zero). Note that a firm's markup is always below $\bar{\mu}$. And the firm's intertemporal FOC becomes

$$\mu^{-1} - \bar{\mu}^{-1} = G(\psi^*) \mathrm{E} \left\{ q \left[(1-\delta) \frac{p'}{p} \left(\mu'^{-1} - \bar{\mu}^{-1} \right) + \delta \frac{\theta(\rho-1)}{\rho} \frac{p'y'}{pb'} \right] \right\},$$

where $q = \beta \left(\frac{\tilde{C}'}{\tilde{C}}\right)^{-\sigma} \frac{\tilde{p}}{\tilde{p}'}$.

B.2 Symmetric Equilibrium

It follows that to have a symmetric equilibrium, it is not sufficient to exclude idiosyncratic productivity. For any $t > 0$, all entrants need to have the same customer capital as the incumbent's, $b_0 = b_t$. Under symmetric equilibrium, the habit-adjusted aggregate consumption and price index equal:

$$\tilde{C} = N^{\xi+1} y b^{\theta}$$
$$\tilde{p} = N^{-\xi} \frac{p}{b^{\theta}},$$

where N is the measure of incumbents. Then the one-period stochastic discount factor is

$$q = \beta \frac{p}{p'} \left(\frac{y}{y'}\right)^{\sigma} \left(\frac{b}{b'}\right)^{\theta(\sigma-1)} \left(\frac{N}{N'}\right)^{\xi(\sigma-1)+\sigma}.$$

Baseline Case: $\sigma = 1$

If a household has a log utility function in consumption $\sigma = 1$ as in the baseline calibration, then the stochastic discount factor simplifies to

$$q = \beta \frac{Npy}{N'p'y'}.$$

And a firm's intertemporal FOC becomes

$$\mu^{-1} - \bar{\mu}^{-1} = \beta G(\psi^*) \mathrm{E} \left\{ (1-\delta) \frac{Ny}{N'y'} \left(\mu'^{-1} - \bar{\mu}^{-1}\right) + \delta \frac{\theta(\rho-1)}{\rho} \frac{Ny}{N'b'} \right\}.$$

By iterating forward, one can obtain

$$\mu_t^{-1} - \bar{\mu}^{-1} = \frac{\delta}{1-\delta} \frac{\theta(\rho-1)}{\rho} \mathrm{E}_t \left\{ \sum_{j=1}^{\infty} \left[\beta(1-\delta)\right]^j \left[\Pi_{j'=0}^{j} G(\psi_{t+j'}^*)\right] \frac{N_t y_t}{N_{t+j} b_{t+j}} \right\}.$$

One special case is that $\delta = 1$, then the equation becomes

$$\mu_t^{-1} - \bar{\mu}^{-1} = \frac{\theta(\rho-1)}{\rho} \beta G(\psi_t^*) \frac{N_t}{N_{t+1}}.$$

Steady State without Entry and Exit

In the steady state without entry and exit, aggregate productivity is at a constant $A_t = A_{ss}$, $\forall t > 0$. Since there is neither firm entry nor exit ($G = 1$), the number of incumbents stays constant at $N_t = N_{ss}$. Then a firm charges a steady state markup:

$$\mu_{ss} = \frac{\rho_{ss}}{\rho_{ss} - 1},$$

where

$$\rho_{ss} = \rho \frac{1 - \beta(1-\delta)}{1 - \beta(1 - \delta + \delta\theta(\rho - 1))} > \rho,$$

which means that a firm's markup is below optimal markup without habit formation $\mu_{ss} < \frac{\rho}{\rho - 1}$.

In the steady state, habit stock is equal to the output $b_{ss} = y_{ss}$, habit-adjusted consumption index is $\tilde{C}_{ss} = N_{ss}^{\xi+1} y_{ss}^{1+\theta}$, and habit-adjusted price index is $\tilde{p}_{ss} = N_{ss}^{-\xi} p_{ss} y_{ss}^{-\theta}$. Hence steady state output y_{ss} is

$$y_{ss} = \left[N_{ss}^{\xi - \sigma(\xi+1) - \nu} \frac{\alpha}{\mu_{ss}\omega} A_{ss}^{(1+\nu)/\alpha} \right]^{\frac{1}{\frac{1+\nu}{\alpha} + (\sigma-1)(1+\theta)}}.$$

And steady-state labor is

$$l_{ss} = \left[N_{ss}^{\xi - \sigma(\xi+1) - \nu} \frac{\alpha}{\mu_{ss}\omega} A_{ss}^{-(\sigma-1)(1+\theta)} \right]^{\frac{1}{1+\nu+\alpha(\sigma-1)(1+\theta)}}.$$

C Numerical Solution and Simulation

In this computational appendix, first I lay out the solution algorithm for the model. To solve the heterogeneous agents model in general equilibrium, I follow the linear forecast rule approach for firms as in Krusell and Smith (1998). Throughout the solution procedure, I provide the technical details that turned out to be useful in the implementation. Also, I provide accuracy statistics regarding the forecast rule proposed below. Finally, I describe the calculation of the impulse response of the aggregate TFP shock in the model.

C.1 Transformation of the Firm's Problem

Bellman equations describing the incumbent's and entrant's problem in the main text are reproduced below. A firm's value functions are normalized by a habit-adjusted price index, and hence the incumbent's value function becomes $\tilde{V} \equiv V/\tilde{p}$:

$$\tilde{V}(b, a, \Lambda) = \max_{y,l,p,b'} \frac{py - w(\Lambda)l}{\tilde{p}(\Lambda)} + \int_{\psi} \max \left\{ 0, -\psi + \mathrm{E}_{a,\Lambda} \left[\frac{\tilde{p}(\Lambda')}{\tilde{p}(\Lambda)} q(\Lambda, \Lambda') \tilde{V}(b', a', \Lambda') \right] \right\} dG(\psi),$$

and the entrant's value function becomes $\tilde{V}_e \equiv V_e/\tilde{p}$:

$$V_e(\psi_e, b_0, \tilde{a}, \Lambda) = \max \left\{ 0, -\psi_e + \mathrm{E}_{\tilde{a},\Lambda} \left[\frac{\tilde{p}(\Lambda')}{\tilde{p}(\Lambda)} q(\Lambda, \Lambda') V(b_0, a', \Lambda') \right] \right\}.$$

As outlined in the main text, the household's optimization problem yields the following two equations for the real wage and the stochastic discount factor in the equilibrium:

$$\frac{w(\Lambda)}{\tilde{p}(\Lambda)} = \omega L(\Lambda)^{\nu} \tilde{C}(\Lambda)^{\sigma}$$
$$q(\Lambda, \Lambda') = \beta \left(\frac{\tilde{C}(\Lambda')}{\tilde{C}(\Lambda)} \right)^{-\sigma} \frac{\tilde{p}(\Lambda)}{\tilde{p}(\Lambda')}.$$

In the calibration, I choose the parameters such that the household has a log utility in consumption ($\sigma = 1$), and an infinite Frisch elasticity of labor supply ($\nu = 0$). With these parameters, the two equations above simplify to

$$\frac{w(\Lambda)}{\tilde{p}(\Lambda)} = \omega \tilde{C}(\Lambda)$$
$$q(\Lambda, \Lambda') = \beta \frac{\tilde{p}(\Lambda)\tilde{C}(\Lambda)}{\tilde{p}(\Lambda')\tilde{C}(\Lambda')}.$$

First, the real wage is now a function of only the habit-adjusted aggregate consumption $\tilde{C}(\Lambda)$. Second, in the firm's problem, we see that the normalized stochastic discount factor simplifies to $\frac{\tilde{p}(\Lambda')}{\tilde{p}(\Lambda)}q(\Lambda,\Lambda') = \beta\frac{\tilde{C}(\Lambda)}{\tilde{C}(\Lambda')}$. Hence, the only aggregate variable that matters to the firm's problem is the habit-adjusted aggregate consumption $\tilde{C}(\Lambda)$, and it can be expressed as the following in log:

$$\log\tilde{C}(\Lambda) = \log A + \log\Omega(\Lambda),$$

where $\Omega(\Lambda) \equiv \left[\int \left(al^{\alpha}(b,a,\Lambda)b^{\theta}\right)^{\frac{\rho-1}{\rho}} d\Gamma(b,a)\right]^{\frac{\rho}{\rho-1}}$, and $l(b,a,\Lambda)$ is the optimal labor demand decision for a firm with a set of state variables $\{b,a,\Lambda\}$ in the equilibirum. Hence, we see that the sufficient variables for the aggregate state Λ are the aggregate TFP A and the distribution of firms $\Gamma(b,a)$ ($\Lambda = \{A,\Gamma\}$).

Steady-State Computation

I outline the computation for the steady state of the model with $A_t = 1$ for all t, hence A_t is omitted from the state variable. Note that idiosyncratic productivities are not omitted. Overall, it is a root finding problem for the habit-adjusted aggregate consumption $\tilde{C}^*$: (1) Given a guess of $\tilde{C}^*$, solve for the firm's optimization problem[15]; (2) compute the firm's policy function of out $y(b,a,\tilde{C}^*)$, and calculate the habit-adjusted aggregate consumption from $\tilde{C}^* = \left[\int \left(y(b,a,\tilde{C}^*)b^{\theta}\right)^{\frac{\rho-1}{\rho}} d\Gamma(b,a,\tilde{C}^*)\right]^{\frac{\rho}{\rho-1}}$; and (3) repeat until the market-clearing $\tilde{C}^*$ is found.

Krusell-Smith Algorithm

To describe the dynamics of the habit-adjusted aggregate consumption $\tilde{C}(\Lambda)$, I need to keep track of the evolution of the firm distribution Γ. Unfortunately, Γ is an infinitely-dimensional object, and it is impossible to keep track of it perfectly in practice. Hence, to lessen the complexity of the problem, I follow the Krusell and Smith (1998) approach and conjecture that $\log\Omega'$ is a linear function of its past variable $\log\Omega$, and the current aggregate TFP $\log A'$. Then the conjecture rule for the habit-adjusted aggregate consumption is

$$\log\hat{\tilde{C}}' = \beta_{\tilde{C}0} + \beta_{\tilde{C}1}\log\tilde{C} + \beta_{\tilde{C}2}\log A' + \beta_{\tilde{C}3}\log A.$$

Below, I will test and discuss the internal accuracy of this forecast rule with several statistics widely used in the literature.

[15] Note that the stochastic discount factor is a constant in the steady state $q_t = \beta\left(\frac{\tilde{C}^*}{\tilde{C}^*}\right)^{-\sigma} = \beta$.

Incumbent's Problem

Now, given the linear forecast rule in $\tilde{C}$, the aggregate state variable for a firm is reduced to $\Lambda = \{A, \tilde{C}\}$. With constraints (12), (13), and (18) substituted into the value function, incumbent's normalized value function $\tilde{V}$ can be written as

$$\tilde{V}(b,a,A,\tilde{C}) = \max_{b'} \Bigg\{ N^{\frac{\xi(\rho-1)-1}{\rho}} b^{\theta(\rho-1)/\rho} \tilde{C}^{1/\rho} \left(\frac{b'-(1-\delta)b}{\delta}\right)^{\frac{\rho-1}{\rho}} - \omega \tilde{C}^{\sigma} (Aa)^{-1/\alpha} \left(\frac{b'-(1-\delta)b}{\delta}\right)^{\frac{1}{\alpha}}$$

$$- \Pr[\psi \leq \psi^*(b,a,A,\tilde{C})] \mathrm{E}\left[\psi | \psi \leq \psi^*(b,a,A,\tilde{C})\right]$$

$$+ \beta \Pr[\psi \leq \psi^*(b,a,A,\tilde{C})] \mathrm{E}_{a,A} \left[\left(\frac{\hat{\tilde{C}}'}{\tilde{C}}\right)^{-\sigma} \tilde{V}(b',a',A',\hat{\tilde{C}}') \right] \Bigg\}.$$

The operating cost ψ is drawn from a log normal distribution $\log \psi \sim N(m_\psi, \sigma_\psi)$, then probability of less than ψ^* is equal to

$$\Pr(\psi \leq \psi^*) = \Phi\left(\frac{\log \psi^* - m_\psi}{\sigma_\psi}\right), \tag{34}$$

where $\Phi(\cdot)$ is a cdf function for standard normal distribution. And the conditional expectation of operating costs is

$$\mathrm{E}(\psi | \psi \leq \psi^*) = \exp\left(m_\psi + \frac{\sigma_\psi^2}{2}\right) \Phi\left(\frac{\log \psi^* - m_\psi - \sigma_\psi^2}{\sigma_\psi}\right) \Big/ \Phi\left(\frac{\log \psi^* - m_\psi}{\sigma_\psi}\right). \tag{35}$$

Entrant's Problem

Similarly, an entrant's normalized value function $\tilde{V}_e$ can be written as

$$\tilde{V}_e(\psi_e, b_0, \tilde{a}, A, \tilde{C}) = \max \left\{ 0, -\psi_e + \beta \mathrm{E}_{\tilde{a},A} \left[\left(\frac{\hat{\tilde{C}}'}{\tilde{C}}\right)^{-\sigma} \tilde{V}(b_0, a', A', \hat{\tilde{C}}') \right] \right\}, \tag{36}$$

and the entrant enters if and only if

$$\psi_e \leq \beta \mathrm{E}_{\tilde{a},A} \left[\left(\frac{\hat{\tilde{C}}'}{\tilde{C}}\right)^{-\sigma} \tilde{V}(b_0, a', A', \hat{\tilde{C}}') \right]. \tag{37}$$

C.2 Solution Algorithm

With the transformed Bellman equations above, I now lay out the outline of the solution algorithm. First, guess initial values for the coefficients of the forecast rule $\left(\beta_{\tilde{C}0}^{(0)}, \beta_{\tilde{C}1}^{(0)}, \beta_{\tilde{C}2}^{(0)}, \beta_{\tilde{C}3}^{(0)}\right)$ to solve the model, and perform the following iterations $m = 1, 2, 3...$ of the solution algorithm:

1. Given the forecast rule $\left(\beta_{\tilde{C}0}^{(m)}, \beta_{\tilde{C}1}^{(m)}, \beta_{\tilde{C}2}^{(m)}, \beta_{\tilde{C}3}^{(m)}\right)$ from the previous iteration, solve the value functions for incumbent $\tilde{V}^m$ and entrant $\tilde{V}_e^m$.
2. With the firm's policy function, simulate the economy for T periods with some arbitrary initial conditions (A_0, Γ_0).
3. Using the simulated variables obtained from the simulation, update the forecast rule $\left(\beta_{\tilde{C}0}^{(m+1)}, \beta_{\tilde{C}1}^{(m+1)}, \beta_{\tilde{C}2}^{(m+1)}, \beta_{\tilde{C}3}^{(m+1)}\right)$ accordingly.
4. Repeat steps (1) to (3) until some convergence criteria is attained.

Below, I explain the implementation of each step of the solution algorithm laid out above in greater detail.

C.2.1 Step 1: Firm Problem

I solve the incumbent problem using policy function iteration on grid points, also known as Howard's improvement algorithm. I use $n_b = 400$ grid points for the choice of customer capital in the next period b'. For the exogenous process of the aggregate TFP A and the idiosyncratic productivity a, I discretize them following the method of Tauchen (1986) and use $n_A = 7$ and $n_a = 11$ grid points, respectively. I denote Π^A and Π^a as the transition matrix probability for the aggregate TFP and the idiosyncratic productivity, respectively. Finally, I choose $n_{\tilde{C}} = 7$ grid points for the habit-adjusted aggregate consumption $\tilde{C}$. And the incumbent uses the forecast rule $\beta_{\tilde{C}}^{(m)} = \left(\beta_{\tilde{C}0}^{(m)}, \beta_{\tilde{C}1}^{(m)}, \beta_{\tilde{C}2}^{(m)}, \beta_{\tilde{C}3}^{(m)}\right)$ to calculate the continuation value. However, given the discrete nature of the solution method, the forecasted habit-adjusted aggregate consumption in the next period $\hat{\tilde{C}}'$ does not fall on the given grid points in general. Hence, I compute the continuation value $\mathrm{E}\tilde{V}'$ off the grid points using linear interpolation. Within each loop for the policy function iteration, I iterate the value function with 100 steps forward. The policy function iteration procedure stops when the supremum norm of the percentage change of the value function is less than 10^{-6}.

For the entrant, the signal $\tilde{a}$ is drawn from the uniform distribution with bounds $[0, \bar{\psi}_e]$. Its draw of productivity for production follows the AR(1) process $\log a = \rho_a \log \tilde{a} + \varepsilon^a$. And each potential entrant also draws an entry cost ψ_e from distribution $G^e(\cdot)$. A potential entrant decides to enter the market in the next period if and only if the entry cost is lower than the continuation value. Once again, I need to use linear interpolation to calculate the continuation value for entrant.

C.2.2 Step 2: Simulation of the Model

I simulate the economy with a period of $T = 700$. I generate a series of realizations for the aggregate TFP $\{A_t\}_{t=1,\dots,T}$, which follows the discrete Markov chain process as discussed above. Also, I initiate the economy with an arbitrary distribution $\Gamma_0(b,a)$.

Throughout the simulation, I follow the histogram-based approach to track the cross-sectional distribution as proposed by Young (2010). This method avoids the sampling error from the Monto Carlo approach. In particular, I keep track of the distribution over the customer capital and idiosyncratic productivity $\Gamma_t(b,a)$. Within each period t, given the firm's policy function $b'(b,a,A,\tilde{C})$, the next period's distribution is determined by

$$
\begin{aligned}
\Gamma_{t+1}(b,a) = &\sum_{j}^{n_a}\sum_{i}^{n_b} \mathbb{1}\left\{b'(b_i,a_j,A_t,\tilde{C}_t)=b\right\} \times \Pi^a(a|a_j) \\
&\times (1-\gamma)\times \Pr\left(\psi \le \psi^*(b_i,a_j,A_t,\tilde{C}_t)\right) \times \Gamma_t(b_i,a_j) \\
&+\sum_{i}^{n_{\tilde{a}}} \mathbb{1} \times \Pr\left(\psi_e \le \psi_e^*(b_0,\tilde{a},A_t,\tilde{C}_t)\right) \times (1-\gamma)\times \Pi_a(a|\tilde{a}_i)\times h(\tilde{a}_i)\times M_t,
\end{aligned}
$$

where M_t is the mass of potential entrants at time t.

Within each period, the incumbent's decision must be consistent with the market clearing condition for output $\tilde{C}_t$. In particular, given a predetermined distribution of firm $\Gamma_t(b,a)$, for any guess of current habit-adjusted aggregate consumption $\dot{\hat{C}}$, an incumbent with customer capital b and idiosyncratic productivity a uses the forecast rule to calculate the continuation value and solves the following problem:

$$
\begin{aligned}
\tilde{V}(b,a,A,\dot{\hat{C}}) = \max_{b'} \Bigg\{ &N^{\frac{\xi(\rho-1)-1}{\rho}} b^{\theta(\rho-1)/\rho}\dot{\hat{C}}^{1/\rho}\left(\frac{b'-(1-\delta)b}{\delta}\right)^{\frac{\rho-1}{\rho}} - \omega\dot{\hat{C}}(Aa)^{-1/\alpha}\left(\frac{b'-(1-\delta)b}{\delta}\right)^{\frac{1}{\alpha}} \\
&-\Pr[\psi\le\psi^*(b,a,A,\dot{\hat{C}})]\mathrm{E}\left[\psi|\psi\le\psi^*(b,a,A,\dot{\hat{C}})\right] \\
&+\beta(1-\gamma)\Pr[\psi\le\psi^*(b,a,A,\dot{\hat{C}})]\mathrm{E}_{a,A}\left[\left(\frac{\dot{\hat{C}}'}{\dot{\hat{C}}}\right)^{-1}\tilde{V}(b',a',A',\dot{\hat{C}}')\right]\Bigg\}.
\end{aligned}
$$

The habit-adjusted aggregated consumption implied by the firm's policy function $b'(b, a, A, \dot{\tilde{C}})$ is

$$\ddot{\tilde{C}} = \left[\sum_{b,a} \left(\frac{b'(b, a, A, \dot{\tilde{C}}) - (1-\delta)b}{\delta} b^{\theta}\right)^{\frac{\rho-1}{\rho}} \Gamma_t(b, a)\right]^{\frac{\rho}{\rho-1}}.$$

The market-clearing condition for output is satisfied if and only if $\ddot{\tilde{C}} = \dot{\tilde{C}}$. In practice, I use the golden section search method to search for the solution.

Lastly, there is one point worth mentioning to speed up the algorithm in the implementation. In practice, it is costly to search for the maximizer of the customer capital b' over the grid for a given $\dot{\tilde{C}}$ every time. Hence, I instead do a linear interpolation of the policy function obtained in step (1) along the grid for $\tilde{C}$ to approximate the solution. The result turns out to be quite close to the one obtained with a search for the maximizer, and the speed gain is substantial at the same time.

C.2.3 Step 3: Forecast-Rule Update

After running the simulation for $T = 700$ periods, we have obtained a series of aggregate TFP and habit-adjusted aggregate consumption $\{A_t, \tilde{C}_t\}_{t=1,\dots,T}$.[16] To update the forecast rule, discard first $T_0 - 1$ periods of variables. This is called the "burning" process, and we try to insulate the effects of initial conditions on the equilibrium outcome as much as possible. I set $T_0 = 200$ in the implementation. Now with the simulated variables $\left\{A_t, \tilde{C}_t\right\}_{t=T_0,\dots,T}$, I run the following OLS regression:

$$\log \tilde{C}_{t+1} = \beta_{\tilde{C}0} + \beta_{\tilde{C}1} \log \tilde{C}_t + \beta_{\tilde{C}2} \log A_{t+1} + \beta_{\tilde{C}3} \log A_t.$$

Denote the set of coefficients obtained from the regression above as $\hat{\beta}_{\tilde{C}} = \left(\hat{\beta}_{\tilde{C}0}, \hat{\beta}_{\tilde{C}1}.\hat{\beta}_{\tilde{C}2}, \hat{\beta}_{\tilde{C}3}\right)$. I update the forecast rule as the weighted sum of $\hat{\beta}_{\tilde{C}}$, and the forecast rule from the previous iteration $\beta_{\tilde{C}}^{(m)}$:

$$\beta_{\tilde{C}}^{(m+1)} = w_{KS}\hat{\beta}_{\tilde{C}} + (1 - w_{KS})\beta_{\tilde{C}}^{(m)},$$

where I choose the weight $w_{KS} = 0.8$ in the implementation.

C.2.4 Step 4: Convergence Criteria

There are several convergence metrics used widely in the literature. The most commonly used metric is the change in the forecast rule regression coefficients. Here, I use the maximum error statistics as proposed by Den Haan (2010). The Den Haan statistic is the maximum

[16] Note that $\tilde{C}_t$ is obtained from the equilibrium search within each period, not from the forecast rule.

difference in log between the $\tilde{C}_t^{KS}$ generated by repeatedly applying the forecast rule $\beta_{\tilde{C}}$ only, and the $\tilde{C}_t$ generated from the equilibrium search within each period in the simulation. The convergence criteria is that the Den Haan statistic is less than equal to 10^{-4}. The details of the maximum Den Haan statistic are described in the subsection below.

C.3 Internal Accuracy Statistics

The prediction rule used by a firm to solve the value function is the following:

$$\log \hat{\tilde{C}}' = \beta_{\tilde{C}0} + \beta_{\tilde{C}1} \log \tilde{C} + \beta_{\tilde{C}2} \log A' + \beta_{\tilde{C}3} \log A.$$

The first metric commonly used in the literature is the R^2 obtained from the regression of the equation above. In my model, the R^2 is greater than 99% in all of the alternative specifications. However, Den Haan (2010) points out the following three main problems associated with the use of R^2: (i) R^2 assesses the goodness of fit conditional on the variables generated by the true law of motion. In other words, it only checks one period ahead forecast error. (ii) The R^2 is an average statistic, and it might hide some large errors. In particular, it will be problematic if an large error occurs when we are interested in the model's behavior in response to a large shock. (iii) The R^2 scales the forecast error by the variance of the dependent variable.

In response to this, Den Haan (2010) proposes to assess the accuracy of the model by calculating the maximum error between the actual habit-adjusted aggregate consumption and the habit-adjusted aggregate consumption generated by iteration of the forecast rule. This method allows the error of the forecast rule to accumulate over the time. Denote $\tilde{C}_t$ as the market-clearing consumption within each period. And denote $\tilde{C}_t^{DH}$ as the outcome of the iteration of the forecast rule $\beta_{\tilde{C}}$,

$$\log \tilde{C}_{t+1}^{DH} = \beta_{\tilde{C}0} + \beta_{\tilde{C}1} \log \tilde{C}_t^{DH} + \beta_{\tilde{C}2} \log A_{t+1} + \beta_{\tilde{C}3} \log A_t,$$

where the initial point of the iteration is equal to the actual habit-adjusted aggregate consumption $\tilde{C}_1^{DH} = \tilde{C}_1$. Then the maximum Den Haan statistic for a given forecast rule $\beta_{\tilde{C}}$ is defined as

$$DH(\beta_{\tilde{C}}) \equiv \max_t |\log \tilde{C}_t - \log \tilde{C}_t^{DH}|.$$

Hence, we can interpret the Den Haan statistic as the maximum percentage error of the forecasted variable. In my model, the Den Haan statistic over 500 periods is 1%, which means that the error of the firm's prediction of $\tilde{C}_t$ is at most 1% over 500 years. The average absolute error defined in the same fashion is much smaller (0.43%).

C.4 Impulse-Response Calculation

In this subsection, I describe the calculation of the impulse response of interested variables to an exogenous shock.

With the firm's policy function obtained from the numerical solution above, I simulate $M = 2000$ economies each with length of periods $T_M = 200$. At $T_{\text{shock}} = 190$, I impose a one-standard deviation negative shock to the aggregate TFP A_t. And after the shock period T_{shock}, the aggregate TFP A_t evolves normally according to the Markov chain.

For any interested variable Y, I denote Y_{mt} as the simulated value from economy m at period t and define the impulse response of Y as

$$\text{IR}_Y = 100 \times \left(\frac{\bar{Y}_t}{\bar{Y}_{T_{\text{shock}}-2}} - 1 \right),$$

where $\bar{Y}_t = \frac{1}{M} \sum_m Y_{mt}$ is the average of Y_{mt} across all the simulated economies at time t. For the presented figures, I set $T_{\text{shock}} = 0$.

D Competition Model

In this section, I use the oligopolistic competition framework introduced by Atkeson and Burstein (2008) for the competition model.

D.1 Household

The representative household inelastically supplies fixed labor L and has a preference C that consists of consumption from a continuum of sectors j:

$$C = \left(\int_0^1 C_j^{\frac{\eta-1}{\eta}} \, dj \right)^{\frac{\eta}{\eta-1}},$$

where C_j is consumption for sector j's good, and η is the elasticity of substitution between any two different sectoral goods. Within each sector j, there are n_j firms producing differentiated goods. The household has a CES-type preference over a finite number of differentiated goods for each sector j:

$$C_j = \left(\sum_{i=1}^{n_j} C_{ij}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}},$$

where C_{ij} is consumption of good i in sector j, and ρ is the elasticity of substitution between any two differentiated goods within a sector. It is assumed that the elasticity of substitution within a sector is higher than the elasticity of substitution across a sector, $\rho > \eta$.

The household chooses consumption $\{C_{ij}\}$ to maximize C subject to the following budget constraint:

$$\int_0^1 \left(\sum_{i=1}^{n_j} P_{ij} C_{ij} \right) dj \leq WL,$$

where P_{ij} is the price of good i in sector j, and W is the nominal wage. The solution to the household's problem gives the demand function for C_{ij}:

$$C_{ij} = \left(\frac{P_{ij}}{P_j} \right)^{-\rho} \left(\frac{P_j}{P} \right)^{-\eta} C, \tag{38}$$

where P_j is sector j's price index defined as

$$P_j \equiv \left(\sum_{i=1}^{n_j} P_{ij}^{1-\rho} \right)^{\frac{1}{1-\rho}},$$

and P is total economy price index defined as

$$P \equiv \left(\int_0^1 P_j^{1-\eta}\right)^{\frac{1}{1-\eta}}.$$

D.2 Firm

Firm i in sector j produces output using labor

$$Y_{ijt} = a_{ijt} l_{ijt},$$

where a_{ijt} is producer-level productivity. Firms engage in Cournot competition within a sector.[17] Taking wage W and demand equation (38) as given, a firm i in sector j chooses its output Y_{ijt} to maximize its profit,

$$\pi_{ijt} = \max_{P_{ijt}, Y_{ijt}} \left(P_{ijt} - \frac{W}{a_{ijt}}\right) Y_{ijt}.$$

The solution to the firm's profit maximization problem is a markup over marginal cost,

$$P_{ijt} = \frac{\varepsilon(S_{ijt})}{\varepsilon(S_{ijt}) - 1} \frac{W}{a_{ijt}},$$

where firm-specific demand elasticity $\varepsilon(S_{ijt})$ is a harmonic weighted average of elasticities of substitution ρ and η,

$$\varepsilon(S_{ijt}) = \left(S_{ijt}\frac{1}{\eta} + (1 - S_{ijt})\frac{1}{\rho}\right)^{-1},$$

where S_{ijt} is a firm's market share in sector j,

$$S_{ijt} = \frac{P_{ijt} Y_{ijt}}{\sum_{i=1}^{n_j} P_{ijt} Y_{ijt}} = \left(\frac{P_{ijt}}{P_{jt}}\right)^{1-\rho}.$$

Since there are a finite number of firms in each sector, the firms are large enough ($S_{ijt} > 0$) to affect industry price index P_{jt}. Also, a firm's markup $\mathcal{M}_{ijt}$ can be expressed as

$$\frac{1}{\mu_{ijt}} = \frac{\rho - 1}{\rho} - \left(\frac{1}{\eta} - \frac{1}{\rho}\right) S_{ijt}. \tag{39}$$

Since $\rho > \eta$, markup is an increasing and convex function of market share.

[17] Bertrand competition generates qualitatively the same results.

D.3 Calibration

I take the 4-digit level market share S_{ijt} in the data to calculate the markup from equation (39). And I choose the elasticity of substituion across sector η to be 1.01 and set the elasticity within sector ρ to be 3.10 so that the average markup is equal to 1.50.

D.4 Quantitative Analysis

To compare the model to the data, I run the regression specified by 31 for each size quintile of firms. In contrast with the data, the competition model predicts that larger firms have more counter-cyclical markups, as shown in Figure 11. This is because the markup is an increasing convex function of the markup in the model. Firms' market shares are counter-cyclical in the data since there are less remaining firms in recessions. This implies counter-cyclical markups, but more so for the large firms due to the convexity property of the competition model.

Also, the model could only explain 12% of the cross-sectional variation of estimated markups in terms of standard deviation, as shown in Table 11. One reason is because the markups are constant for most of the firms. The markup of the 5th percentile is not much different from the 75th percentile, and it only starts slightly increasing at the 95th percentile. This could also be attributed to the convexity of the mapping between the markup and market share.

E Sticky-Price Model

In this section, I examine a sticky-price model in which a firm needs to pay a fixed cost for price adjustment.

E.1 Household

The representative household has an additively separable preference over consumption and labor and maximizes the following:

$$\max \mathrm{E}_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma}}{1-\sigma} - \omega \frac{L_t^{1+\psi}}{1+\psi} \right) \right\},$$

where $\frac{1}{\sigma}$ is the inter temporal elasticity of substitution (IES), ω is the disutility parameter from labor, and $\frac{1}{\psi}$ is the Frisch elasticity of labor supply. And C_t is the Dixit-Stiglitz aggregator of differentiated goods consumption over varieties i,

$$C_t = \left(\int_0^1 c_{it}^{\frac{\theta-1}{\theta}} di \right)^{\frac{\theta}{\theta-1}}.$$

The budget constraint for the household is

$$\int_0^1 p_{it} c_{it} di + \mathrm{E}_t[Q_{t,t+1} B_{t+1}] \leq B_t + W_t L_t + \int_0^1 \pi_{it} di.$$

A complete set of Arrow-Debreu state-contingent assets is traded so that B_{t+1} is a random variable that delivers payoffs in period $t+1$. $Q_{t,t+1}$ is the stochastic discount factor used to price them.

The first-order conditions of the household's maximization problem is

$$\frac{W_t}{P_t} = \omega \frac{L_t^{\psi}}{C_t^{-\sigma}},$$
$$Q_{t,t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\sigma} \frac{P_t}{P_{t+1}},$$

where P_t is the aggregate price level. The demand equation for each good is the following:

$$c_{it} = \left(\frac{p_{it}}{P_t} \right)^{-\theta} C_t.$$

Finally, I assume that the aggregate nominal value-added $S_t \equiv P_t C_t$ follows an exogenous random walk:

$$\log S_t = \log S_{t-1} + \mu_S + \eta_t, \quad \eta_t \sim N(0, \sigma_S).$$

We can think of this as the central bank having a targeted path of nominal value-added, and it does so by adjusting interest rate accordingly.

E.2 Firms

Each firm produces output c_{it} using a technology in labor l_{it},

$$c_{it} = a_{it} l_{it},$$

where a_{it} is firm-specific idiosyncratic productivity, which follows an AR(1) process

$$\log a_{it} = \rho_a \log a_{it-1} + \epsilon_{it}, \quad \epsilon_t \sim N(0, \sigma_a).$$

To change its price, a firm must pay a fixed cost κ in units of labor. Hence, a firm's nominal profit equals to

$$\pi_{it} = \left(p_{it} - \frac{W_t}{a_{it}}\right)\left(\frac{p_{it}}{P_t}\right)^{-\theta} C_t - \kappa W_t I_{p_{it} \neq p_{it-1}}.$$

E.2.1 Krusell-Smith Forecast Rule

To solve the model in general equilibrium, it is necessary to keep track of the distribution of firms over idiosyncratic productivities and prices, and thus determine the aggregate price level. Here, I assume that the aggregate price level itself is self predictable. In particular, I assume that each firm perceives a Krusell-Smith type law of motion for S_t/P_t,

$$\log \frac{S_t}{P_t} = \gamma_0 + \gamma_1 \log \frac{S_t}{P_{t-1}}.$$

Given this conjecture, a firm's state variables are: (i) last period's individual price over the nominal value-added $\frac{p_{it-1}}{S_t}$, (ii) idiosyncratic productivity a_{it}, and (iii) ratio of nominal value-added over aggregate price level $\frac{S_t}{P_t}$. And a firm's problem can be written recursively in real term as

$$V\left(\frac{p_{it-1}}{S_t}, a_{it}, \frac{S_t}{P_t}\right) = \max_{p_{it}} \left\{\frac{\pi_{it}}{P_t} + \mathrm{E}_t\left[Q_{t,t+1} V\left(\frac{p_{it}}{S_{t+1}}, a_{it+1}, \frac{S_{t+1}}{P_{t+1}}, \kappa_{it+1}\right)\right]\right\}.$$

E.3 Recursive Competitive Equilibrium

A recursive competitive equilibrium is a law of motion (γ_0, γ_1), a set of price-level paths $\{P_t\}$, and a set of wage paths $\{W_t\}$ that are consistent with the:

1. Household utility maximization problem;
2. Firm profit maximization problem;
3. Goods market clearing;
4. Arrow-Debreu market clearing; and
5. Evolution of nominal aggregate demand S_t and idiosyncratic productivity a_{it}.

E.4 Calibration

One period in the model equals one month in the data. The monthly discount factor is $\beta = 0.997$. For the representative household, I assume log utility in consumption $\sigma = 1$ and an infinite Frisch elasticity of labor supply $\psi = 0$ as in Hansen (1985) and Rogerson (1988). Hence, the real wage is a linear function of the aggregate consumption $W_t/P_t = \omega C_t$; this means that we do not need to keep the aggregate labor supply as a state variable.

For elasticity of substitution, I set $\theta = 3$ to match average markup of 1.50. The growth rate and standard deviation of value-added S_t are calibrated to match those of French data. A firm's idiosyncratic productivity is set to be $\rho_a = 0.7$, a value often used in the sticky-price literature. I jointly set the standard deviation $\sigma_a = 0.024$ and the adjustment cost $\kappa = 0.28\%$ to match the price adjustment frequency 13.7% and the size of price change 4.7% as documented for French firms in Gautier and Le Bihan (2018).

E.5 Quantitative Analysis

The reason that a standard New Keynesian model could generate countercyclical markup is the following: Under monopolistic competition and constant consumer price elasticity θ, a firm's optimal pricing strategy is a constant markup $\frac{\theta}{\theta-1}$ over marginal cost. However, with price stickiness, a procyclical marginal cost implies that in a boom, the gap between the price and the marginal cost shrinks, and hence decreases in the markup. And the menu cost is independent of the firm size, so small firms would adjust prices less frequently.

As shown by results of regression 31 in Figure 11, the menu cost model matches the heterogeneous cyclicality well. However, Table 11 shows that it explains less than 4% of the markup standard deviation. This is because the desired markup $\frac{\theta}{\theta-1}$ is the same for all firms.

F Empirically Comparable Variables

For the calibration, I need to define several variables that are consistent with the measurement in the data. The reasons are twofold. First, the construction of consumer price index (CPI) in the real world suffers from the well-known new-product bias since it does not put much weight on utility gain from product variety. Even if it adjusts for love-of-variety effect, it does so at a lower frequency than in my model. Second, as seen from (14), the welfare-consistent price index is adjusted by habit stocks, which is different from the standard composite price index. Hence, I need to adjust for these two biases to match moments of aggregate output.

For any variable J_t in the model, I denote J_t^D for a data-consistent counterpart of it. Then, the CPI $\tilde{p}_t^D$ is given by $\tilde{p}_t^D \equiv N_t^{\frac{1}{\rho-1}} \left(\int_{i \in I_t} p_{it}^{1-\rho} di \right)^{1/(1-\rho)}$, and a data-consistent counterpart for aggregate output is $X_t^D \equiv (\tilde{p}_t X_t)/p_t^D$.[18] With some algebra, it can be written as

$$X_t^D = \underbrace{N_t^{-\frac{1}{\rho-1}}}_{\xi\text{-bias}} \times \underbrace{\left[\frac{\int_{i \in I_t} \left(\mu_{it} l_{it}^{1-\alpha} a_{it}^{-1} b_{it}^{-\theta} \right)^{1-\rho} di}{\int_{i \in I_t} \left(\mu_{it} l_{it}^{1-\alpha} a_{it}^{-1} \right)^{1-\rho} di} \right]^{\frac{1}{1-\rho}}}_{\theta\text{-bias}} \times X_t,$$

which has two bias components, ξ-bias and θ-bias. These stand for bias due to love-of-variety and deep habit formation, respectively. Bilbiie, Ghironi, and Melitz (2012) emphasize the ξ-bias in their quantitative analysis over the business cycle. In the real world, a statistical agent adjusts for the new product bias in the CPI calculation, but at a slow pace.[19] θ-bias measures the welfare difference due to good-specific habit formation. And to see the relation of dynamics between X_t^D and X_t more clearly, I restrict to a symmetric equilibrium case,[20] then the θ-bias can be simplified to b^θ, and growth in aggregate output is

$$\Delta \log X_t^D = -\frac{1}{\rho - 1} \Delta \log N_t - \theta \Delta \log b_t + \Delta \log X_t.$$

[18] This can be seen by considering a more general case of the Dixit-Stiglitz preference:

$$X_t = N_t^{\xi - \frac{1}{\rho-1}} \left[\int_{i \in I_t} \left(c_{it} b_{it}^{\theta} \right)^{\frac{\rho-1}{\rho}} di \right]^{\frac{\rho}{\rho-1}},$$

where ξ measures the degree of love-of-variety. Its corresponding aggregate price indicator is $\tilde{p}_t = N_t^{-\xi + \frac{1}{\rho-1}} \left(\int_{i \in I_t} p_{it}^{1-\rho} di \right)^{1/(1-\rho)}$. $\xi = \frac{1}{\rho-1}$ corresponds to the usual Dixit-Stiglitz case, while $\xi = 0$ if there is no love-of-variety.

[19] The UK added smartphone into the basket of goods in 2011.

[20] Note that even if all firms have the same productivity, they might still differ in size, due to the fact that an entrant's customer capital is different from an incumbent's, and it needs time to accumulate customer capital to catch up with the incumbent. Hence, to guarantee a symmetric equilibrium, one needs a time-varying initial customer capital for an entrant, which is equal to an incumbent's.

In the simulation, the number of incumbents N_t is procyclical, and b_t is procyclical with a lag of one period, hence X_t^D tends to underestimate the true growth rate of output in absolute terms.

References

[1] Den Haan, W. J. Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents. *Journal of Economic Dynamics and Control 34*, 1 (2010), 79–99.

[2] Gautier, E., and Le Bihan, H. Shocks vs menu costs: Patterns of price rigidity in an estimated multi-sector menu-cost model. *Working Paper* (2018).

[3] Krusell, P., and Smith, Jr, A. A. Income and wealth heterogeneity in the macroeconomy. *Journal of political Economy 106*, 5 (1998), 867–896.

[4] Tauchen, G. Finite state markov-chain approximations to univariate and vector autoregressions. *Economics letters 20*, 2 (1986), 177–181.

[5] Young, E. R. Solving the incomplete markets model with aggregate uncertainty using the krusell–smith algorithm and non-stochastic simulations. *Journal of Economic Dynamics and Control 34*, 1 (2010), 36–41.

G Tables

Table 8: Heterogeneity in Markup Cyclicality ϕ: First-difference Regression

Dep. Variable $\Delta \log \mu_{it}$	(1) All	(2) Large	(3) Small	(4) All	(5) All
$\Delta \log Y_t$	-1.19	-1.02	-1.23	-1.23	-1.24
	(0.25)	(0.22)	(0.25)	(0.25)	(0.26)
$\Delta \log Y_t \times \text{Large}_i$				0.21	
				(0.11)	
$\Delta \log Y_t \times s_{it}$					4.00
					(1.04)
N	66696	12452	54244	66696	66696
adj. R^2	0.01	0.01	0.01	0.01	0.01

Note: μ_{it} is the firm-level inverse of material share of revenue. $\log Y_t$ is log aggregate output quadratically detrended. $Large_i$ is an indicator for a firm with more than 1% average market share within a 4-digit industry. s_{it} stands for market share in a 4-digit industry. The standard errors are clustered at the time level and are reported in parentheses.

Table 9: Heterogeneity in Markup Cyclicality ϕ: Cross-country Analysis

Dep. Variable $\log \mu_{it}$	(1) France	(2) Germany	(3) Italy	(4) Spain
$\log Y_t$	-1.21	-1.34	-1.88	-0.49
	(0.16)	(0.11)	(0.39)	(0.38)
$\log Y_t \times \text{Large}_i$	0.45	0.45	0.54	0.35
	(0.12)	(0.09)	(0.28)	(0.05)
Firm FE	Yes	Yes	Yes	Yes
N	96507	62844	409837	269211
adj. R^2	0.89	0.83	0.86	0.77

Note: μ_{it} is the firm-level inverse of material share of revenue. $\log Y_t$ is log aggregate output quadratically detrended. $Large_i$ is an indicator for a firm with more than 1% average market share within a 4-digit industry. All specifications include firm fixed effects (FEs). The standard errors are clustered at the time level and are reported in parentheses.

Table 10: Firm Age and Financial Variables

Dep. Variable $\log \mu_{it}$	(1)	(2)	(3)	(4)	(5)	(6)
$\log Y_t$	-1.21	-0.81	-0.96	-1.46	-0.95	-1.08
	(0.16)	(0.19)	(0.16)	(0.22)	(0.16)	(0.20)
$\log Y_t \times s_{it}$	3.52	3.78	3.61	3.73	3.54	3.58
	(1.61)	(1.49)	(1.62)	(1.55)	(1.63)	(1.67)
$\log Y_t \times \log \mathrm{Age}_{it}$		-0.31				
		(0.20)				
$\log Y_t \times \mathrm{LIQD}_{it}$			-0.14			
			(0.04)			
$\log Y_t \times \mathrm{CASH}_{it}$				1.95		
				(1.57)		
$\log Y_t \times \mathrm{CURR}_{it}$					-0.10	
					(0.04)	
$\log Y_t \times \mathrm{SOLV}_{it}$						-0.22
						(0.26)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
N	96507	96507	96493	96368	96493	96255
adj. R^2	0.89	0.89	0.89	0.90	0.89	0.89

Note: μ_{it} is the firm-level inverse of material share of revenue. $\log Y_t$ is log aggregate output quadratically detrended. s_{it} stands for market share in a 4-digit industry. Age_{it} stands for a firm's age. $LIQD_{it}$ stands for liquidity ratio, $CASH_{it}$ for cash ratio, $CURR_{it}$ for current ratio, and $SOLV_{it}$ for solvency ratio. All financial variables are winsorized at 1% and 99% levels. The standard errors are clustered at the time level and are reported in parentheses.

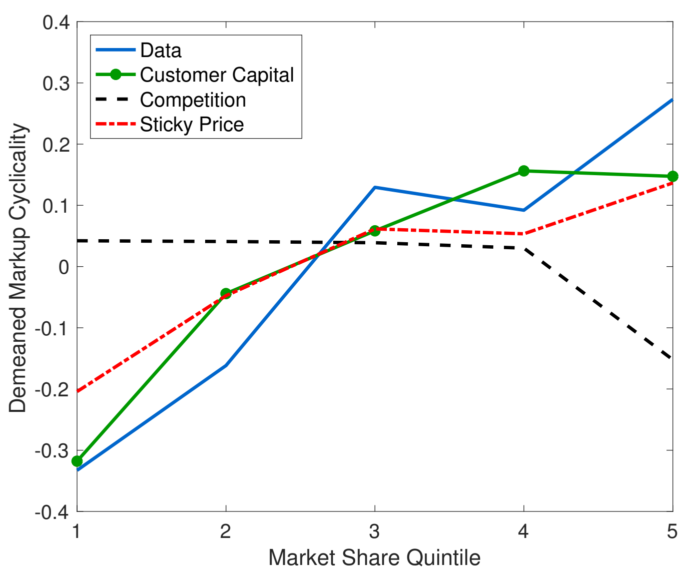
Table 11: Markup Estimates: Data vs Alternative Models

Moments	(1) DLW	(2) Customer Capital	(3) Competition	(4) Sticky Price
Mean	1.58	1.50	1.50	1.50
Std. Log	0.52	0.31	0.06	0.02
p5	0.58	0.88	1.48	1.45
p25	0.83	1.18	1.48	1.48
p50	1.13	1.40	1.48	1.50
p75	1.68	1.76	1.49	1.53
p95	4.01	2.36	1.56	1.56
p75/p25	2.01	1.50	1.01	1.03

Note: This table shows De Loecker-Warzynski markup estimates and model-simulated markups. Column 1 shows De Loecker-Warzynski markup estimates with a Cobb-Douglas specification. Column 2 shows the results of the baseline customer capital model. Column 3 shows the results of the Atkeson-Burstein (2008) model. Column 4 shows the results of the menu-cost model.

H Figures

Figure 11: Markup Cyclicality by Firm Size: Data vs Alternative Models



Note: The figure plots the coefficient estimate of regression 31 for each quintile of market share. The figure compares the case of empirical estimates of markup and three alternative cases of model simulations. The solid blue line shows data estimates, the green line with circles shows the simulation of customer capital, the black dashed line shows the simulation of the Atkeson-Burstein (2008) model, and the red dash-dot line shows the simulation of the menu-cost model. All values are demeaned within each case.