

Customer Capital, Markup Cyclical, and Amplification

Sungki Hong*

Princeton University

January 13, 2017

[\[Link to the Latest Version\]](#)

Abstract

This paper studies the importance of firm-level price markup dynamics for business cycle fluctuations. The first part of the paper uses 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 countercyclical with an average elasticity of -0.9 with respect to real GDP, in line with the earlier industry-level evidence. Importantly, I find substantial heterogeneity in markup cyclical across firms, with small firms having significantly more countercyclical markups than large firms. In the second part of the paper, I develop a general equilibrium model that matches these empirical findings and explore its implications for business cycle dynamics. In particular, I embed *customer capital* (due to deep habits as in Ravn, Schmitt-Grohe, and Uribe 2006) into a standard Hopenhayn (1992) model of firm dynamics with entry and exit. A key feature of the model is that a firm's decision about markups becomes dynamic – firms accumulate customer capital in the periods of fast growth by charging low markups, and choose to exploit it by charging high markups in the downturns. In particular, during recessions, the endogenous higher exit probability for smaller firms implies that they place lower weight on future profits, leading them to charge higher markups. This mechanism serves to endogenously increase the dispersion of firm sales and employment in recessions, a property that is consistent with the data. I further show that the resulting input misallocation amplifies both the volatility and persistence of the exogenous productivity shocks driving the business cycle.

*Email: sungki@princeton.edu. I am deeply grateful to my advisers Oleg Itskhoki, Richard Rogerson, Jakub Kastl, and Mark Aguiar for their continual guidance and support. I am also indebted to Jan De Loecker, John Kim, Nobu Kiyotaki, Ben Moll, Edurado Morales, Dmitry Mukhin, Ezra Oberfield, Esteban Rossi-Hansberg, Chris Sims, and all seminar participants at Princeton University for useful suggestions and comments.

1 Introduction

Understanding dynamics of price-cost markups is an important issue in business cycle analysis. Chari, Kehoe, and McGrattan (2007) find that the movements of the *labor wedge*¹ account for much of aggregate fluctuations, and Bilts, Klenow, and Malin (2016) find that variation in markups explains at least half of the variation in the labor wedge. Also, endogenous markup variation is an important propagation mechanism for business cycle models to generate a more volatile employment response. Moreover, markup cyclicity 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 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 the 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 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 firm 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 sectors². The dataset contains both

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

²I use France as the benchmark country since it has longer time period (2003 – 2013) available for online download

production data and financial balance sheet data. However, the dataset provides 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 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 measurement of true costs of labor inputs suffers from several frictions, including but not limited to hiring and firing cost, overhead costs, and etc (as discussed by Rotemberg and Woodford 1999). Also, I check robustness for functional forms for 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 -1.2 with respect to real GDP with 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% - 1.2%. This is in line with the recent findings by Bils, Klenow, and Malin (2015) using US industry-level data. Furthermore, taking full advantage of firm level data, I find substantial heterogeneity of markup cyclicity among firms. In particular, if I split firms based on their revenue share within own industries, I find that small firms' markups are 50% 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 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 firm's characteristics that matters 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 cyclicity remains significant, controlling for the age effect. The other possibility is that firm size represents 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

from WRDS than other three countries.

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-Grohe, and Uribe 2006) into a standard Hopenhayn (1992) model of firm dynamics with entry and exit. The driving force of 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 the booms, and low in the recessions, resulting in countercyclical markups. Also, the endogenous firm entry and exit decision also affects firms pricing. During recessions, the endogenous probability of 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 exit affects them more, resulting in more countercyclical markups.

Taken together, the model is able to replicate the two main empirical findings qualitatively, and quantitatively as well: 1) firms have countercyclical markups, and 2) small firms have more countercyclical markups than large firms. Furthermore, while the model can explain around one-third to two-thirds of the magnitude of markup elasticity with respect to aggregate output, it closely captures the heterogeneity of markup cyclicity between small and large firms as estimated in the data.

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 increase in dispersion of firm size. This implication has support in the data. In the model, small firms’ deflated sales are 13% more responsive to aggregate output than large firms’, while the difference is 22% in the data. Notice that the model generates the countercyclical dispersion of firm size without relying on exogenous second moment shock in productivity growth as in Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2014). Indeed, 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 not only on 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 misallocations 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 effects 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 cyclicity of markup. Bilal (1987), Rotemberg and Woodford (1999), and Nekarda and Ramey (2013) use labor input to test the markup cyclicity at the aggregate level. Other studies infer the cyclicity using inventories (Kryvtsov and Midrigan 2012), retail prices (Stroebel and Vavra 2014), and advertising expenses (Hall 2014). I follow Bilal, Klenow, and Malin (2013) 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 cyclicity at the firm level. And the paper is also related to the literature estimating heterogeneity in markup cyclicity. Bilal, Klenow, and Malin (2012) study how markup cyclicity 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 raise prices more during the recent financial crisis. My work exclusively explores the heterogeneity of markup cyclicity due to firm size, which endogenously affects 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, Steinsson 2003). Other models try to ground the source of markup variation. In the New Keynesian model, sticky price with a procyclical marginal cost generates 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 model³. 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, firm’s pricing in my model is driven by intertemporal decision of customer capital accumulation.

³Moreira (2015) builds a similar model to this paper to study effects of initial business conditions on entrants’ size. However, there are several difference between her work and mine. First, the focus of my paper is markup cyclicality, while her paper focuses on size of new born businesses. Second, in terms of methodology, it is a partial equilibrium model in her 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. Also, I can quantify the welfare implication of the model.

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 misallocation of inputs can lead to 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 cyclicity along 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 the next section, 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. In Section 6, I describe alternative preference specification. I then conclude in Section 7.

2 Empirical Study of Markup Cyclicity

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 discuss potential economic mechanisms.

2.1 Estimation of Markup

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

$$\mu_{it} = \frac{P_{it}}{MC_{it}}. \quad (1)$$

In practice, we seldom observe neither price nor marginal cost in data. To overcome this problem, I follow the insights from Hall (1986) and De Loecker and Warzynski's (2012) and procedures in the literature that relies on firm's cost minimization condition with respect to static inputs that do not have 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:

$$Y_{it} = F_{it}(M_{it}, L_{it}, K_{it}, A_{it}), \quad (2)$$

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

$$\mathcal{L}(M_{it}, L_{it}, K_{it}, \lambda_{it}) = P_{it}^M M_{it} + W_{it} L_{it} + R_{it} K_{it} + \lambda_{it}(Y_{it} - F_{it}(M_{it}, L_{it}, K_{it}, A_{it})), \quad (3)$$

where P_{it}^M , W_{it} , and R_{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} . In large, 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 firm's dynamic decisions. The firm's optimality condition for the material input M_{it} is

$$P_{it}^M = \lambda_{it} \frac{\partial F_{it}}{\partial M_{it}}. \quad (4)$$

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

$$\mu_{it} = \theta_{it}^M (\alpha_{it}^M)^{-1}, \quad (5)$$

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, firm's decision about dynamic inputs involves solving the full dynamic problem. For example, if there is adjustment costs for labors, such as hiring and firing costs, a firm has to consider

how labor decision affects not only today's outcome, but also future's expected profits. Hence, the wedge between the output elasticity and labor share captures an additional wedge caused by the dynamic decisions. In the 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 the adjustment costs, the markup is actually countercyclical.

Due to the reasons stated above, I use material input to examine markup cyclicity. Furthermore, I run my analysis at the firm level, which is different from previous analysis conducted at the industry level. Industry-level markup is a revenue-weighted average markup, and thus the results includes the compositional effect of firm sizes. However, as I discuss below, it is important to examine markup cyclicity 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 of production function, I restrict to two functional forms of production function commonly used in the literature: Cobb-Douglas and Translog. For the case of Cobb-Douglas form, the production function has the following form:

$$y_{it} = \theta_m m_{it} + \theta_l l_{it} + \theta_k k_{it} + a_{it}, \quad (6)$$

where lower case letters denote the natural logarithm of variables. Note that all the output elasticity terms θ are constants, and hence there is no need for estimation since I am mostly interested in the dynamics of markup, not level. 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 Translog form, which allows for more flexibility in terms of production function estimation. Specifically, the Translog production function is:

$$\begin{aligned} y_{it} = & \theta_m m_{it} + \theta_l l_{it} + \theta_k k_{it} + \theta_{mm} m_{it}^2 + \theta_{ll} l_{it}^2 + \theta_{kk} k_{it}^2 \\ & + \theta_{ml} m_{it} l_{it} + \theta_{mk} m_{it} k_{it} + \theta_{lk} l_{it} k_{it} + a_{it}. \end{aligned} \quad (7)$$

In this case, the output elasticity is a function of inputs⁴, hence ignoring the nature of time varying elasticity might bias the result of markup cyclical. 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 Akerberg, 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 set, I do not have firm-specific price deflator, rather I deflate the firm revenue with industry-level price index. Hence, the firm specific price deflator enters the error term, and potentially causes bias in the estimation⁵. 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 dataset⁶. This dataset has recently been used in many studies⁷. I focus on manufacturing sector for it suits the estimation of production function. I only use the firm labeled as “unconsolidated” to avoid double-counting the sales with sales in its parent company. The dataset contains 12,392 firm observations on average each year from 2003 to 2013⁸.

I now define the variables used in the empirical analysis. An *industry* is defined as a 4-digit NACE Rev. 2 code. I measure 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 input* M_{it} with total material costs. I use the number of employees as measure of *labor input* L_{it} . I measure the *capital input*

⁴For example, output elasticity with respect to material is

$$\frac{\partial f_{it}}{\partial m_{it}} = \theta_m + 2\theta_{mm}m_{it} + \theta_{ml}l_{it} + \theta_{mk}k_{it},$$

which is a function of material, labor, and capital.

⁵De Loecker and Goldberg (2014) has a thorough discussion of potential issues with not controlling firm specific output and input prices.

⁶The four countries include France, Germany, Italy, and Spain. I use France as my baseline case, since other countries have shorter time period available for online download.

⁷For example, Gopinath, Kalemli-Ozcan, Karabarbounis, and Villegas-Sanchez (2015) use the same dataset to study capital misallocation in European countries.

⁸I understand that the time period is relatively short for business cycle study. However, this is the best dataset I can get for firm-level study at this stage. Currently, I am applying for the access to US Census dataset, which covers longer time period, and I would do the same analysis on it as soon as I get the access.

K_{it} with tangible fixed assets. To control for the input quality across firms in 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 2-digit industry level⁹.

I take business cycle measures from Eurostat, and Federal Reserve website. We use four measures of business cycle: (1) real GDP, (2) total employment, (3) total working hours, and (4) recession period defined by NBER. All business cycle variables are quadratically detrended.

2.3 Empirical Modeling Strategy

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

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

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

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

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

2.4 Empirical Results

Cobb-Douglas

I begin my empirical analysis with the case of Cobb-Douglas production. With this specification, the movement in markup is proportional to the inverse of material share. I evaluate the markup cyclicity 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. The column (1) shows that markup is countercyclical and

⁹Details of data cleaning process will be available in online appendix.

statistically significant for all firms. The elasticity of markup with respect to aggregate output is -1.2 . This means that one-percent increase of output from its non-linear trend causes the markup to decline by 1.2 percent.

Moreover, with micro-level data, I find that there is substantial heterogeneity in markup cyclicity. One interesting and important aspect to look at is whether the markup cyclicity depends on firm size. Markups of small and large firms potentially move differently for several reasons: market power, pricing frictions, and 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 4-digit industry. Column (2) and (3) report the results for subsample of large and small firms, respectively. I find that large firms markup elasticity to be small, equal to -0.9 , and small firms have larger elasticity of -1.3 . This implies that small firms' markups fluctuate 50% more than large firms' along business cycle. Recently, Bils, Klenow, and Malin (2016) use intermediate input to estimate markup cyclicity at industry level and find that elasticity with respect to real GDP to be 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 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 Table 2. I find the same results as with fixed-effect regression: markup is countercyclical, and small firms' markups fluctuate more.

Translog

Although Cobb-Douglas is a convenient assumption for estimation, it has a strong restriction in terms of specification that output elasticities to be 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 Translog production function, which allows for higher order and interaction terms of inputs. However, as discussed above, since I do not have firm-specific output deflator,

the estimations of Translog production function is biased. Nevertheless, I use this alternative specification for robustness check.

I report the fixed-effect regression results in Table 3. The results are consistent with 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 business cycle, possibly due to different factor prices, adjustment costs and etc. Second, the cyclical difference between small and large firms are larger. This means that firms adjust its 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.

Robustness Check

Here, I investigate whether the results above are robust to alternative samples and. 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 robustness check. Results are reported in Table 4. All the results are robust except for Spain.

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 cyclical could be due to age effect. I check this by adding an additional interaction term between aggregate output and age. The results are in column (1) of Table 5. The coefficient of the interaction term with market share remains strongly positive and statistically significant, after controlling for age. Hence, the size effect is important for firm's pricing behavior.

Also, a recent study by Gilchrist, Schoenle, Sim, and Zakrajsek (2015) finds that financially constrained firms raise markups more to avoid using costly external financing to pay fixed payment of operation during the Great Recession¹⁰. And firm size might be proxying the financial conditions of the firms. Therefore, I use the following four financial variables to control for financial constraints, which are *cash ratio* – a ratio of cash and cash equivalents to current liabilities, *current ratio* – a ratio of current assets to current liabilities, *liquidity ratio* – a ratio of liquid assets to current liabilities, and *solvency ratio* – a ratio of cashflow to total liabilities. The results are reported in column (2) to (5). First, we see that the size effect remains strongly positive. Second, the coefficients of

¹⁰An earlier study by Chevalier and Scharfstein (1996) find that more financially constrained supermarkets have more countercyclical markups.

interaction term with all financial variables are small and statistically insignificant.

Overall, these results suggest that firm size plays a prominent role in determining the markup cyclicity. Therefore, any theory or model of markup cyclicity has to be consistent with the documented size effect.

2.4.1 Potential Mechanisms

Next, I investigate which model can successfully replicate the two empirical findings that not only firms have countercyclical markups, but also small firms' markups fluctuate more.

In Hong (2016), I test the markup behavior of the following two prominent models in the literature: (i) an oligopolistic competition model, and (ii) a New Keynesian model with heterogeneous price stickiness. First, I explore the Atkeson and Burstein (2008) model of oligopolistic competition, in which markups are an increasing function of firm market shares. Coupled with an exogenous *uncertainty* shock as in Bloom (2009), i.e. a second-moment shock to firm productivities in recessions, this model results in a countercyclical average markup, as in the data. However, in contrast with the data, this model predicts that smaller firms reduce their markups. Second, I calibrate both Calvo and menu cost models of price stickiness to match the empirical heterogeneity in price durations across small and large firms, as in Goldberg and Hellerstein (2011). I find that both models can match the average counter-cyclicity of markups in response to monetary shocks. Furthermore, since small firms adjust prices less frequently, they exhibit greater markup counter-cyclicity, consistent with the empirical patterns. Quantitatively, however, only the menu cost model, through endogenous selection of firms adjusting prices, matches the extent of the heterogeneity in markup cyclicity in the data. Also, the markup cyclicity depends on the source of shock. In response to productivity shocks, both sticky price models imply pro-cyclical markup behavior.

Therefore, I propose a new general equilibrium model that features firm's endogenous accumulation process of customer capital, and entry and exit decision. In the model, firms' markup decisions become dynamics, and the decisions of firms closer to the extensive margin of exit are rather different from others. I describe the main intuition and mechanism of the model below.

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. As 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 customer capital idea in a tractable way, I adopt the framework of *deep habits* model developed by Ravn, Schmitt-Grohe, 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 set-up is able to generate heterogenous markup cyclicalities. 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 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.

Many recent papers have enriched the *deep habits* model to study firm dynamics. There are two papers closely related to my model. One paper is by Gilchrist, Schoenle, Sim, and Zakrajsek (2015), in which firms face financial constraint while making pricing decisions in customer-markets. The other paper is Moreira (2015), which uses the similar framework of customer capital and firm entry-exit dynamics. It studies the procyclical entrant size in a partial equilibrium model. However, my model differs in the following aspects: (i) My model is a general equilibrium model with heterogeneous firms, hence I can study the GE effects, the interaction among firms, and welfare implications over business cycle; (ii) When making decisions to maximize profits, no firms are subject to either nominal rigidity in price setting, or financial frictions; (iii) While Moreira (2015) studies

long-run implication on entrant's size, my model focuses on markup dynamics and its implications in short run.

3.1 Household

Time is discrete and is indexed by $t = 1, 2, 3, \dots$, 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

$$E_t \sum_{\tau=0}^{\infty} \beta^\tau \left[\frac{X_{t+\tau}^{1-\sigma}}{1-\sigma} - \omega \frac{L_{t+\tau}^{1+\nu}}{1+\nu} \right], \quad (10)$$

where X_t is habit-adjusted consumption bundle, and L_t is labor. Household has constant relative risk aversion σ , disutility of labor ω , and inverse Frisch elasticity of labor supply ν .

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

$$X_t = \left[\int_{i \in I_t} \left(c_{it} b_{it}^\theta \right)^{\frac{\rho-1}{\rho}} di \right]^{\frac{\rho}{\rho-1}}, \quad (11)$$

where c_{it} denotes household's consumption of variety i . 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 consumption¹¹. External habit stock b_{it} evolves according to

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

where $\delta \in [0, 1]$ governs both depreciation rate of past habit stock, and conversion rate of consumption into habit stock¹². Finally, $\rho > 1$ governs the intratemporal elasticity of substitution across habit-adjusted consumption, and θ governs the degree of habit formation.

Household's demand for variety i is given by:

$$c_{it} = \left(\frac{p_{it}}{\bar{p}_t} \right)^{-\rho} b_{it}^{\theta(\rho-1)} X_t, \quad (13)$$

¹¹Nakamura and Steinsson (2011) study the case when the household takes habit formation as internal, and find that firms face time-inconsistency problem in this case.

¹²Paciello, Pozzi, and Trachter (2015) microfound this law of motion for customer flow in a model, where customers constantly search for goods with search costs to buy from a firm which provides lower price. We can see (12) as a reduced form of this search process.

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

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

The household is subject to the following budget constraint:

$$\tilde{p}_t X_t + E_t(q_{t,t+1} D_{t+1}) = D_t + w_t L_t + \Phi_t, \quad (15)$$

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

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

$$\frac{w_t}{\tilde{p}_t} = \omega L_t^\nu X_t^\sigma, \quad (16)$$

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

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

3.2 Firm

Incumbent

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

$$y_{it} = A_t a_{it} l_{it}^\alpha, \quad (18)$$

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 technology¹³. A_t and a_{it} follow the AR(1) processes respectively:

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

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

¹³Aggregate 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 supply-side shock or demand-side shock. I have constructed a partial equilibrium model with an exogenous aggregate demand, and all the results in this paper hold. This is in contrast to a New Keynesian model, in which markup is counter-cyclical against demand-side shock, while pro-cyclical against supply-side shock.

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 $\Lambda_t = \{A_t, \Gamma_t\}$.

Each period t , incumbent i faces demand (13), and takes into account that the habit stock b_{it} evolves as (12). After production, 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-adjusted 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

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

s.t. constraints (12), (13), and (18) satisfy. (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 firm-specific threshold value of operating cost implicitly defined by

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

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

The mass of exiting firms at time t is

$$N_t^x = \int \int \Pr[\psi \geq \psi^*(b, a, \Lambda)] d\Gamma_t(b, a), \quad (23)$$

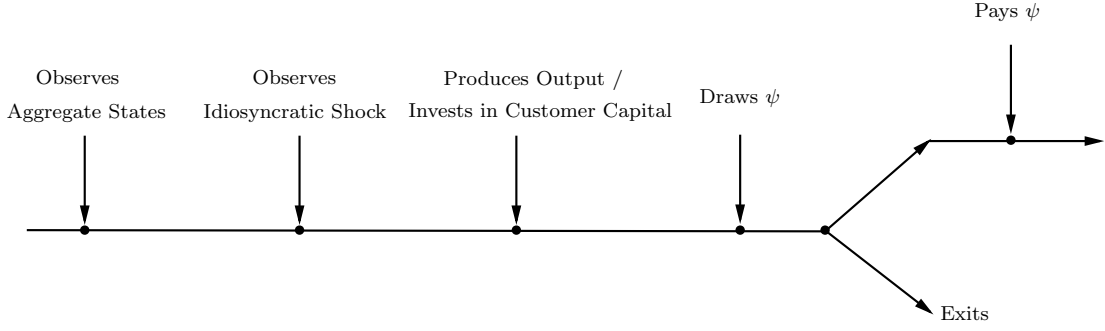
where x denotes *exit*. And the mass of incumbents continue to next period is

$$N_t^c = N_t - N_t^x, \quad (24)$$

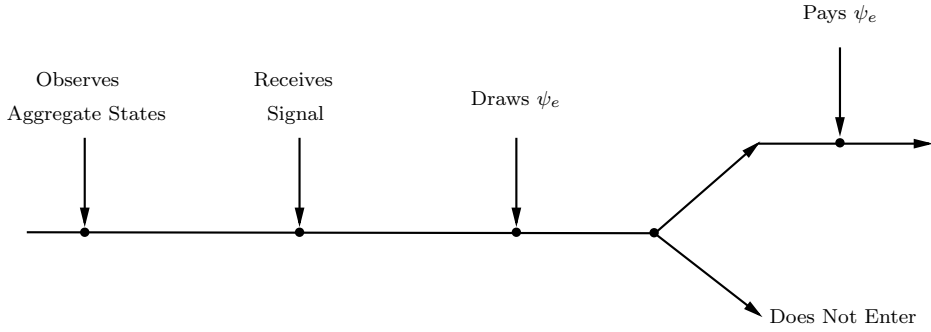
where c denotes *continuation*.

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

(i) Incumbent



(ii) Potential Entrant



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):

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

where μ_{it} is firm i 's price markup over marginal cost, $\bar{\mu} \equiv \frac{\rho}{\rho-1}$ is monopolistic firm's constant price markup, $G(\psi_{it}^*)$ is firm i 's probability of staying in the market in the next period $t + 1$, $q_{t,t+1}$ 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 $\mu_{it} < \bar{\mu}$ ¹⁴. This is because a firm has an incentive to *invest*

¹⁴One can easily see this by iterating (25), and get the following:

$$\mu_{it}^{-1} - \bar{\mu}^{-1} = \frac{\delta}{1 - \delta} \frac{\theta(\rho - 1)}{\rho} \text{E}_t \left\{ \sum_{j=1}^{\infty} q_{t,t+j} (1 - \delta)^j \left[\prod_{j'=0}^j G(\psi_{it+j'}^*) \right] \frac{p_{it+j} y_{it+j}}{p_{it} b_{it+j}} \right\}$$

Since all the right-hand-side terms are positive, we have $\mu_{it} < \bar{\mu}$.

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 $\delta = 1$ ¹⁵. 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:

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

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 $q_{t,t+1}$; (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 cyclicity. 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 markups¹⁶. Since the stochastic discount factor is common across all firms, it affects all firms by the same magnitude¹⁷.

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. In downturns, small firms are less likely to stay in the market than 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 meaningless 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 year 2005 – 2010 for France. The reasons of not using the whole sample

¹⁵I choose $\delta = 0.19$ in the quantitative exercise.

¹⁶This result can be overturned in other models. Opp, Parlour, and Walden (2014) study a theoretical model of price wars among oligopolistic firms as in Rotemberg and Saloner (1986). They extend the analysis in a general equilibrium, and find that if stochastic discount factor is highly pro-cyclical, present value of future cooperation could be higher than current profit from deviation, thus resulting in procyclical markup.

¹⁷With a representative household with access to complete market, the stochastic discount factor is the same for all firms. This assumption can be relaxed to assume heterogenous households with incomplete market, then, depending on the ownership structure of the firm, the stochastic discount factor will be different across firms. This might be an interesting extension future work, but it is not the main focus of this paper.

period are twofold. First, the data tends 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 exit dummy a noisy measure. I choose year 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 (6). Column (1) is 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 10%-level. The reason that it is not significant at 5% is probably due to two reasons. First, censoring of samples in the data makes the measure of exit noisy. Second, firm's exit decision might contain more randomness in normal periods, hence the exit dummy does not fully reflect the firm's optimal decision. Hence, I interact the aggregate output with both exit dummy and recession dummy, equal to one if year is either 2008 or 2009. The results are in column (3). We see that the coefficient of exit dummy during recessions is negative and significant. This result suggests that exit decision during recessions are more endogenous.

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 features heterogenous firms but without entry and exit, and I find that the difference of markup cyclicity 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 good of variety i in the economy. A potential entrant can only enter into the production line of variety i absent of an incumbent ¹⁸.

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,

¹⁸There is a one-to-one mapping between variety and 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.

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 incumbent's idiosyncratic productivity AR(1) process, $\log \tilde{a} \sim N(0, \sigma_a / (1 - \rho_a^2))$ ¹⁹. And its actual idiosyncratic productivity upon entry a' follows the same process as 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 $G^e(\psi_e)$ ²⁰. All entrants are endowed with the same level of initial customer capital b_0 ²¹. Note that even though the entrants draw their productivities from the same distribution as incumbents, they are still smaller because they have less initial product demand b_0 .

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

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

A potential entrant enters the economy if and only if the entry cost ψ_e is less than or equal to a threshold $\psi_e^*(b_0, \tilde{a}, \Lambda)$, which is implicitly given by the following:

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

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

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

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) entrant's distribution Γ^e , and incumbent's distribution Γ such that

¹⁹There are other choices of potential entrants' productivities distribution in the literature. They include, but not limited to, uniform distribution in Hopenhayn and Rogerson (1993), exponential distribution in Lee and Mukoyama (2013), and pareto distribution in Clementi and Palazzo (2013).

²⁰Given the discrete nature of my solution technique, randomness of entry cost helps smooth the response of number of entrants along business cycle, and hence facilitates the search for general equilibrium solution. Clementi, Khan, Palazzo, and Thomas (2014) use the same assumption.

²¹This assumption can be relaxed by allowing the firms to endogenously choose the level of initial customer capital by paying some sunk costs. We can think of the sunk costs as advertising costs of its product.

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 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 firm's problem.
3. Aggregate profit is given by

$$\Phi(\Lambda) = \int [p(b, a, \Lambda)y(b, a, \Lambda) - w(\Lambda)l(b, a, \Lambda)] d\Gamma(b, a). \quad (31)$$

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} ,

$$\Gamma^e(\mathcal{A}) = M \int \int \int \mathbb{1}\{a' \in \mathcal{A}\} \times \mathbb{1}\{\psi_e \leq \psi_e^*(b_0, \tilde{a}, \Lambda)\} \times d\Pi_a(a'|\tilde{a}) \times dh(\tilde{a}) \times dG^e(\psi_e), \quad (32)$$

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

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

$$\begin{aligned} \Gamma'(\mathcal{B} \times \mathcal{A}) = & \int \int \int \mathbb{1}\{b'(b, a, \Lambda) \in \mathcal{B}\} \times \mathbb{1}\{a' \in \mathcal{A}\} \times \mathbb{1}\{\psi \leq \psi^*(b, a, \Lambda)\} \\ & \times d\Gamma(b, a) \times d\Pi_a(a'|a) \times dG(\psi) + \mathbb{1}\{b_0 \in \mathcal{B}\} \times \Gamma^e(\mathcal{A}), \end{aligned} \quad (33)$$

where ψ^* is given by (22).

3.4 Equilibrium Computation

In this section, I sketch the outline of the equilibrium computation. The solution algorithm I employ is closely related to the ones in Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2014), and Clementi and Palazzo (2013). For readers who are interested in the detail of the implementation, please refer to Appendix C.

First, firm's problem can be expressed in terms of the habit-adjusted price index. In particular, incumbent and entrant's value function can be normalized by the habit-adjusted price index as

$\tilde{V} \equiv V/\tilde{p}$, and $\tilde{V}_e \equiv V_e/\tilde{p}$, respectively. Hence, the normalized value functions for an incumbent becomes

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

and the normalized value function for an entrant becomes:

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

The advantage of the transformation above is that the expression of the real wage $w(\Lambda)/\tilde{p}(\Lambda)$ is given by the household's FOC (16), and the real stochastic discount factor is simply

$$\frac{\tilde{p}(\Lambda')}{\tilde{p}(\Lambda)} q(\Lambda, \Lambda') = \beta \left(\frac{X(\Lambda')}{X(\Lambda)} \right)^{-\sigma}, \quad (34)$$

which only depends on the ratio of the habit-adjusted aggregate consumption X .

In the Appendix C, I show that the vector of aggregate state variables Λ actually consists of the aggregate productivity A , and the distribution of firm $\Gamma(b, a)$ over the customer capital b and the idiosyncratic productivity a , denotes as $\Lambda = \{A, \Gamma\}$. In general, the firm distribution Γ is an infinitely-dimensional object, hence it is computational impossible to perfectly keep track of it. Instead, following the strategy as suggested in Krusell and Smith (1998), and its application in Clementi and Palazzo (2013), in which firms face similar entry and exit problems as in mine, I conjecture that the law of motion for habit-adjusted aggregate consumption is given by the following forecast rule:

$$\log X_{t+1} = \beta_{X0} + \beta_{X1} \log X_t + \beta_{X2} \log A_{t+1} + \beta_{X3} \log A_t. \quad (35)$$

As in Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2014), I assume that the inverse of Frisch elasticity of labor supply $\nu = 0$, which implies that real wage is a linear function of habit-adjusted aggregate consumption $w_t/\tilde{p}_t = \omega X_t^\sigma$. This means that I do not need an extra forecast rule for the aggregate labor (or the real wage). Therefore, the state variables become $\{b, a, A, X\}$, and a firm uses the forecast rule (35) when calculating the continuation value in the future. In the solution algorithm, I keep updating the forecast rule coefficient $\beta^X = \{\beta_{X0}, \beta_{X1}, \beta_{X2}, \beta_{X3}\}$, until some accuracy level is reached. The definition of the accuracy statistics and its assessment is described in the Appendix C.

3.5 Empirically Comparable Variables

For the calibration, I need to define several variables that is 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 for moment matching of aggregate output.

From now on, for any variable J_t in the model, I denote J_t^D for 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 data-consistent counterpart for aggregate output is $X_t^D \equiv (\tilde{p}_t X_t) / p_t^{D22}$. With some algebra, it can be written as

$$X_t^D = \underbrace{N_t^{-\frac{1}{\rho-1}}}_{\xi\text{-bias}} \times \underbrace{\left[\frac{\int_{i \in I_t} \left(\mu_{it} l_{it}^{1-\alpha} a_{it}^{-1} b_{it}^{-\theta} \right)^{1-\rho} di}{\int_{i \in I_t} \left(\mu_{it} l_{it}^{1-\alpha} a_{it}^{-1} \right)^{1-\rho} di} \right]^{\frac{1}{1-\rho}}}_{\theta\text{-bias}} \times X_t, \quad (36)$$

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 business cycle. In the real world, a statistical agent adjusts for the new product bias in the CPI calculation, but at a slow pace²³. 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 case²⁴, then the θ -bias can be simplified to b^θ , and growth in aggregate output is:

$$\Delta \log X_t^D = -\frac{1}{\rho-1} \Delta \log N_t - \theta \Delta \log b_t + \Delta \log X_t. \quad (37)$$

In the simulation, 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 term.

²²This can be seen by considering a more general case of 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 as discussed in Benassy (1996). 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.

²³The UK added smartphone into the basket of good in 2011.

²⁴Note that even if all the firms have the same productivity, they might still differ in sizes, due to the fact that entrant's customer capital is different from 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 entrant, which is equal to incumbent's.

4 Calibration

One main objective of the quantitative exercise is to see whether the model can match two key moments in the empirical findings—the magnitude and the heterogeneity of markup cyclicality. To do so, 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 7. To be comparable with the data, I set one period in the model equal to one year. The annual discount factor equals to $\beta = 0.96$. I assume log utility in consumption, $\sigma = 1$. For labor supply, I assume inverse Frisch elasticity $\nu = 0$ as in Hansen (1985) and Rogerson (1988). And as explained above, this simplifies the equilibrium computation²⁵. I set the production elasticity of labor input as $\alpha = 0.7$. I choose 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 $\theta = 1.490$, 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 steady state markup is actually around $\mu = 1.2$, close to what most literature has found.

4.2 Fitted Parameters

For the remaining parameters that I choose to match the empirical moments are listed in Table 8. Since calibration of the full set of parameters in a general equilibrium requires significant amount of time, I split the calibration procedure into two steps. In the first step, I calibrate the parameters govern the cross-sectional moments of firms. They are the parameters regarding the distribution of entry cost, the relative size of entrants, the distribution of operating cost, and idiosyncratic productivities. I calibrate these 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 the entrants have the same average

²⁵For robustness check, I have tried a more realistic Frisch elasticity $\nu = 2$ and laid out the detail of computation procedure in the appendix. My main result is robust to this alternative value.

productivities as the incumbents. I set 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 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} is also autocorrelated across periods. And this value is very close to the estimate in Foster et al. (2008)²⁶.

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.

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 impulse response analysis to analyze variables of interest. I show that my model is able to generate heterogeneity in markup cyclicity between small and large firm, as observed in the data. Moreover, in contrast to a standard business cycle model, my model predicts countercyclical firm size dispersion with only a first moment shock in aggregate productivity. And due to the increased misallocation of inputs caused by the firm size dispersion in recessions, the fluctuations of the measured TFP is amplified. Finally, I compare the baseline model to the one with homogeneous firms to see the difference.

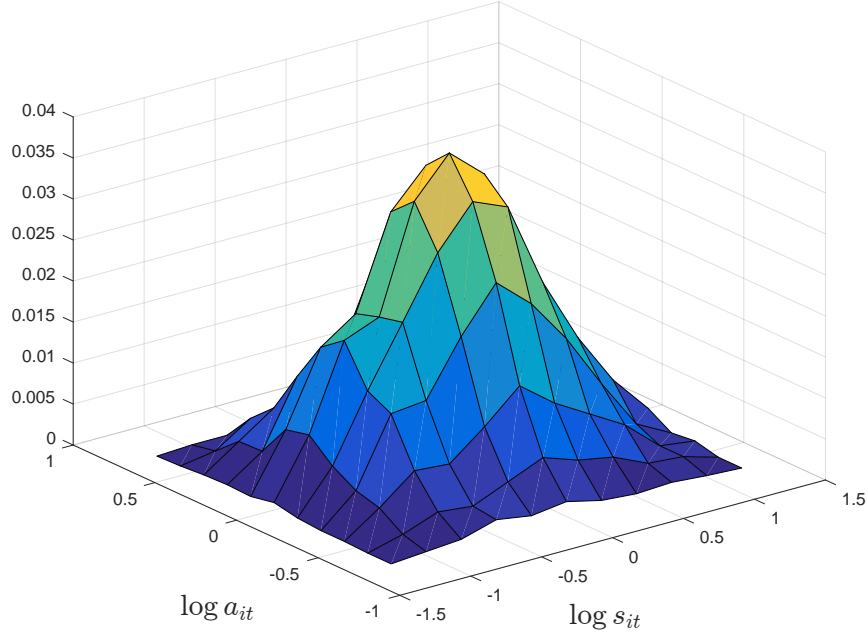
5.1 Steady State Results

Figure 2 displays the stationary distribution of firms over idiosyncratic productivity a_{it} and market share in revenue $s_{it} = \frac{p_{it}y_{it}}{\sum_i p_{it}y_{it}}$ ²⁷. In contrast to Hopenhayn (1992), where there is a one-to-one mapping between idiosyncratic productivity and size (Unless stated 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 is no such adjustment costs in this model, and firms are simply constrained by the level of current

²⁶With firm-level output data, they find that the persistence of physical idiosyncratic productivity to be 0.814

²⁷The market share in the figure is already demeaned.

Figure 2: Stationary Distribution of Firms



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

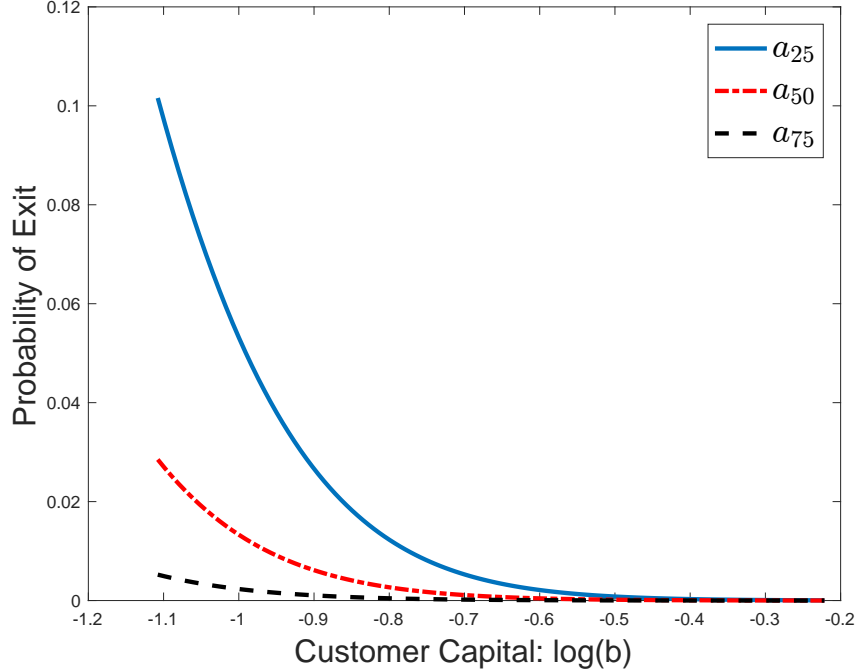
demand (b_{it}). A firm with high productivity but low demand can only grow in size slowly by accumulating its customer capital.

Figure 3 displays the steady-state exit probability of firms. I plot exit probability for firms with 25th percentile, 50th percentile, and 75th 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 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. Also, the increase of exit probability is smooth because the operating cost ψ is drawn from a continuous distribution²⁸.

Lastly, Figure 4 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 its output

²⁸There will be discrete jump if the operating cost is instead a singleton.

Figure 3: Exit Probability of Firms: $1 - G(\psi^*)$



to catch up with optimal size implied by its high a , while high b firms are shrinking to adjust to its optimal size implied by its new low a . The second force is that the productivity process is mean-reverting, and there is positive, though low²⁹, correlation between customer capital and productivity $\rho_{\log a_{it}, \log b_{it}} = 0.35$. Overall, the revenue growth rate is quite small, since the model is calibrated to match the high persistence of market share in the data (0.96).

5.2 Results with Aggregate Fluctuations

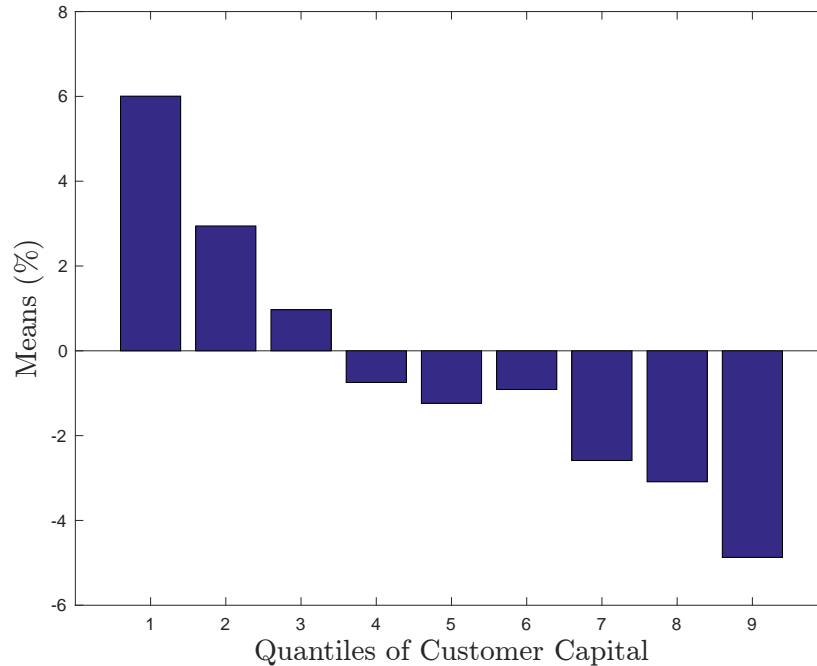
5.2.1 Heterogeneity in Markup Cyclicalilty

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 cyclicalilty, which is the key moment to target in this paper. To compare my model to the data, for each period t in both data and simulation, I first split firms into three categories based on their market share s_{it} : $\mathcal{C} \in \{Small, Middle, Large\}$. Each category \mathcal{C} consists of a tercile of firms³⁰. And then, for each category, I separately run the following first

²⁹It is low because the model is calibrated such that the entrants have the same average productivities as the incumbents, which are relatively larger in size.

³⁰Since the first difference regression involves the variable at time t and $t - 1$, I categorize firms based on their average market share at time t and $t - 1$, $\frac{s_{it} + s_{it-1}}{2}$.

Figure 4: Revenue Growth Rate $\frac{p_{it+1}y_{it+1}}{p_{it}y_{it}}$



difference regression³¹:

$$\Delta \log \mu_{it} = \phi_C \Delta \log X_t^D + \varepsilon_{it}. \quad (38)$$

My results are summarized in Figure 5. 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 cyclicity. For the markup cyclicity of large firms, the model matches the counterpart for Translog production function case. The result is encouraging, since the only moments that I target in the calibration are static moments, but the model successfully captures the heterogenous dynamics of markups.

Second, the model captures one-third of average cyclicity when compared to Cobb-Douglas case, and two-thirds for Translog. This possibly implies that other models of markup cyclicity are still important. For example, one element that the model lacks is sticky price. It will be an important extension to add this feature into the model for future research, since understanding the interactions between the sticky price and the cyclical variations in desired markups. Nevertheless,

³¹Note 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 .

there are possible options within the current framework to match the magnitude. I discuss the alternative functional form specifications from that would help towards the right direction below.

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 2000 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 afterwards. For each period t , I calculate the average value of interested variables and normalized to give the percentage deviation from their pre-shock values.

Figure 6 depicts the heterogeneous responses of firms. The middle panel depicts the markup responses of small and large firms, where small and large are defined as above. 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 its transmission into firm's output dynamics. The bottom panel shows the output responses of small and large firms. Since a small firm raises markup more than a large firm, its output declines relatively more. Hence, the dispersion of firm output increases endogenously³². And note that the output gap between 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 output in the data. However, we can test the dispersion by looking at deflated value-added³³. 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:

$$\Delta \log \left(\frac{p_{it} y_{it}}{p_t^D} \right) = \phi_C^{py} \Delta \log X_t^D + \varepsilon_{it}. \quad (39)$$

where $C \in \{Small, Middle, Large\}$ as above. The results are in Table 10. We can see that large firms tend to have less volatile sales, both in the model and data. In the model, small firms' deflated sales are 13% more volatile than large firms', while the difference is 22% in the data. Hence, the model can explain 60% of heterogeneity in output cyclicity.

In Figure 7, 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

³²The same result holds with market share and labor.

³³Since my model does not incorporate intermediate inputs, hence it is more natural to compare the sales in the model to value-added in the data.

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 Terry (2014) find that 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 dispersion of productivity works against the dispersion of firm size.

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:

$$Z_t \equiv \frac{X_t}{L_t^\alpha}, \quad (40)$$

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:

$$Z_t = A_t \times \underbrace{E(a)_t}_{\text{Compositional Effect}} \times \underbrace{\left[\int \left(\frac{a}{E(a)_t} b^\theta \right)^{\frac{\rho-1}{\rho}} \left(\frac{l(b, a, \Lambda_t)}{L_t} \right)^{\alpha \frac{\rho-1}{\rho}} d\Gamma_t(b, a) \right]^{\frac{\rho}{\rho-1}}}_{\text{Misallocation Effect}}. \quad (41)$$

As seen above, the difference between Z_t and A_t are due to two terms associated with compositional and misallocation effect. First, $E(a)_t$ is the mean of incumbents' idiosyncratic productivity. This variable is time varying due to composition of firms over business cycles. High-productivity firms survive during recessions, and low-productivity firms enter the economy during booms. So $E(a)_t$ 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 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 8. 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 *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, $Z_t^D/E(a)_t$. 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.

Lastly, to show the importance of heterogeneity of firm dynamics in the model, I compare the baseline model to a model that features homogeneous firms without entry and exit. In Figure 9, without heterogeneity, we see that the measured TFP perfectly overlaps with the true aggregate productivity. Hence, all the amplified dynamics of the measured TFP comes from firm heterogeneity.

6 Alternative Functional Form Specifications

As shown in the quantitative analysis, the model successfully matches the heterogeneity of markup cyclicity, but only a half of the magnitude. In this section, I discuss two potential ways to improve on this dimension.

6.1 GHH Preference

The volatility of stochastic discount factor $q_{t,t+1}$ determines the magnitude of markup cyclicity. One way to increase its volatility is to increase the value of relative risk aversion from the value used in the baseline calibration, $\sigma = 1$. However, a high σ implies a strong wealth effect on labor supply. In fact, the aggregate labor becomes counter-cyclical in the simulation when $\sigma = 2$, an upper bound suggested by Chetty (2006). Therefore, to reduce the wealth effect on labor supply, one option would be to use the non-separable utility function introduced by Greenwood, Hurcowitz, and Huffman (1988)³⁴ as below:

$$U(X_t, L_t) = \frac{1}{1 - \sigma} \left(X_t - \omega \frac{L_t^{1+\nu}}{1 + \nu} \right)^{1-\sigma}. \quad (42)$$

³⁴GHH preference is one special case of a more general preference introduced by Jaimovich and Rebelo (2009). The other special case is the one proposed by King, Plosser, and Rebelo (1988).

Under GHH, the household's intratemporal and intertemporal conditions become the following:

$$\frac{w(\Lambda)}{\tilde{p}(\Lambda)} = \omega L^\nu(\Lambda), \quad (43)$$

$$q(\Lambda, \Lambda') = \beta \left(\frac{X(\Lambda') - \omega \frac{L^{1+\nu}(\Lambda')}{1+\nu}}{X(\Lambda) - \omega \frac{L^{1+\nu}(\Lambda)}{1+\nu}} \right)^{-\sigma} \frac{\tilde{p}(\Lambda)}{\tilde{p}(\Lambda')}. \quad (44)$$

Hence, wealth effect on labor supply is entirely eliminated. In contrast to the baseline case, now we need two forecast rules for habit-adjusted aggregate consumption X_t and aggregate labor L_t respectively:

$$\log X_{t+1} = \beta_{X0} + \beta_{X1} \log X_t + \beta_{X2} \log A_{t+1} + \beta_{X3} \log A_t, \quad (45)$$

$$\log L_t = \beta_{L0} + \beta_{L1} \log X_t + \beta_{L2} \log A_t. \quad (46)$$

This procedure takes significantly more time for computation, and I find that the increase in markup cyclicity is modest at this stage³⁵.

6.2 Quasi-Difference Deep Habits

In the baseline model, the household's aggregate consumption depends on the quasi-product of product-specific consumption and habit stock. Instead, we can consider one that depends on the quasi-difference of consumption and habit stock, as introduced as the baseline case in Ravn et al. (2006). Specifically, the household has the following preference over consumption:

$$X_t = \left[\int_{i \in I_t} (c_{it} - \theta b_{it})^{\frac{\rho-1}{\rho}} di \right]^{\frac{\rho}{\rho-1}}. \quad (47)$$

Then household's demand for consumption c_{it} is given by

$$c_{it} = \left(\frac{p_{it}}{p_t} \right)^{-\rho} X_t + \theta b_{it}. \quad (48)$$

As before, the past habit stock b_{it} affects today's demand. So firms still take into account that current price decision affects future demand when making decisions. If the present value of future profit is high, firms have incentive to invest in customer capital, and vice versa. And the cyclicity of stochastic discount factor determines the present value. This is called *intertemporal effect*. However,

³⁵For this model, I assume that the constant relative risk aversion $\sigma = 2$, and the inverse of Frisch elasticity of labor supply $\nu = 2$, as suggested by Chetty, Guren, Manoli, and Weber (2011). The gain in average markup cyclicity is not large, probably due to strong cyclicity of labor compared to consumption, which reduces the volatility of stochastic discount factors.

now the habit stock b_{it} enters as an additive term into the equation. This generates an extra effect called *price-elasticity effect*. The first term in the equation above is the usual price-elastic term with price elasticity ρ , and the second term is price inelastic. The total price elasticity is the weighted average between ρ and 0. An increase in X_t in boom induces the price elasticity to increase, and hence the markup goes down.

However, before using the quasi-difference form of deep habits to increase the magnitude of markup cyclical, one needs to solve one issue. Due to the perfectly inelastic term, a monopolist actually has an incentive to set an infinite price. Schmitt-Grohe and Uribe (2007) show that any finite deviation from the symmetric equilibrium price is suboptimal, but this does not need to hold in general. One simple but not desirable solution is to enforce a price cap for firms. Finding a good remedy for this issue is beyond the scope of this paper and is left for future research.

7 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 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 endogenous firm entry and exit. In the model, a negative aggregate productivity shock increases small firms' exit probabilities more than large firms', and thus small firms increase 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 the business cycles than large firms, their pricing, and thus output responses are more volatile. Hence, during recessions, small firms output decline 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 the 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 contradictory evidence of markup cyclical while using static inputs (material) and dynamics inputs (labor). The second one is that the incorporation would allow for a decomposition analysis of labor wedge into product

market wedge (markup) and labor market wedge, but the model in this paper attributes all 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 to obtain a better monetary policy. Adding these two features will be important extensions for my future research agenda.

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 and domestic prices: how large are strategic complementarities? Tech. rep., National Bureau of Economic Research, 2016.
- [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] BENASSY, J.-P. Taste for variety and optimum production patterns in monopolistic competition. *Economics Letters* 52, 1 (1996), 41–47.
- [7] BILBIIE, F., GHIRONI, F., AND MELITZ, M. J. Endogenous entry, product variety, and business cycles. *Journal of Political Economy* 120, 2 (2012), 304–345.
- [8] BILS, M. The cyclical behavior of marginal cost and price. *The American Economic Review* (1987), 838–855.
- [9] 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.
- [10] BILS, M., KLENOW, P. J., AND MALIN, B. A. Resurrecting the role of the product market wedge in recessions. Tech. rep., National Bureau of Economic Research, 2016.
- [11] BLANCHARD, O., AND GALÍ, J. Real wage rigidities and the new keynesian model. *Journal of Money, credit and Banking* 39, s1 (2007), 35–65.
- [12] BLOOM, N. The impact of uncertainty shocks. *Econometrica* 77, 3 (2009), 623–685.

- [13] 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).
- [14] CHANG, Y., KIM, S.-B., KWON, K., AND ROGERSON, R. Individual and aggregate labor supply in a heterogeneous agent economy with intensive and extensive margins. *Unpublished Manuscript* (2014).
- [15] CHARI, V. V., KEHOE, P. J., AND MCGRATTAN, E. R. Business cycle accounting. *Econometrica* 75, 3 (2007), 781–836.
- [16] CHETTY, R. A new method of estimating risk aversion. *The American Economic Review* 96, 5 (2006), 1821–1834.
- [17] CHETTY, R., GUREN, A., MANOLI, D., AND WEBER, A. Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *The American Economic Review* 101, 3 (2011), 471–475.
- [18] 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.
- [19] CLARIDA, R., GALI, J., AND GERTLER, M. The science of monetary policy: A new keynesian perspective. *Journal of Economic Literature* 37 (1999), 1661–1707.
- [20] 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).
- [21] CLEMENTI, G. L., AND PALAZZO, B. Entry, exit, firm dynamics, and aggregate fluctuations. Tech. rep., National Bureau of Economic Research, 2013.
- [22] DE LOECKER, J., AND WARZYNSKI, F. Markups and firm-level export status. *The American Economic Review* 102, 6 (2012), 2437–2471.
- [23] 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.
- [24] 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.

- [25] EDMOND, C., AND VELDKAMP, L. Income dispersion and counter-cyclical markups. *Journal of Monetary Economics* 56, 6 (2009), 791–804.
- [26] FOSTER, L., HALTIWANGER, J., AND SYVERSON, C. The slow growth of new plants: Learning about demand? *Economica* 83, 329 (2016), 91–129.
- [27] GILCHRIST, S., SCHOENLE, R., SIM, J., AND ZAKRAJSEK, E. Inflation dynamics during the financial crisis. *FEDS Working Paper* (2015).
- [28] GOPINATH, G., KALEMLI-OZCAN, S., KARABARBOUNIS, L., AND VILLEGAS-SANCHEZ, C. Capital allocation and productivity in south europe. Tech. rep., National Bureau of Economic Research, 2015.
- [29] GREENWOOD, J., HERCOWITZ, Z., AND HUFFMAN, G. W. Investment, capacity utilization, and the real business cycle. *The American Economic Review* (1988), 402–417.
- [30] HALL, R. E. Market structure and macroeconomic fluctuations. *Brookings papers on economic activity*, 2 (1986), 285–322.
- [31] HALL, R. E. What the cyclical response of advertising reveals about markups and other macroeconomic wedges. *Hoover Institution, Stanford University* (2014).
- [32] HANSEN, G. D. Indivisible labor and the business cycle. *Journal of monetary Economics* 16, 3 (1985), 309–327.
- [33] HOPENHAYN, H., AND ROGERSON, R. Job turnover and policy evaluation: A general equilibrium analysis. *Journal of political Economy* (1993), 915–938.
- [34] HOPENHAYN, H. A. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society* (1992), 1127–1150.
- [35] HSIEH, C.-T., AND KLENOW, P. J. Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics* 124, 4 (2009), 1403–1448.
- [36] JAIMOVICH, N., AND FLOETOTTO, M. Firm dynamics, markup variations, and the business cycle. *Journal of Monetary Economics* 55, 7 (2008), 1238–1252.
- [37] JAIMOVICH, N., AND REBELO, S. Can news about the future drive the business cycle? *The American economic review* 99, 4 (2009), 1097–1118.

- [38] 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).
- [39] KING, R. G., PLOSSER, C. I., AND REBELO, S. T. Production, growth and business cycles: I. the basic neoclassical model. *Journal of monetary Economics* 21, 2-3 (1988), 195–232.
- [40] 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.
- [41] KRUSELL, P., AND SMITH, JR, A. A. Income and wealth heterogeneity in the macroeconomy. *Journal of political Economy* 106, 5 (1998), 867–896.
- [42] KRYVTSOV, O., AND MIDRIGAN, V. Inventories, markups, and real rigidities in menu cost models. *The Review of Economic Studies* (2012), rds028.
- [43] LEAHY, J. A survey of new keynesian theories of aggregate supply and their relation to industrial organization. *Journal of Money, Credit and Banking* 43, s1 (2011), 87–110.
- [44] LEE, Y., AND MUKOYAMA, T. Entry, exit, and plant-level dynamics over the business cycle. *Working Paper* (2013).
- [45] LEVINSOHN, J., AND PETRIN, A. Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies* 70, 2 (2003), 317–341.
- [46] MOREIRA, S. Firm dynamics, persistent effects of entry conditions, and business cycles. *Working Paper* (2015).
- [47] NAKAMURA, E., AND STEINSSON, J. Price setting in forward-looking customer markets. *Journal of Monetary Economics* 58, 3 (2011), 220–233.
- [48] NEKARDA, C. J., AND RAMEY, V. A. The cyclical behavior of the price-cost markup. Tech. rep., National Bureau of Economic Research, 2013.
- [49] OLLEY, G. S., AND PAKES, A. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica* 64, 6 (November 1996), 1263–97.
- [50] OPP, M. M., PARLOUR, C. A., AND WALDEN, J. Markup cycles, dynamic misallocation, and amplification. *Journal of Economic Theory* 154 (2014), 126–161.

- [51] PACIELLO, L., POZZI, A., AND TRACHTER, N. Price dynamics with customer markets. *Working Paper* (2014).
- [52] PETERS, M. Heterogeneous mark-ups, growth and endogenous misallocation. *Working Paper* (2013).
- [53] RAVN, M., SCHMITT-GROHÉ, S., AND URIBE, M. Deep habits. *The Review of Economic Studies* 73, 1 (2006), 195–218.
- [54] RESTUCCIA, D., AND ROGERSON, R. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics* 11, 4 (2008), 707–720.
- [55] ROGERSON, R. Indivisible labor, lotteries and equilibrium. *Journal of Monetary Economics* 21, 1 (1988), 3–16.
- [56] ROTEMBERG, J. J., AND SALONER, G. A supergame-theoretic model of price wars during booms. *The American Economic Review* 76, 3 (1986), 390–407.
- [57] ROTEMBERG, J. J., AND WOODFORD, M. Markups and the business cycle. In *NBER Macroeconomics Annual 1991, Volume 6*. MIT Press, 1991, pp. 63–140.
- [58] ROTEMBERG, J. J., AND WOODFORD, M. The cyclical behavior of prices and costs. *Handbook of macroeconomics* 1 (1999), 1051–1135.
- [59] SCHMITT-GROHÉ, S., AND URIBE, M. Why the monopolist does not deviate from the symmetric equilibrium in a model with good-specific subsistence points. *Note, Duke University* (2007).
- [60] 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.
- [61] 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.
- [62] STEINSSON, J. Optimal monetary policy in an economy with inflation persistence. *Journal of Monetary Economics* 50, 7 (2003), 1425–1456.
- [63] STROEBEL, J., AND VAVRA, J. House prices, local demand, and retail prices. Tech. rep., National Bureau of Economic Research, 2014.

- [64] TAUCHEN, G. Finite state markov-chain approximations to univariate and vector autoregressions. *Economics letters* 20, 2 (1986), 177–181.
- [65] 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.

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:

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

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 general 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 Akerberg, Caves, and Frazer’s (2015; hereafter ACF) assumption and approach for estimation, and my procedure is the following:

1. Given the timing assumption in ACF³⁶, firm’s demand function for material is

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

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

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

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.

³⁶ACF 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:

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

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 input decision before realizing the “unanticipated” shock ϵ_{it} .

B FOCs for Incumbent's Problem

B.1 First Order Conditions for Incumbents

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 firm's idiosyncratic productivity a from the state variables. Then, incumbent's transformed value function becomes the following:

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

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:

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

Note that normalization of the exit value to be 0 helps simplify the last FOC with Leibniz integral rule by a lot.

Combining first three FOCs yields the following

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

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

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

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

B.2 Symmetric Equilibrium

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

$$\begin{aligned} X &= N^{\frac{\rho}{\rho-1}} y b^\theta \\ \tilde{p} &= N^{-\frac{1}{\rho-1}} \frac{p}{b^\theta}, \end{aligned}$$

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

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

Baseline Case: $\sigma = 1$

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

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

And firm's intertemporal FOC becomes:

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

By iterating forward, one can obtain:

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

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, aggregate productivity is at a constant $A_t = A_{ss}$, $\forall t > 0$. Since there is neither firm entry nor exit, the number of incumbents stays constant and I normalize it to $N_t = 1$, $\forall t > 0$. Then a firm charges a steady state markup:

$$\mu_{ss} = \frac{\rho_{ss}}{\rho_{ss} - 1}, \tag{49}$$

where

$$\rho_{ss} = \rho \frac{1 - \beta(1 - \delta)}{1 - \beta(1 - \delta + \delta\theta(\rho - 1))} > \rho, \quad (50)$$

which means that firms' 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 $X_{ss} = y_{ss}^{1+\theta}$, and habit-adjusted price index is $\tilde{p}_{ss} = p_{ss}y_{ss}^{-\theta}$. Hence steady state output y_{ss} is

$$y_{ss} = \left[\frac{\alpha}{\mu_{ss}\omega} A_{ss}^{(1+\nu)/\alpha} \right]^{\frac{1}{\frac{1+\nu}{\alpha} + (\sigma-1)(1+\theta)}}. \quad (51)$$

And steady state aggregate labor is

$$L_{ss} = \left[\frac{\alpha}{\mu_{ss}\omega} A_{ss}^{-(\sigma-1)(1+\theta)} \right]^{\frac{1}{1+\nu+\alpha(\sigma-1)(1+\theta)}}. \quad (52)$$

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 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 Firm's Problem

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

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

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

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

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

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

In the calibration, I choose the parameters such that the household has 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

$$\begin{aligned} \frac{w(\Lambda)}{\tilde{p}(\Lambda)} &= \omega X(\Lambda) \\ q(\Lambda, \Lambda') &= \beta \frac{\tilde{p}(\Lambda) X(\Lambda)}{\tilde{p}(\Lambda') X(\Lambda')}. \end{aligned}$$

First, the real wage is now a function of only habit-adjusted aggregate consumption $X(\Lambda)$. Second, in the firm's problem, we see that the normalized stochastic discount factor simplifies to

$\frac{\bar{p}(\Lambda')}{\bar{p}(\Lambda)}q(\Lambda, \Lambda') = \beta \frac{X(\Lambda)}{X(\Lambda')}$. Hence, the only aggregate variable matters to firm's problem is the habit-adjusted aggregate consumption $X(\Lambda)$, and it can be expressed as the following in log

$$\log X(\Lambda) = \log A + \log \Omega(\Lambda),$$

where $\Omega(\Lambda) \equiv \left[\int (al^\alpha(b, a, \Lambda)b^\theta)^{\frac{\rho-1}{\rho}} d\Gamma(b, a) \right]^{\frac{\rho}{\rho-1}}$, 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 equilibrium. 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 X^* : (1) Given a guess of X^* , solve for firm's optimization problem³⁷; (2) Compute firm's policy function of out $y(b, a, X^*)$, and calculate the habit-adjusted aggregate consumption from $X^* = \left[\int (y(b, a, X^*)b^\theta)^{\frac{\rho-1}{\rho}} d\Gamma(b, a, X^*) \right]^{\frac{\rho}{\rho-1}}$; (3) Repeat until the market-clearing X^* is found.

Krusell-Smith Algorithm

To describe the dynamics of the habit-adjusted aggregate consumption $X(\Lambda)$, 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

$$\log \hat{X}' = \beta_{X0} + \beta_{X1} \log X + \beta_{X2} \log A' + \beta_{X3} \log A. \quad (53)$$

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

³⁷Note that the stochastic discount factor is a constant in the steady state $q_t = \beta \left(\frac{X^*}{X^*} \right)^{-\sigma} = \beta$.

Incumbent's Problem

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

$$\begin{aligned} \tilde{V}(b, a, A, X) = \max_{b'} & \left\{ b^{\theta(\rho-1)/\rho} X^{1/\rho} \left(\frac{b' - (1-\delta)b}{\delta} \right)^{\frac{\rho-1}{\rho}} - \omega X^\sigma (Aa)^{-1/\alpha} \left(\frac{b' - (1-\delta)b}{\delta} \right)^{\frac{1}{\alpha}} \right. \\ & - \Pr[\psi \leq \psi^*(b, a, A, X)] \mathbb{E}[\psi | \psi \leq \psi^*(b, a, A, X)] \\ & \left. + \beta \Pr[\psi \leq \psi^*(b, a, A, X)] \mathbb{E}_{a,A} \left[\left(\frac{\hat{X}'}{X} \right)^{-\sigma} \tilde{V}(b', a', A', \hat{X}') \right] \right\}. \end{aligned}$$

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

$$\Pr(\psi \leq \psi^*) = \Phi \left(\frac{\log \psi^* - m_\psi}{\sigma_\psi} \right), \quad (54)$$

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

$$\mathbb{E}(\psi | \psi \leq \psi^*) = \exp \left(m_\psi + \frac{\sigma_\psi^2}{2} \right) \Phi \left(\frac{\log \psi^* - m_\psi - \sigma_\psi^2}{\sigma_\psi} \right) / \Phi \left(\frac{\log \psi^* - m_\psi}{\sigma_\psi} \right). \quad (55)$$

Entrant's Problem

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

$$\tilde{V}_e(\psi_e, b_0, \tilde{a}, A, X) = \max \left\{ 0, -\psi_e + \beta \mathbb{E}_{\tilde{a}, A} \left[\left(\frac{\hat{X}'}{X} \right)^{-\sigma} \tilde{V}(b_0, a', A', \hat{X}') \right] \right\}, \quad (56)$$

and the entrant enters if and only if

$$\psi_e \leq \beta \mathbb{E}_{\tilde{a}, A} \left[\left(\frac{\hat{X}'}{X} \right)^{-\sigma} \tilde{V}(b_0, a', A', \hat{X}') \right]. \quad (57)$$

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 $(\beta_{X0}^{(1)}, \beta_{X1}^{(1)}, \beta_{X2}^{(1)}, \beta_{X3}^{(1)})$ to solve the model, and perform the following iterations $m = 1, 2, 3, \dots$ of the solution algorithm

1. Given the forecast rule $(\beta_{X0}^{(m)}, \beta_{X1}^{(m)}, \beta_{X2}^{(m)}, \beta_{X3}^{(m)})$ 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 $(\beta_{X0}^{(m+1)}, \beta_{X1}^{(m+1)}, \beta_{X2}^{(m+1)}, \beta_{X3}^{(m+1)})$ accordingly.
4. Repeat from step (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 $n_X = 7$ grid points for the habit-adjusted aggregate consumption X . And the incumbent uses the forecast rule $\beta_X^{(m)} = (\beta_{X0}^{(m)}, \beta_{X1}^{(m)}, \beta_{X2}^{(m)}, \beta_{X3}^{(m)})$ 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{X}' does not fall on the given grid points in general. Hence, I compute the continuation value $E\tilde{V}'$ 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 $b'(b, a, A, x)$, the next period's distribution is determined by

$$\begin{aligned}\Gamma_{t+1}(b, a) &= \sum_j^{n_a} \sum_i^{n_b} \mathbb{1} \{b'(b_i, a_j, A_t, X_t) = b\} \times \Pi^a(a|a_j) \\ &\quad \times (1 - \gamma) \times \Pr(\psi \leq \psi^*(b_i, a_j, A_t, X_t)) \times \Gamma_t(b_i, a_j) \\ &\quad + \sum_i^{n_{\tilde{a}}} \mathbb{1} \times \Pr(\psi_e \leq \psi_e^*(b_0, \tilde{a}_i, A_t, X_t)) \times (1 - \gamma) \times \Pi_a(a|\tilde{a}_i) \times h(\tilde{a}_i) \times M_t,\end{aligned}$$

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 X_t . In particular, given a predetermined distribution of firm $\Gamma_t(b, a)$, for any guess of current habit-adjusted aggregate consumption \dot{X} , an incumbent with customer capital b and idiosyncratic productivity a uses the forecast rule to calculate the continuation value, and solves the following problem

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

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

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

The market clearing condition for output is satisfied if and only if $\ddot{X} = \dot{X}$. In practice, I use 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{X} every time. Hence, I instead do a linear interpolation of the policy function obtained in

step (1) along the grid for X to approximate the solution. The result turns out to be quite close to the one obtained with 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 $\{A_t, X_t\}_{t=1, \dots, T}$ ³⁸. 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 $\{A_t, X_t\}_{t=T_0, \dots, T}$, I run the following OLS regression:

$$\log X_{t+1} = \beta_{X0} + \beta_{X1} \log X_t + \beta_{X2} \log A_{t+1} + \beta_{X3} \log A_t.$$

Denote the set of coefficients obtained from the regression above as $\hat{\beta}_X = (\hat{\beta}_{X0}, \hat{\beta}_{X1}, \hat{\beta}_{X2}, \hat{\beta}_{X3})$. I update the forecast rule as the weighted sum of $\hat{\beta}_X$, and the forecast rule from the previous iteration $\beta_X^{(m)}$:

$$\beta_X^{(m+1)} = w_{KS} \hat{\beta}_X + (1 - w_{KS}) \beta_X^{(m)},$$

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

C.2.4 Step 4: Convergence Criteria

There are several convergence metrics widely used 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 statistics is the maximum difference in log between the X_t^{KS} generated by repeatedly applying the forecast rule β_X only, and the X_t generated from the equilibrium search within each period in the simulation. The convergence criteria is that the Den Haan statistics is less than equal to 10^{-4} . The detail of the maximum Den Haan statistics is described in the subsection below.

C.3 Internal Accuracy Statistics

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

$$\log \hat{X}' = \beta_{X0} + \beta_{X1} \log X + \beta_{X2} \log A' + \beta_{X3} \log A.$$

³⁸Note that x_t is obtained from the equilibrium search within each period, not from the forecast rule.

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 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 statistics, and might hide some large errors. In particular, it will be problematic if a 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 be accumulated over the time. Denote X_t as the market-clearing consumption within each period. And denote X_t^{DH} as the outcome of the iteration of the forecast rule β_X ,

$$\log X_{t+1}^{DH} = \beta_{X0} + \beta_{X1} \log X_t^{DH} + \beta_{X2} \log A_{t+1} + \beta_{x3} \log A_t,$$

where the initial point of the iteration is equal to the actual habit-adjusted aggregate consumption $X_1^{DH} = X_1$. Then the maximum Den Haan statistics for a given forecast rule β_X is defined as

$$DH(\beta_X) \equiv \max_t |\log X_t - \log X_t^{DH}|.$$

Hence, we can interpret the Den Haan statistics as the maximum percentage error of the forecasted variable. In my model, the Den Haan statistics over 500 periods is 1%, which means that the error of the firm's prediction of X_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 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 $T_{\text{shock}} = 0$.

C.5 A More Realistic Frisch Elasticity of Labor Supply

In my benchmark model, I have assumed that the inverse of Frisch elasticity of labor supply to be zero, $\nu = 0$. However, this choice of parameter is based on complexity of computation, not economic findings. Chang et. al. (2011) suggest a Frisch elasticity of 0.75 in a representative model, Chang et. al. (2014) suggest a value slightly larger than unity, Chetty et al. (2011) suggest 2 that reconciles the micro and macro elasticity of labor supply. In this extension exercise, I set $\nu = 2$.

Steady State Computation

The computation for the steady state is similar to the one when $\nu = 0$, except that the real wage now becomes

$$\frac{w}{\tilde{p}} = \omega L^\nu X^\sigma,$$

which depends on both aggregate labor supply l , and habit-adjusted aggregate consumption X . Hence, I need to find roots for both X^* and L^* in the following way: (1) Given a guess of (X^*, L^*) , solve for firm's optimization problem; (2) Compute firm's policy function for output $y(b, a, X^*, L^*)$, and labor demand $l(b, a, X^*, L^*)$, and calculate $X^* = \left[\int (y(b, a, X^*, L^*) b^\theta)^{\frac{\rho-1}{\rho}} d\Gamma(b, a, X^*, L^*) \right]^{\frac{\rho}{\rho-1}}$, and $L^* = \int l(b, a, X^*, L^*) d\Gamma(b, a, X^*, L^*)$; (3) Repeat until the market-clearing X^* and L^* are found.

Augmented Krusell-Smith Algorithm

To compute the numerical solution for the general equilibrium model, the main challenge now is that, for a finite Frisch elasticity, we need to consider the additional labor supplied in each period. Aggregate labor supplied $L(\Lambda)$ depends on the aggregate state variable Λ , and this determines the real wage jointly with the habit-adjusted aggregate consumption $X(\Lambda)$. Firms take the real wage as given, and solve for the optimization problem. The solution of the optimization problem leads to the policy function for labor demand, $L(b, a, L(\Lambda), X(\Lambda))$. At the equilib-

rium, the labor supplied $L(\Lambda)$ must be equal to the labor demand integrated over all incumbents $L(\Lambda) = \int L(b, a, X(\Lambda), L(\Lambda))d\Gamma(b, a, X(\Lambda), L(\Lambda))$.

From the equilibrium condition above, we see that the aggregate labor only depends on the current aggregate state Λ , hence I conjecture the following log-linear function for aggregate labor

$$\log L_t = \beta_{L0} + \beta_{L1} \log X_t + \beta_{L2} \log A_t. \quad (58)$$

Hence a firm uses β_L to compute current profit, and β_X to calculate the continuation value for its optimization problem. In the simulation, I need to search for both X_t^* and L_t^* that clear the good and labor markets for each period t .

C.6 GHH Preference

Under GHH preference, the equilibrium computation is quite similar to the finite Frisch elasticity case ($\nu > 0$). One only needs to change the real wage and the stochastic discount factor into the following

$$\frac{w(\Lambda)}{\tilde{p}(\Lambda)} = \omega L^\nu(\Lambda), \quad (59)$$

$$q(\Lambda, \Lambda') = \beta \left(\frac{X(\Lambda') - \omega \frac{L^{1+\nu}(\Lambda')}{1+\nu}}{X(\Lambda) - \omega \frac{L^{1+\nu}(\Lambda)}{1+\nu}} \right)^{-\sigma} \frac{\tilde{p}(\Lambda)}{\tilde{p}(\Lambda')}. \quad (60)$$

And the rest of the implementation is exactly the same as the finite Frisch elasticity case.

D Tables

Table 1: Heterogeneity in Markup Cyclicity ϕ : Cobb-Douglas

Dep. Variable	(1)	(2)	(3)	(4)	(5)
$\log \mu_{it}$	All	Large	Small	All	All
$\log Y_t$	-1.201*** (0.186)	-0.868*** (0.240)	-1.259*** (0.178)	-1.242*** (0.179)	-1.241*** (0.181)
$\log Y_t \times \text{Large}_i$				0.286** (0.0817)	
$\log Y_t \times s_{it}$					3.612*** (0.723)
N	118757	16008	102749	118757	118757
adj. R^2	0.965	0.972	0.964	0.965	0.965

Standard errors in parentheses

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

Note: $\log Y_t$ is log aggregate output quadratically detrended. Large_i is an indicator for a firm with more than 1% market share within a 4-digit industry. s_{it} stands for market share in a 4-digit industry. The regression includes firm fixed effect, and control variables: market share, number of employees, HHI, and age. Standard errors are clustered at time level.

Table 2: Markup Cyclicalilty (Material Costs): Cobb-Douglas

Dep. Variable	(1)	(2)	(3)	(4)	(5)
$\Delta \log \mu_{it}$	All	Large	Small	All	All
$\Delta \log Y_t$	-1.198*** (0.257)	-0.983** (0.303)	-1.238*** (0.251)	-1.234*** (0.252)	-1.224*** (0.259)
$\Delta \log Y_t \times \text{Large}_i$				0.234** (0.0663)	
$\Delta \log Y_t \times s_{it}$					2.263*** (0.362)
N	80703	12378	68325	80703	80703
adj. R^2	0.026	0.024	0.027	0.026	0.026

Standard errors in parentheses

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

Note: $\log Y_t$ is log aggregate output. Large_i is an indicator for a firm with more than 1% market share within a 4-digit industry. s_{it} stands for market share in a 4-digit industry. The regression includes control variables: market share, number of employees, HHI, and age. Standard errors are clustered at time level.

Table 3: Heterogeneity in Markup Cyclicalities ϕ : Translog

Dep. Variable	(1)	(2)	(3)	(4)	(5)
$\log \mu_{it}$	All	Large	Small	All	All
$\log Y_t$	-0.712*** (0.162)	-0.326 (0.188)	-0.828*** (0.172)	-0.842*** (0.171)	-0.822*** (0.173)
$\log Y_t \times \text{Large}_i$				0.546** (0.190)	
$\log Y_t \times s_{it}$					6.554** (1.837)
N	52999	11932	41067	52999	52999
adj. R^2	0.902	0.895	0.903	0.902	0.902

Standard errors in parentheses

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

Note: $\log Y_t$ is log aggregate output quadratically detrended. Large_i is an indicator for a firm with more than 1% market share within a 4-digit industry. s_{it} stands for market share in a 4-digit industry. The regression includes firm fixed effect, and control variables: market share, number of employees, HHI, and age. Standard errors are clustered at time level.

Table 4: Heterogeneity in Markup Cyclicalit ϕ : Cobb-Douglas

Dep. Variable	(1)	(2)	(3)	(4)
$\Delta \log \mu_{it}$	France	Germany	Italy	Spain
$\Delta \log Y_t$	-1.234*** (0.252)	-1.400*** (0.154)	-2.121*** (0.172)	-1.311 (0.706)
$\Delta \log Y_t \times \text{Large}_i$	0.234*** (0.0663)	0.457*** (0.0365)	0.513*** (0.129)	0.757 (0.472)
N	80703	37054	197502	238712
adj. R^2	0.026	0.059	0.076	0.017

Standard errors in parentheses

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

Note: $\log Y_t$ is log aggregate output. Large_i is an indicator for a firm with more than 1% market share within a 4-digit industry. The regression includes control variables: market share, number of employees, HHI, and age. Standard errors are clustered at time level.

Table 5: Age and Financial Variables

Dep. Variable	(1)	(2)	(3)	(4)	(5)
$\log \mu_{it}$					
$\log Y_t$	-1.010*** (0.128)	-1.176*** (0.170)	-1.012*** (0.139)	-1.045*** (0.129)	-1.043*** (0.151)
$\log Y_t \times s_{it}$	3.769*** (0.656)	3.599*** (0.931)	3.617*** (0.743)	3.550*** (0.786)	3.645*** (0.738)
$\log Y_t \times \text{Age}_{it}$	-0.0655 (0.0514)				
$\log Y_t \times \text{CASH}_{it}$		-1.346 (3.548)			
$\log Y_t \times \text{CURR}_{it}$			-0.135 (0.0952)		
$\log Y_t \times \text{LIQD}_{it}$				-0.158 (0.129)	
$\log Y_t \times \text{SOLV}_{it}$					-0.508 (0.431)
N	118757	116431	116458	116406	116205
adj. R^2	0.965	0.965	0.965	0.965	0.965

Standard errors in parentheses

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

Note: $\log Y_t$ is log aggregate output quadratically detrended. s_{it} stands for market share in a 4-digit industry. Age_{it} stands for firm's age. CASH_{it} stands for cash ratio, CURR_{it} for current ratio, LIQD_{it} for liquidity ratio, and SOLV_{it} for solvency ratio. The regression includes control variables: market share, number of employees, HHI, and age, and corresponding financial variables. Standard errors are clustered at time level.

Table 6: Heterogeneity in Markup Cyclicity: Exit Indicator

Dep. Variable	(1)	(2)	(3)
$\log \mu_{it}$			
$\log Y_t$	-1.356*** (0.191)	-1.338*** (0.196)	-1.339*** (0.197)
$\log Y_t \times s_{it}$	3.902*** (0.327)	3.878*** (0.328)	3.878*** (0.331)
$\log Y_t \times \text{EXIT}_{it}$		-0.271 (0.128)	0.0244 (0.266)
$\log Y_t \times \text{EXIT}_{it} \times \text{REC}_t$			-0.435* (0.187)
N	54111	54111	54111
adj. R^2	0.973	0.973	0.973

Standard errors in parentheses

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

Note: $\log Y_t$ is log aggregate output quadratically detrended. s_{it} stands for market share in a 4-digit industry. Age_{it} stands dummy for firm exit, which is equal to one if firms exits in the next period. The regression includes control variables: market share, number of employees, HHI, and age, EXIT. Standard errors are clustered at time level.

Table 7: 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	1.490
Depreciation Rate of Customer Capital	δ	0.19

Table 8: Fitted Parameters

Parameter	Symbol	Value
Entry Cost	$\bar{\psi}_e$	0.12
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

Table 9: Targeted Moments in Data

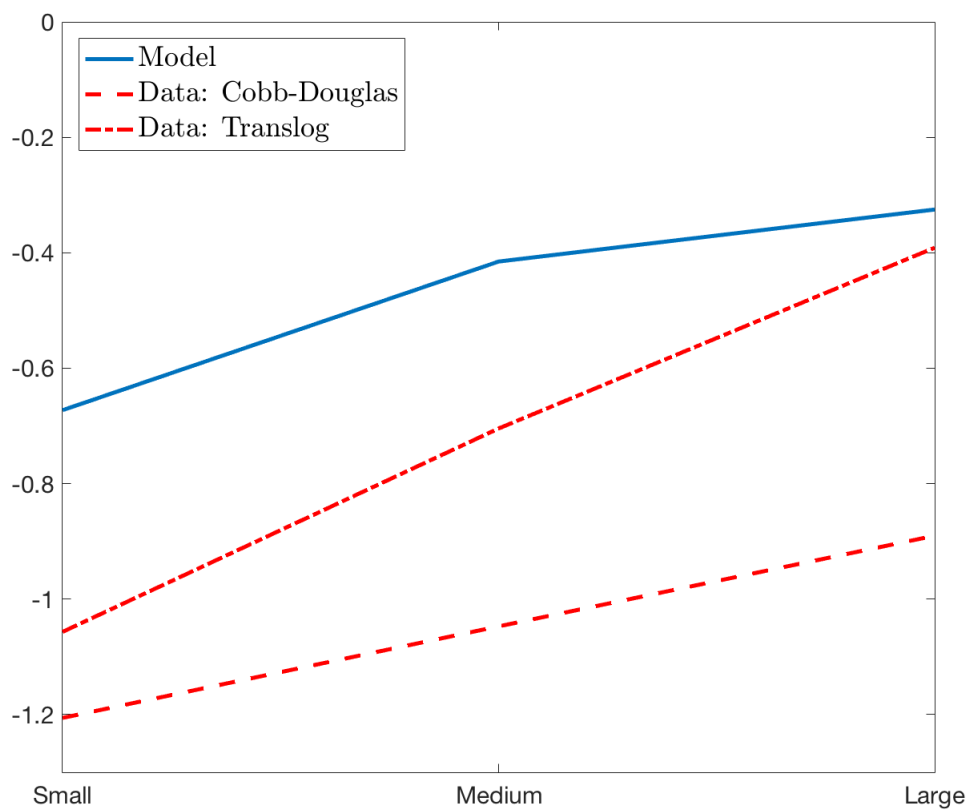
Target	Data	Model
Average Exit (Entry) Rate	7.5%	8.5%
Entrant's Relative Size	51%	54%
Exiter's Relative Size	41%	43%
$corr(s_{it}, s_{it-1})$	0.96	0.94
IQR of $std(\Delta \log s_{it})$	0.22	0.22
Autocorr. of Aggregate Output	0.609	0.608
Std. Dev. of Aggregate Output	0.027	0.027

Table 10: Heterogeneity in Sales and Labor Cyclicity

	(1)	(2)	(3)
	Small	Medium	Large
Model: $\Delta \log \left(\frac{p_{it} y_{it}}{p_t^D} \right)$	1.1284	1.0517	1.0036
Data: $\Delta \log \left(\frac{p_{it} y_{it}}{p_t^D} \right)$	1.3725	1.3650	1.1249

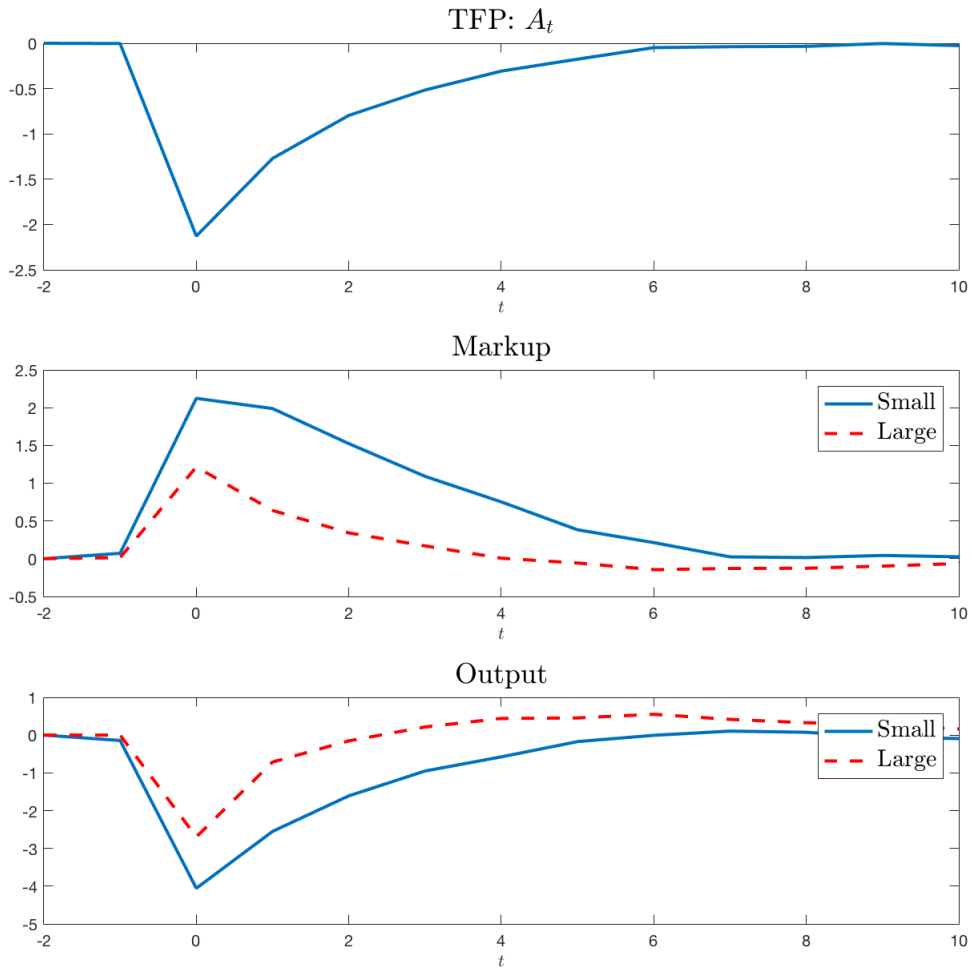
E Graphs

Figure 5: Markup Cyclicity by Firm Size: Model vs Data



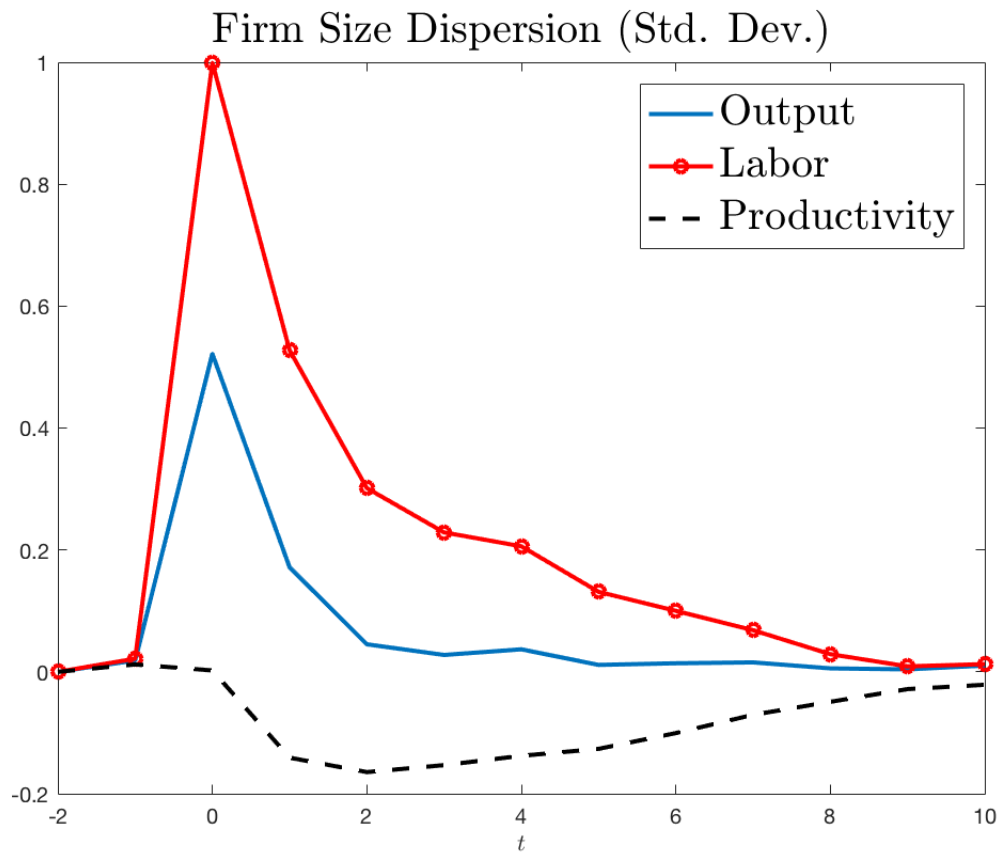
Note: Markup cyclicity with respect to aggregate consumption X_t^D . Firms in regression samples are split into small, medium, and large categories by market share.

Figure 6: Impulse Response to TFP A_t



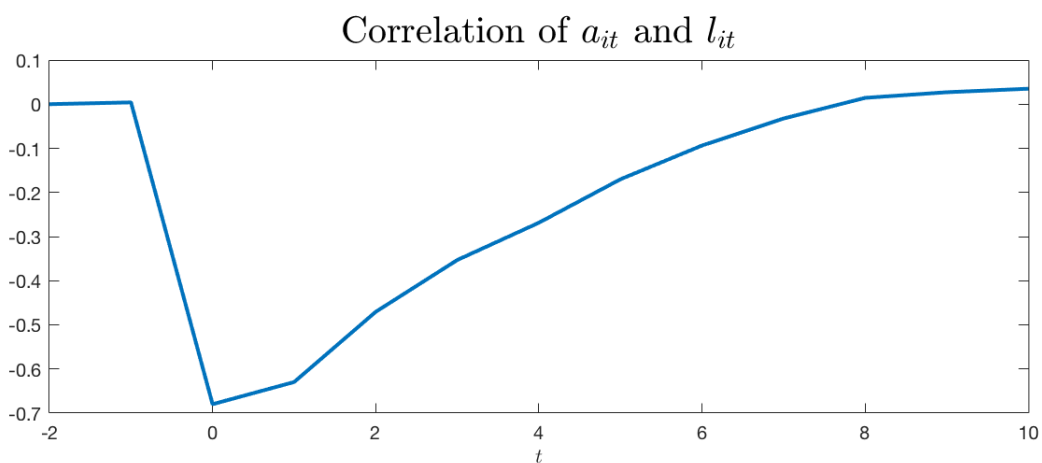
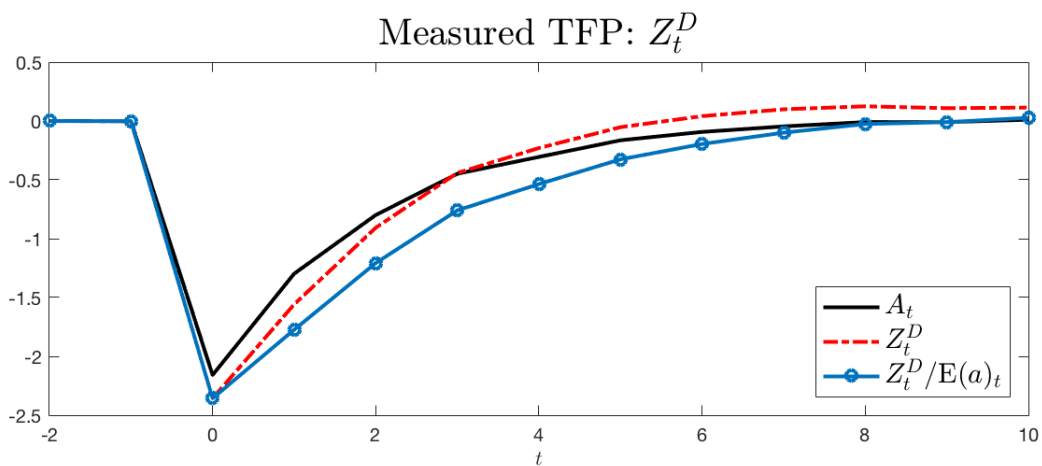
Note: Heterogeneous Responses of Small and Large Firms.

Figure 7: Dispersion of Firm Size and Productivity



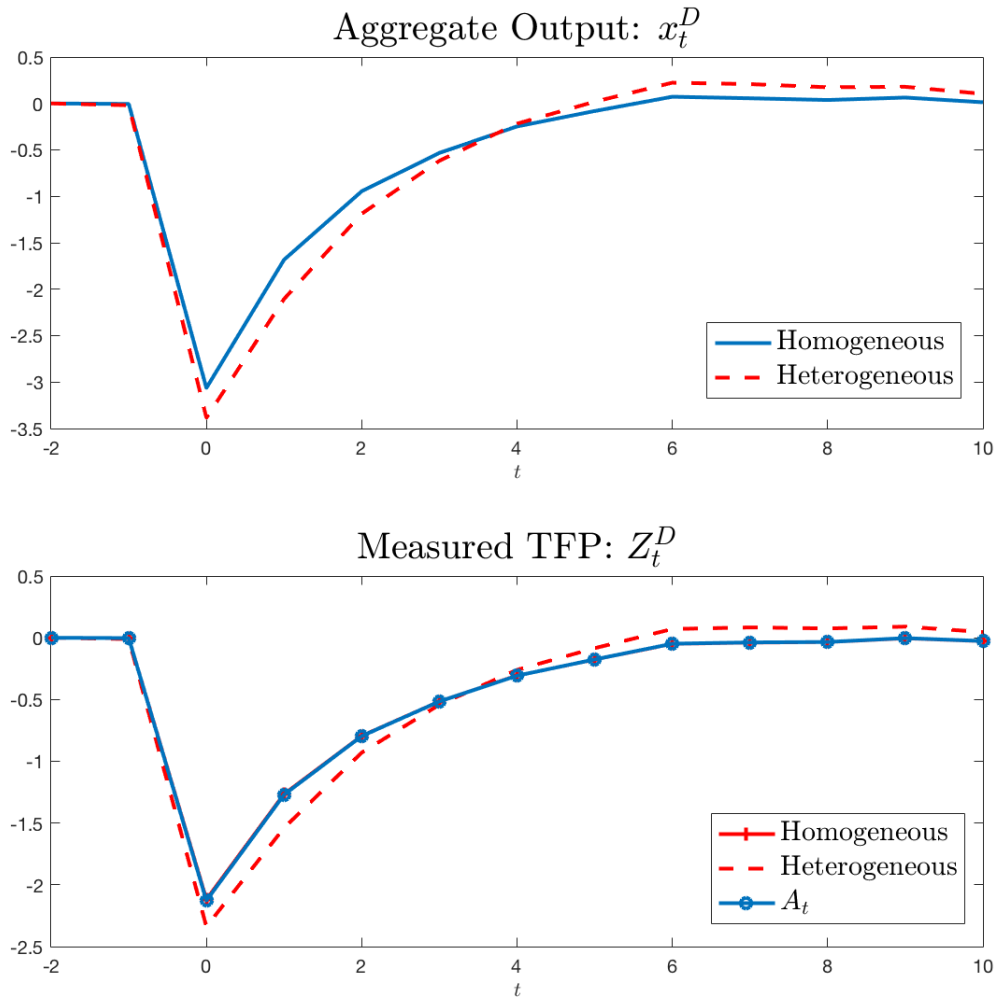
Note: The solid line depicts the standard deviation of output y_{it} . The dotted-solid line depicts the standard deviation of labor l_{it} . The dashed line depicts the standard deviation of idiosyncratic productivity a_{it} .

Figure 8: Comparison of Aggregate Productivity A_t and Measured TFP Z_t^D



Note: The top panel depicts the time profile of aggregate productivity A_t , the measured TFP Z_t^D , and the measured TFP net of compositional effect $Z_t^D/E(a)_t$. The dashed line depicts the standard deviation of idiosyncratic productivity a_{it} . The bottom panel depicts the correlation between idiosyncratic productivity a_{it} and labor l_{it} .

Figure 9: Comparison: Homogeneous vs Heterogeneous Firm Model



Note: The Heterogenous case stands for the baseline case with heterogenous agents with entry and exit. The Homogenous case stands for the case with homogeneous firms without entry and exit.