MARKUP, CUSTOMER BASE, AND FIRM DYNAMICS

CHANWOO KIM*

January 25, 2022

This paper studies the implication of firm-level market power by demand accumulation in markup cyclicality and lifecycle growth. I document two pieces of new empirical evidence: (i) individual firm markup is countercyclical to aggregate productivity and monetary policy shocks, and (ii) smaller firms have more countercyclical markups in response to the shocks. To explain the new evidence, I propose a firm dynamics model with the customer base and endogenous entry and exit. The model is quantitatively consistent with the data and can endogenously match age-dependent growth of firms.

*Research Department, Bank of Korea, 67, Sejong-daero, Jung-gu, Seoul, 04514, Rep. of KOREA, Email: ckim@bok.or.kr. This paper is the first chapter of my PhD thesis. I am grateful to my supervisor, Morten Ravn, for invaluable support and comments. I would like to thank Boragan Aruoba, Wei Cui, Martin Ellison, Jordi Gali, Simon Gilchrist, Gregor Jarosch, Patrick Kehoe, Keith Kuester, Sangyoon (Tim) Lee, Yoonsoo Lee, Francesco Lippi, Ralph Luetticke, Albert Marcet, Aureo De Paula, Facundo Piguillem, Franck Portier, Matthew Read, Tatsuro Senga, Vincent Sterk, Stephen Terry, Dajana Xhani, and participants in seminars at UCL, Bank of Canada, Sogang University, Korea University, and Korea International Economic Association for their thoughtful suggestions and comments. The view expressed in this paper does not represent the official view of Bank of Korea. All errors are mine.**JEL Codes: E3, E5, D2, D5, L1, L2, M1.**

Growing body of literature studies lifecycle firm behavior[1] and the implications of the heterogeneous firm behavior on macrodynamics[2]. This paper adds to this line of literature using microlevel data and a model that emphasizes individual firm behavior. Specifically, I start with new empirical evidence that the firm-level impulse response of markup is countercyclical to aggregate supply and monetary policy shocks. I further document that firms with small customer base show more countercyclical markup. Since I am not aware of any theory that explains the empirical result[3], I provide a theory that can explain the empirical evidence. Furthermore, the paper provides an interesting result that the proposed theory can endogenously explain the lifecycle growth and exit rates of firms.

Countercyclical markup plays a key role in the amplification of shocks in macro models by shifting the labor demand curve[4] in the direction of the shocks. Examples are firm entry and exit models (Jaimovich and Floetotte 2008) to productivity shocks and New Keynesian models (Christiano, Eichenbaum, and Evans 2005; Smets and Wouter 2007) to demand shocks. Given the importance of the markup cyclicality in macro models, researchers try to measure markup cyclicality using different approaches. However, studies tend to find little agreement. Therefore, I first document that the impulse response of firm-level markup is *countercyclical* to aggregate productivity and monetary policy shock using a panel version of local projection. Relative to existing literature, this approach is more granular and model consistent.

Using the granularity of my data, I further find that the impulse responses of firms with *small* demand bases have *more* countercyclical markup to the aggregate

1. For example, Foster, Haltiwanger, and Syverson (2008, 2016); Haltiwanger, Jarmin, and Miranda (2013)

2. For instance, Khan and Thomas (2008); Sedlacek and Sterk (2017); Ottonello and Winberry (2020)

3. For example, New Keynesian models have procyclical markup to productivity shock since price is rigid and marginal cost is flexible. I further am not aware of a paper that studies customer base dependent markup cyclicality.

4. "Countercyclical markup is like salt in cooking" (Basu, 2016) summarizes the importance of markup in macro models.

supply and demand shocks. Existing studies tend to focus on the heterogeneous response of sales (Gertler and Gilchrist 1994; Crouzet and Mehrotra 2018) or employment (Moscarini and Postel-Vinay 2012) but not markup[5]. If firm-level markup depends on the size of the demand base, then the firm distribution plays an important role in the amplification of shocks.

To explain new empirical evidence, I propose a firm dynamics model under monopolistic competition. Additional feature of the model is the demand accumulation mechanism in the form of deep habits (Ravn, Schmitt-Grohe, and Uribe 2006) and with endogenous entry and exit (Clementi et al. 2014). When I take the model to match salient features of US firm dynamics, the model can *quantitatively* match the empirical evidence without targeting any moments related to the impulse responses. The key mechanism is the tradeoff between the current profit and the future benefit from demand accumulation.

The other important finding of my model is that it can *endogenously* explain the *age-dependent growth rate* of firms. Understanding the firm's growth mechanism is not only an interesting question itself (Gibrat 1931) but is also helpful in understanding the lower frequency movement of the key variables (Sedlacek and Sterk 2017). In contrast to the pioneering study of Gibrat (1931), which claims that a firm's growth speed is independent of its size, recent studies show that small or young firms grow faster than big and old firms (Dunne, Roberts, and Samuelson 1989; Luttmer 2007, 2011; Decker et al. 2014; Pugsley, Sedlacek, and Sterk 2019). However, to my knowledge, no one has taken the model to the data to show that the model can endogenously match the data. I show that the model can closely match the lifecycle growth rate of firms in the data without targeting any moments related to the growth rate.

To study the impulse responses of firm-level markup, I use a unique combination of existing literature. I first identify the firm-level markups from COMPUS-

5. Hong (2019) is the only paper I am aware of that studies size-dependent markup. However, he studies the current correlation between GDP and firm-level markup.

TAT data using the production approach in the line of Hall (1986), De Loecker and Warzynski (2012), and De Loecker and Eeckhout (2017). Then, I take the identified aggregate supply and monetary policy shock shocks from the literature, i.e., Fernald (2014) and Gertler and Karadi (2014). Finally, I employ a panel version of local projection (Jorda, 2005), using the firm-level markup and aggregate shocks to recover nonlinear impulse response functions to aggregate shocks. To explore demand base-dependent markup cyclicality, I use the mean group estimator. Since I cannot find the direct data on the demand base, I use sales as a proxy for the demand base under the assumption that the demand base and the sales are positively correlated.

My model is a firm dynamics model with customer markets and endogenous entry and exit. The key difference between my model and a standard firm dynamics model is that a firm is concerned with both its productivity and its customer base under monopolistically competitive environment. In a standard firm dynamics model, a firm is concerned only with productivity[6]. Customer base in my model is a group of loyal customers that buy the product of a firm repeatedly. In other words, I model a type of behavior that consumers buy Nike because everyone else bought Nike. This deep habit preference assumption provides the foundation for a demand curve that shifts outward as the customer base accumulates. Furthermore, I model that the entry and exit of a firm are subject to idiosyncratic and aggregate economic conditions. Since exit is endogenous to the amount of customer stock a firm has, the customer base plays two roles, i.e., the demand base and insurance (Gilchrist et al. 2017).

In the model, firms face a dynamic tradeoff between the current profit and the future value of the customer base. The customer base is a fraction of past sales; therefore, a firm's pricing problem becomes dynamic. Firms can invest in the customer base by charging low markup today to harvest from the customer base by charging high markup in the future. This invest and harvest incentive is at the heart

6. One can include factor input as an individual state variable; however, I abstract from this margin.

of markup determination.

My model can endogenously match the lifecycle behavior of firms. Young firms choose to grow fast: their incentive to invest is higher than that of old firms because young firms have lower customer base. Since their demand base is low, young firms want to invest in the customer base using their prices. As a firm grows and customer base accumulates, net benefit of lowering markup decreases. Therefore, markup increases as a firm ages.

The model shows countercyclical markup in response to productivity shock and monetary policy shock due to the dynamic tradeoff. When there is a positive supply shock, it is good time to invest in customer base since a firm's marginal cost is low; therefore, firms want to invest further in the customer base by lowering markup. When there is an expansionary monetary policy shock, firms decrease markup to attract more customers since the current demand is larger than the future demand. This insight is consistent with a search-theoretic customer base model with an endogenous opportunity cost of search (Paciello, Pozzi, and Trachter, 2019).

I find that the markups of lower customer base firms are more countercyclical to productivity and monetary policy shocks. For positive productivity shock, larger customer base firms have less incentive to decrease markup since lowering markup is more costly given a bigger customer base. For expansionary monetary policy shock, the exit risk of large customer base firms decreases less than that of small firms; therefore, big firms decrease markup less than small firms. This finding implies that the aggregate response to shock is affected by the firm distribution.

Literature Review. Given the importance of markup cyclicality in macro models, researchers try to measure markup cyclicality using different approaches. Depending on the aggregation level of markup and the measure of the business cycle, the existing research can be summarized into three categories. The first line of research investigates the correlation between a certain measure of aggregate markup and a measure of business cycles (Bils, 1987; Rotemberg and Woodford, 1991). The second line of research considers the static correlation between firm-level markup and aggregate output (Hong, 2019). The third line of research studies the impulse response of aggregate markup to aggregate shocks (Nekarda and Ramey, 2019). My result adds to this line of literature by studying impulse response of firm-level markup and aggregate shocks.

This paper is related to strand of literature that combines the firm dynamics literature with the customer base studies. Small but growing literature study the firm dynamics with an emphasis on the demand side, such as the models of Arkolakis (2010), Dinlersoz and Yorukoglu (2012), and Sedlacek and Sterk (2017). Customer base models start from Phelps and Winter (1970) and the models used to explain macro- and international economics, such as Rotemberg and Woodford (1991, 1995), Drozd and Nosal (2012), Gourio and Rudanko (2014). My model contributes to this line of literature by studying the aggregate response of firm-level markup and by showing that the model can match the selection and the growth of a firm in the data.

This paper is also related to studies that firms or products penetrate into new markets(Melitz, 2003; Arkolakis, 2010; Fitzgerald, Haller, and Yedid-Levi, 2017). The results in this paper adds to this strands of works by providing empirical and theoretical results that how US listed firms react to aggregate shocks while they grow.

Two independent studies explored similar environments. Hong (2019) combined deep habits and firm dynamics. My work shares Hong's results that markup is countercyclical to aggregate productivity shock. However, by assuming decreasing return to scale, Hong cannot isolate the effect of the demand accumulation mechanism. Gilbukh and Roldan (2021) also studied the firm-dynamics model with demand accumulation under a product search and match environment. They found that markup is procyclical to aggregate supply and demand shocks. However, in their directed search model, the customer only considers the present value of the utility from the match while ignoring the current price of the product. This property is due to the linear utility function of the buyers and sellers, limiting the role of price to being simply allocative. Moreover, due to the block recursive property, the agents' payoff is independent of the firm distribution.

Roadmap. In the next section, I establish two pieces of new empirical evidence. To do so, I first explain the production approach to identify an individual firm's markup. Then, I show how to measure the impulse response of markup using an individual firm's markup decisions. After documenting the empirical evidence, I present a demand-driven firm dynamics model with endogenous entry and exit. Using the model, I show the analysis at a steady state and with aggregate shocks. Finally, I conclude.

I. DATA ANALYSIS

I use the production approach to obtain an individual firm's markup in the line of Hall (1986) and De Loecker and Warzynski (2012). Using the identified markup from the production approach, I execute local projection (Jorda 2005) to find the impulse response function upon aggregate shocks.

In this section, I closely follow De Loecker and Eeckhout (2017) to estimate an individual firm's markup[7]. The production approach builds on the insight that, in a perfectly competitive market ant under the Cobb-Douglas production function, the output elasticity of a variable input is equal to its expenditure share of total revenue. Therefore, the gap between the two is viewed as a markup that comes from imperfect competition.

Production approach is useful in many aspects. First, it can be applied to the general environment since it does not require any assumption of a market structure or demand system. Second, it relies on cost minimization. Therefore, the approach

7. Recent paper by Bond, ... doubts the identification power of the production approach. Despite their concern, I still find the approach is useful in obtaining firm level markup.

can be used if there exists a static input factor., which will be useful in the theory aspect. Third, the method does not require observing or measuring the user cost of capital.

Since the method is widely used in the recent literature, I attempt to be concise in explaining the framework. For details, please see De Loecker and Eeckhout (2017) or De Loecker and Warzynski (2012).

I.A. *Production Approach*

Consider an economy that consists of a continuum of firms that want to minimize their cost. Firm i's cost minimization problem is as follows:

$$\mathscr{L}_{it} = \min_{L_{it},K_{it}} W_{it}L_{it} + r_{it}K_{it} - \Lambda_{it}(Q_{it}(L_{it},K_{it}) - \bar{Q}_{it})$$

where W denotes wage, L is labor supply, r is the user cost of capital, k is capital, q is production quantity, and $\bar{Q}$ is scalar. I assume that labor is a variable input. footnote¡It can be extended the assumptions to include many variable inputs and many fixed or dynamic inputs.¿ By differentiating the labor input, I obtain the optimal labor input demand condition.

$$\frac{\partial \mathscr{L}_{it}}{\partial L_{it}} = W_{it} - \Lambda_{it}\frac{\partial Q_{it}}{\partial L_{it}} = 0$$

I note that Λ_{it} is a measure of marginal cost. Intuitively, the above equation shows that marginal cost equals the cost of hiring one unit of labor over the marginal labor productivity ($MC_{it} = \frac{W_{it}}{MPL_{it}}$). The definition of output elasticity to variable cost is

$$\theta^l_{it} = \frac{\partial Q_{it}}{\partial L_{it}}\frac{L_{it}}{Q_{it}} = \frac{1}{\Lambda_{it}}\frac{W_{it}L_{it}}{Q_{it}}$$

By rearranging the definition of output elasticity, I obtain an equation for markup.

$$\mu_{it} \equiv \frac{P_{it}}{MC_{it}} = MPL_{it}\frac{P_{it}Q_{it}}{W_{it}L_{it}}\frac{L_{it}}{Q_{it}} = \theta_j^L \frac{P_{it}Q_{it}}{W_{it}L_{it}} \tag{1}$$

The equation (1) consists of two parts: Output elasiticity and the expenditure share of labor cost. The share of labor cost to total sales is easily found in firms' financial statements; therefore, I focus on how to recover the output elasticity of variable input from the data.

Control function approach, by using some economic structure allows me to circumvent the endogeneity problem between input choice and unobserved productivity. Relative to other approaches that do not use economic structure, I do not need to find an instrument that is very difficult to find. Furthermore, it does not need relatively strong assumptions such as such as fixed productivity, perfectly competitive input and output markets. The approach also works even if a subset of inputs is dynamic. To test possible identification concern raised by a recent paper(Gandhi, Navarro, and Rivers, 2020), I estimate the output elasticity using dynamic panel approach (Blundell and Bond, 1998) and finds robust results.

This paper exploits a following industry-specific Cobb-Douglas production function for main results.

$$\begin{aligned} Q_{it} &= L_{it}^{\theta_j^l} K_{it}^{\theta_j^k} \exp(\omega_{it}) \\ \tilde{q}_{it} &= \theta_j^l l_{it} + \theta_j^k k_{it} + \omega_{it} + \varepsilon_{it} \end{aligned} \tag{2}$$

where i denotes an individual firm, j denotes the industry, and $\omega_{it} = h(l_{it}, k_{it})$ is idiosyncratic productivity that follows an AR(1) process. The second equation is obtained by taking the log of the first equation and adding measurement error (ε_{it})[8]. I notice that the function is written in the form of value-added production function. However, I can interpret the function as the Leontief gross output production function, in which intermediate input is proportional to the output (Ackerberg, Caves,

8. One can think of it as an independent and identically distributed (IID) productivity shock that is unknown at the point of production decision

and Frazer 2015)[9]. Furthermore, the the result is robust to a general production function, e.g., the translog[10].

To estimate the equation (2), I use two-step GMM. The first step involves purging measurement error and productivity. Specifically, I regress sales on labor, capital, time dummies, and a constant. Then, I set $\hat{q}_{it}$ as the true output and obtain productivity (ω_{it}) by calculating $\omega_{it} = \hat{q}_{it} - \hat{\theta}^l_j l_{it} - \theta^k_j k_{it} -$ constant. The second step is GMM. I regress the obtained productivity on its lag: the residual (ξ_{it}) is the shock to productivity. Then, I use two-moment conditions to find two parameters.

$$E[\xi_{it}(\hat{\theta}^l_j, \hat{\theta}^k_j) l_{i,t-1}] = 0$$

$$E[\xi_{it}(\hat{\theta}^l_j, \hat{\theta}^k_j) k_{i,t}] = 0$$

The key assumptions in the estimation are that the past variable input use is (i) independent of the current period productivity shock and (ii) related to the current period variable input use. The timing guarantees the first assumption, and the AR(1) process of productivity supports the second assumption. The last step is to adjust for measurement error.

$$\mu_{it} \equiv \frac{P_{it}}{\Lambda_{it}} = \theta^l_j \frac{P_{it} Q_{it}}{W_{it} L_{it} \exp(\varepsilon_{it})}$$

In this approach, the Solow residual is the sum of idiosyncratic productivity and measurement error; therefore, to find the "true" quantity, I need to eliminate the measurement error component using the residual from the first stage.

I.B. Data

I choose COMPUSTAT data from Wharton Research Data Services (WRDS). The choice of COMPUSTAT is based on following three reasons. First, the database is the only publicly available source that covers firms in all industries. Second,

9. Using this specification, the model does not suffer from the functional dependence problem. See Appendix for details.

10. The result can be found in Appendix.

COMPUSTAT provides detailed financial statement variables that works with the empirical strategy. Third, the database covers a significant fraction of the economy. For example, dataset that covers only manufacturing accounts for less than 10% of GDP while COMPUSTAT covers approximately 30% of employment (Davis et al. 2007), and more fraction of GDP. For data-cleaning, I follow De Loecker and Eeckhout (2017)[11].

For total factor productivity (TFP) shock, I use Fernald's (2014) utilization-adjusted productivity for the US business sector. The utilization-adjusted productivity is developed to consider the fact that standard TFP includes the change in factor use, such as labor effort and the workweek of capital. The approach finds data on inputs using careful growth accounting as used by the Bureau of Labor Statistics (BLS). Unobserved utilization, i.e., labor effort and capital utilization, is estimated using hours per worker as a proxy under the assumption that firms optimize their input choice.

I use a high-frequency identification approach to identify the monetary policy (MP) shock. This approach is useful to address a possible forecast problem in identifying the monetary policy shock. Specifically, I use high-frequency identification shock (Ramey 2018). Ramey claims that Gertler and Karadi's (2015) high-frequency identification shock[12] has serial correlation, which comes from the method that Gertler and Karadi used to convert the announcement day shocks to a monthly series. Therefore, I use Romer and Romer's (2004) method to generate annual shocks[13] following the suggestion of Ramey (2018). The high-frequency identification approach was pioneered by Cook and Hahn (1989) and is widely used in the literature[14]. Under the assumption that most of the information related to monetary policy is revealed around the FOMC meeting, the approach uses the change

11. For detailed procedure, Essentially, I present the data-cleaning procedure in Appendix.

12. I choose Eurodollar six-month future data since it has the longest sample period.

13. The procedure is as follows: First, create a cumulative daily monetary policy shock series. Second, take the difference between the end-of-the-year level and the beginning-of-the-year level of the cumulative shock series.

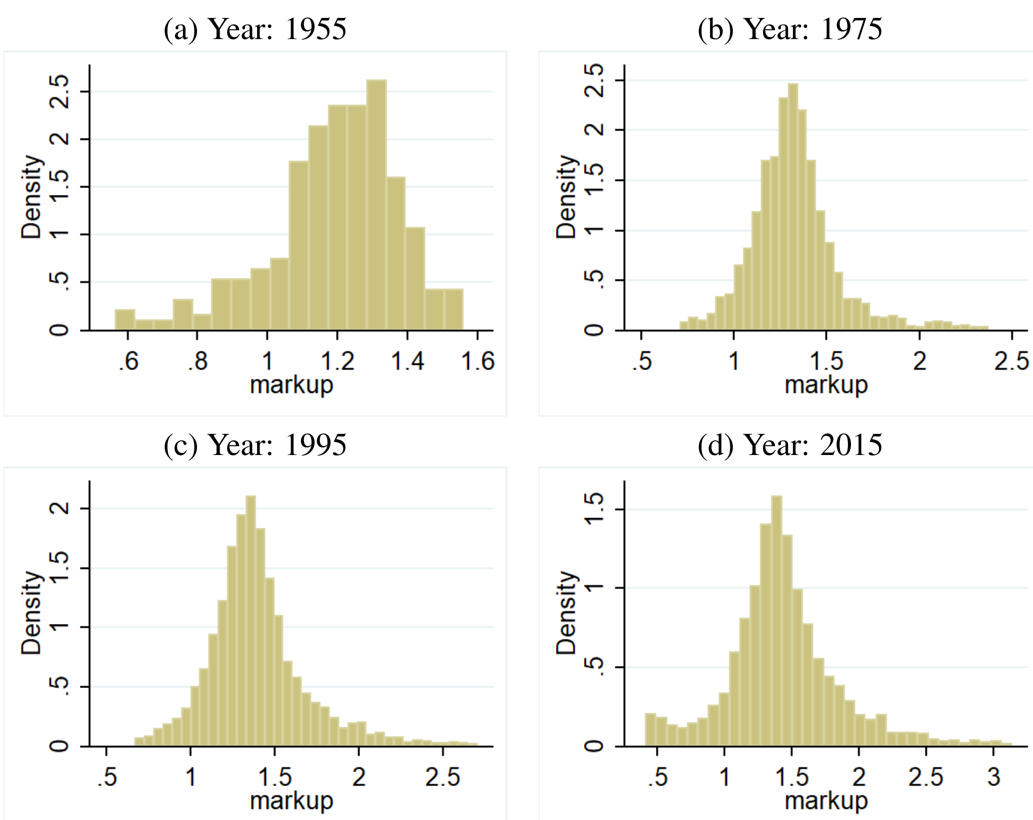
14. For example, see Kuttner (2001), Cochrane and Piazzesi (2002), Gurkaynak, Sack, and Swanson (2005), Campbell, Evans, Fisher, and Justiniano (2012), and Nakamura and Steinsson (2018).

in the bond price within a small time window. I normalize the shock so that the increase in the shock reflects the expansionary monetary policy.

I.C. Markup Distribution

In this section, I describe how markup distributions have evolved over the last couple of decades (figure 1). Over time, the mean and the variance of the markup distributions increase, whereas the skewness of the distributions decreases. I note that markup can even be lower than one.

FIGURE 1: CHANGE IN MARKUP DISTRIBUTION



Note: Markup distributions are truncated at 1st and 99th percentile

I.D. Markup Response to Aggregate Shocks

In this section, I show the markup response to aggregate shocks using a panel version of local projection (Jorda 2005; Jord, Schularick, and Taylor, 2016). To estimate the impulse responses, I first take the log of all variables except shocks, age, and market share. I then use an industry-specific[15] quadratic time series to eliminate the trend for relevant firm-level variables and use a quadratic time series to remove the macrotrend for GDP. Last, I estimate dynamic responses of markup to aggregate shocks. All robustness exercises are summarized in the Appendix.

Impulse Response Analysis. Using the identified shocks from the data and the literature, I find the response of markup to aggregate productivity and monetary policy. Specifically, I regress

$$\text{Markup}_{i,t+h} = \gamma_1^h + \gamma_2^h \text{Shock}_t + \gamma_3^h \text{Markup}_{i,t-1} + \gamma_4^h \text{Control}_{i,t-1} + \gamma_i^h + e_{i,t+h}$$

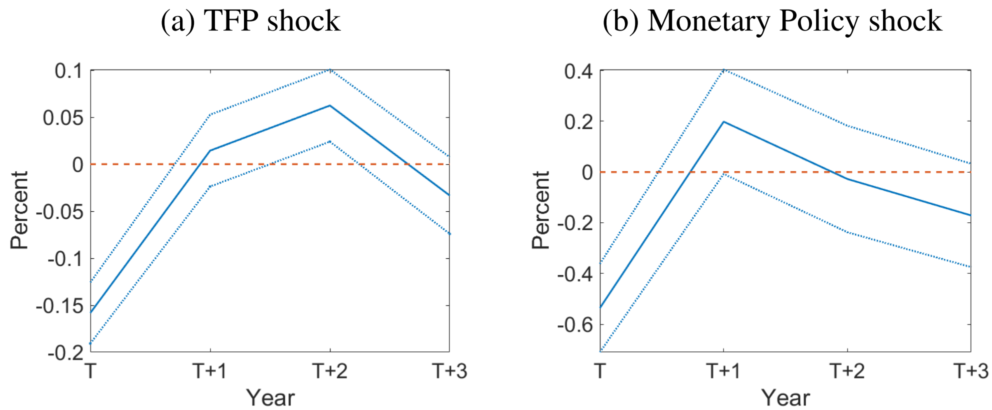
γ_2 captures the average cross-sectional percent change in markup due to an aggregate shock. The control variables are GDP, firm size, market share, age, productivity, and sales effort. Since my data are annual, I set $h = \{0, 1, 2, 3\}$.

I find that the individual average markup is countercyclical to productivity and monetary policy shocks (figure 2). The solid line in the figure represent the level of coefficients (γ_2), and the dotted lines show the 95% confidence interval. The regression tables are provided in Appendix for detailed results of the regression.

The left panel of figure 2 shows that a 1% increase in TFP can cause a 0.15% average decrease in individual markup on impact and that the effect disappears in the next year. The right panel of figure 2 illustrates the response of the firm-level markup to a one-unit change in the six-month future Eurodollar price due to expansionary monetary policy shock. It shows that the average firm-level markup decreases by 0.5%, and the effect disappears the next year.

15. I choose two-digit industry due to sample numbers. For robustness, I use the first difference for detrending and obtain a similar result.

FIGURE 2: IMPULSE RESPONSE OF MARKUP



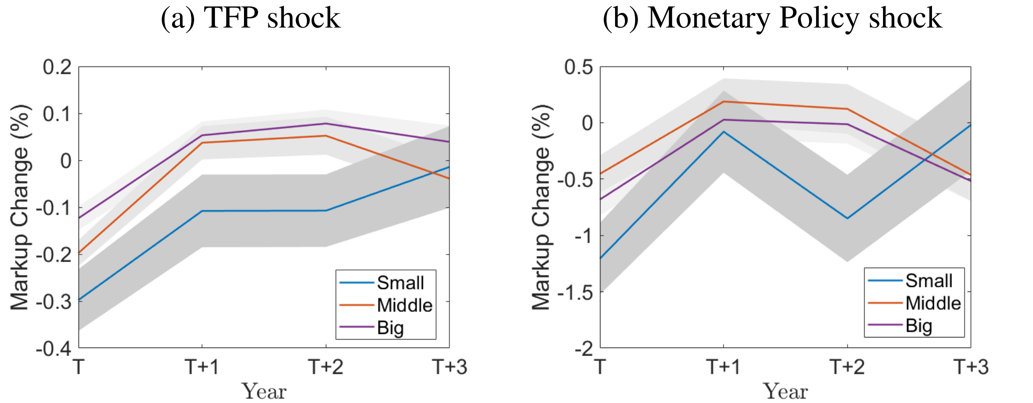
Note: Solid lines are firm-level responses, and dotted lines are 95th percentile confidence intervals. I use heteroskedasticity and autocorrelation robust standard errors to calculate the interval.

Size-Dependent Markup Response. In this section, I study how the markup response differs depending on the customer stock of a firm. To proceed, I first pool the data and divide it into three bins according to the revenue of each data point. Under the assumption that revenue and customer base are positively correlated, I use revenue[16] as a proxy for the customer base. Then, I use the mean group estimator to find the size-dependent markup responses.

I find that the markup of a smaller firm is more countercyclical to productivity and monetary policy shocks (figure 3). The solid lines are the average response of firms in the group, and the shaded area is the 68th percentile confidence interval. For the TFP shocks and the monetary policy shock (the left and right panels of figure 3, respectively), the response of smaller firms is different from that of medium and large firms. Furthermore, the difference tends to persist for some years after. In the model section, I explain the mechanism behind the data.

16. Here, I use current revenue. The result is robust to the use of lagged revenue.

FIGURE 3: SIZE-DEPENDENT RESPONSE OF MARKUP



Note: Solid lines are firm-level responses, and shaded areas are 68th percentile confidence intervals. I use heteroskedasticity and autocorrelation robust standard errors to calculate the interval.

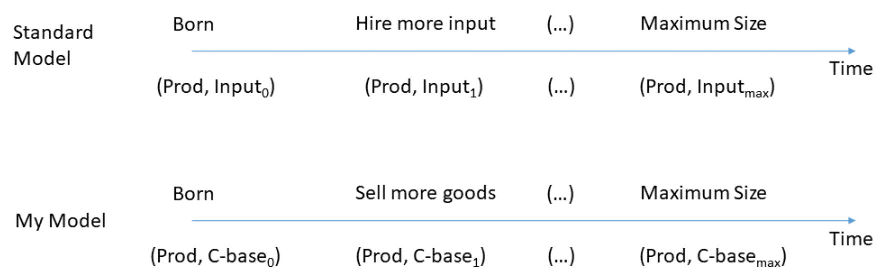
II. MODEL

In this section, I propose a model that captures the salient features of the data. In the model, firms entry and exit decision depend on their own characteristics as well as macro economic conditions(Hopenhayn and Rogerson 1993; Clementi et al. 2014). Also, households form past external habit at good level which I consider as customer base(Ravn, Schmitt-Grohe, and Uribe 2006). The customer base, which means a group of loyal customers who buy the good repeatedly, is a fraction of the past sales quantity in the model.

The key distinction between my model and standard firm dynamics models are in demand specification. A firm operates in perfectly competitive goods market and characterized by a productivity and factor inputs such as labor and capital. In this paper, a firm is in monopolistic competitive goods market and characterized by productivity and a customer base (figure 4). To grow, firms want to accumulate the customer base and use single price to invest in the customer stock.

Customer base is modelled as deep habit, or households' habit formation at good level. In other words, a fraction of households become a loyal customer of a spe-

FIGURE 4: MODEL COMPARISON



cific good. Hence, demand is derived from the households' optimization problem differently from search theory based models. Since a fraction of current customer becomes future customer, firms need to compare the value of current profit to attract one more customer today and exploit from the customer in the future. This invest and harvest incentive is at the heart of firms' pricing decision.

In addition to deep habits, the model also contains an endogenous exit. The endogenous exit implies that firms' survival probability changes depending on the size of the customer base. A firm with a greater customer base can survive longer upon a series of adverse shocks since it still has loyal customers. Therefore, the customer base serves as a demand base as well as insurance (Gilchrist et al. 2017).

II.A. Setup

Environment. In this economy, there are three types of agents: a continuum of identical households, a continuum of heterogeneous incumbent firms, and a continuum of ex-post heterogeneous potential entrants. Households consume a product, supply labor, and trade bonds. Firms produce goods, hire labor, set prices, and accumulate customer base. Incumbent firms are heterogeneous to productivity and the customer base. Entrants are ex-ante homogeneous but differ after they draw random idiosyncratic productivity. A collection of households owns firms.

Agents interact in three markets: a monopolistically competitive goods market, a perfectly competitive labor market, and a complete financial market. Two idiosyncratic shocks and two aggregate shocks are considered. The idiosyncratic shocks are operating cost shock, and productivity shock. The two aggregate shocks are TFP shock and monetary policy shock[17].

Preference. Preference differs from a standard model in that utility depends on the habit stock of each product (Ravn, Schimitt-Grohe, and Uribe 2006). In other words, households form a past external habit regarding an individual product, which is often called "Catching up with the Joneses at good level.". To be more concrete, I denote households are $j \in [0,1]$ and a variety of consumption goods indexed by $i \in [0,M_t], M_t < 1$. Households' utility depends on habit-adjusted consumption bundle, $\tilde{c}_{jt}$, and the amount of labor, n_{jt}.

$$U_t^j = E_t \sum_{s=t}^{\infty} \beta^{s-t} [\frac{1}{1-\sigma} \tilde{c}_{jt}^{1-\sigma} - \omega n_{jt}]$$

where

$$\tilde{c}_{jt} = f(\{c_{ijt}, h_{it-1}\}_i)$$

The external habit can be considered as a type of brand equity. Given that my data is firm level, external habit is more consistent with the data than internal habit. I also find that since preference depends on external habit, there is no time consistency concern since atomistic households cannot affect the aggregate habit for each good[18]. Since the habit-adjusted consumption basket depends on a predetermined level of habit, which implies that a good's market demand depends on sales history. Thus, firms compare the future benefit of the current profit to the benefit of the customer base.

Technology. Firms produce goods using a constant return to scale technology[19], and labor is the only input. Relative to a standard firm dynamics model that relies on

17. In the model, I assume central banks can adjusted real interest rates directly. Monetary policy shock is exogenous change in real interest rates.

18. See Nakamura and Steinsson (2011) and Rudanko (2017) for a detailed discussion.

19. This assumption guarantees that size-dependent pricing is due to the demand factor, in contrast to Hong (2019) and Gilbukh and Roldan (2017).

decreasing return to scale to control firm growth, supply side is neutral by relying on constant return to scale technology. Firms' production function is following:

$$y_i = e^A e^{z_i} n_i \tag{3}$$

where aggregate productivity A is an AR(1) process, idiosyncratic productivity z_i is markov chain which I discretize with seven grids, and n_i denotes labor input.

$$A_{i,t} = \gamma_A A_{i,t-1} + \varepsilon_A, \varepsilon_A \sim N(0, \sigma_A) \tag{4}$$

$$z_{i,t} = \text{markov chain} \tag{5}$$

Firms face independent and identically distributed (IID) random operating costs (ζ_i).

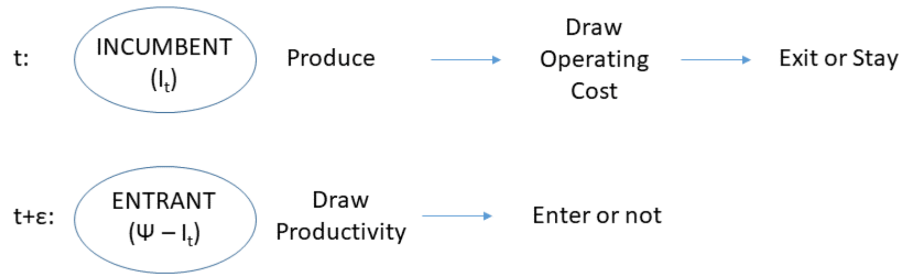
$$\zeta_i \sim N(\mu_\zeta, \sigma_\zeta) \tag{6}$$

The operating cost captures any shock on a firm's cash flow.

Firm Dynamics. In the economy, there exists a fixed number($= \Psi$) of potential goods (Clementi et al., 2014). Since a firm can produce only one good, potential entrants are determined as the number of potential goods. The timing of entry and exit is following: First, incumbent firms produce. Second, incumbent firms draw operating costs. Third, incumbent firms exit if they are hit by an exogenous exit shock or the value of the firm is lower than the operating costs. Fourth, entrants draw their productivity. Fifth, entrants decide whether to enter. To enter, entrants must have pay fixed entry cost.

Households' Problem. Households earns income from labor supply, n, and receives dividend, d, from firms. There exists complete financial market that allows households to save using risk-free one period bond, b. Hence, households' budget constraint is following:

FIGURE 5: TIMING OF THE FIRM'S DECISION



$$\tilde{p}_j\tilde{c}_j + E_t[\frac{b'_j}{(1+r)e^Q}] = b_j + Wn_j + d_j \tag{7}$$

where $\tilde{p}_j = g(p_{ij}, h_{i,-1})$ is the habit-adjusted price for habit adjusted consumption bundle $\tilde{c}_j$. Q is bond price shock, r is return on bonds, W is wage for unit labor supply. Under the assumption that monetary authority targets real interest rates directly, I can consider Q as a monetary policy shock.

$$Q_{i,t} = \gamma_Q Q_{i,t-1} + \varepsilon_Q, \varepsilon_Q \sim N(0, \sigma_Q) \tag{8}$$

Given homothetic and weakly separable preference, one can consider the households' problem as a two-stage budget problem. In the first-stage problem, households choose the amounts of habit-adjusted consumption basket ($\tilde{c}$), labor (n_t^j), and risk-free bonds (b_t^j) to maximize their discounted expected utility given the prices and aggregate state variables, $F = \{A, Q, M\}$:

$$V(F_{-1}) = \max_{\tilde{c}_j, n_j, b_j} [\frac{1}{1-\sigma}\tilde{c}_j^{1-\sigma} - \omega n_j + \beta EV(F)]$$

subject to the budget constraint (Equation 7) and the laws of motion for other aggregate state variables (Equations 4, 8, and 13-15).

The equilibrium conditions are

$$
\begin{aligned}
[\tilde{c}_j]: &\qquad \lambda_j = (\tilde{c}_j)^{-\sigma}\frac{1}{\tilde{p}_j} \\
[n_j]: &\qquad \lambda_j = \frac{\omega}{W} \\
[b'_j]: &\qquad \frac{1}{1+r} = \beta e^{Q} E \frac{\lambda'_j}{\lambda_j}
\end{aligned}
$$

where λ^j is the Lagrange multiplier related to the budget constraint. In the second stage, households solve the following cost minimization problem given $\tilde{c}^j$ and $\{\tilde{p}_i, h_{i,-1}\}_i$.

$$\min_{\{c_{ij}\}} \int_0^I \tilde{p}_{ij} c_{ij} di$$

subject to the habit-adjusted consumption bundle chosen at the first stage.

$$\tilde{c}_j = f(\{c_{ij}, h_{i,-1}\}_i)$$

Using the symmetry of the households, I integrate over the individual demand function from the cost minimization problem to obtain a demand function for each good.

$$c_i = c(\frac{p_i}{\tilde{P}}, \tilde{C}, h_{i,-1}) \tag{9}$$

where $\tilde{C} = \int_j \tilde{c}^j dj$ is the aggregate of habit-adjusted consumption and $\tilde{P} = \int_j \tilde{p}_j dj$ is the aggregate habit-adjusted price. Relative to a plain vanilla model, there exists $\tilde{C}$, a measure of aggregate demand, and $h_{i,-1}$ due to preference assumption.

Incumbent Firm's Problem. Incumbent firms have two idiosyncratic state variables ($= S$) and three aggregate state variables ($= F$). The idiosyncratic state variables are their productivity and customer capital, and the aggregate state variables are two aggregate shocks (supply and demand) and the distribution of firms.

The current customer base (h) is the sum of the customer base depreciated from the last period and the fraction of the current period sales quantity.

$$h_i = (1-\delta)h_{i,-1} + \delta c_i \tag{10}$$

δ is a measure of how fast the customer base adjusts. Since the acquisition and the depreciation of customer capital are set at the same speed, the maximum level of the customer base a firm can sustain is equal to the output ($h_i^* = y_i^*$).

Firms choose price to maximize the discounted stream of habit-adjusted real profit.

$$V(z_{-1}, h_{-1}; \mathrm{F}_{-1}) = \max_{p_i}\{\frac{p_i}{\tilde{P}}y_i - \frac{W}{\tilde{P}}n_i + \max_{\text{exit,stay}}[0, -\frac{e^{\zeta}}{\tilde{P}} + \frac{1}{1+r}EV(z, h; \mathrm{F})]\}$$

subject to the production function (Equation 3), operating cost distribution (Equation 6), demand function (Equation 9) and the laws of motion for the state variables (Equations 4, 5, 8, 9, and 13-15). I note that $F = \{M, A, Q\}$ represents aggregate state variables. A cut-off level of operating cost is the level that equates the habit-adjusted real operating costs and the discounted value of the next period.

$$\frac{\zeta^*}{\tilde{P}} = \frac{1}{1+r}EV(S'; F')$$

Survival probability ($G(\zeta^*)$) is obtained using the property of log-normal distribution.

$$G(\zeta^*) \equiv \Pr(\zeta \leq \zeta^*) = \Phi(\zeta \leq \frac{\log(\zeta^*) - \mu_{\zeta}}{\sigma_{\zeta}})$$

where Φ is a standard normal distribution.

Entrants' Problem. After incumbents make their production and exit decisions, the potential entrants make their entry decisions. Before entry decision, entrants draw their productivity from the long-run distribution of the idiosyncratic productivity process. Once they enter, entrants do not produce any output in the first period and become incumbents at the next period.

Potential entrants must pay fixed entry cost and pay one time advertising cost to set the initial customer base. One can randomly give initial customer base or some fixed initial customer base since the advertising assumption does not change any result of the model. I assume that each advertisement is a posting that contains information about the presence of a good in a market. All consumers are aware of the product once the advertisement is out, but only a fraction of consumers are attracted to the good by the advertisement. Therefore, the amount of advertising labor input determines the quality of the advertisement, which determines the initial customer base.

$$h_{k,0} = \alpha_1 y_{k,a} \tag{11}$$

The advertisement production function is in a generic form.

$$y_{k,a} = a(z_k, n_{k,a}) \tag{12}$$

Aspiring entrants will enter if the expected value of entry exceeds the sum of the advertising cost.

$$\hat{V}(z_k;\mathrm{F}_{-1}) = \max_{enter,not}\{\max_{h_{k,0}}\{-\frac{W}{\tilde{P}}n_{k,a} + \frac{1}{1+r}EV(z_i, h_{k,0};\mathrm{F})\}, 0\}\}$$

subject to the initial customer base condition (Equation 11), advertising production (Equation 12), and the constraints that incumbents face (Equations 3-10 and 13-15). The optimal level of the initial habit, $h_0^*(z;\mathrm{F})$, is implicitly defined by equating the value of entering and not entering.

$$\frac{W}{\tilde{P}}n^*(z_i, h_0^*) = \frac{1}{1+r}E_{\mathrm{F}'}V_{t+1}(z_i', h_0^*;\mathrm{F}')$$

Distribution Updating. The firm distribution (M) is updated by exogenous productivity shock, endogenous habit choice, and firm entry and exit. The current mass of firms is the sum of surviving incumbents and new entrants.

$$M(S;F) = M_i(S;F) + M_e(S;F) \tag{13}$$

$$M_i(S;F) = (1-\vartheta)G(\zeta^*)\int\int \mathbb{1}(z=z)\mathbb{1}(h=h^*)dM(z_{-1},h_{-1};F_{-1}) \tag{14}$$

$$M_e(S;\mathrm{F}) = M_a\int \mathbb{1}(z=z)\mathbb{1}(h=h_0^*)\mathbb{1}(\frac{W}{\tilde{P}}n_i \leq -\frac{\kappa}{\tilde{P}} + \Lambda E_{\mathrm{F}}V_{t+1}(z_i,h_0^*;\mathrm{F}))dG(z) \tag{15}$$

where M_e denotes the actual entrants' distribution, $M_a = \Psi - M_i$ is the aspiring entrants, and $h_0^*(z_{-1};\mathrm{F})$ and $G(\zeta^*)$ are implicitly defined by the exit and entry conditions.

II.B. Equilibrium

Recursive monopolistic competition equilibrium with entry and exit of firms consists of $\{\{p_{it}\}_i, W_t, r_t\}$, $\{\{c_{ijt}\}_i, n_{jt}, b_{j,t+1}\}_j$, $\{V_t, p_{it}, h_{it}, y_{it}, n_{it}\}_i$, $\{\hat{V}_t, y_{a,k,t}, n_{kt}, h_{0,k,t}\}_k$, and $\{M_t, M_{i,t}, M_{e,t}\}$ such that

1. Households maximize their utility and observe their budget constraint.
2. Policy functions ($\{p_{it}, h_{it},\ y_{it}, n_{it}\}_i$) and the exit condition solve the incumbents' problem.
3. Policy functions ($\{y_{akt}, n_{kt}, h_{0kt}\}_k$) and the entry condition solve the entrants' problem.
4. Incumbents exit if the operating cost is higher than the expected next period value, i.e., exit if $\zeta < \zeta^*$, where $\frac{\zeta^*}{\tilde{P}} = \frac{1}{1+r}EV(S';F')$
5. Entrants enter if the value of entry is higher than the entry cost, i.e., if $\frac{W}{\tilde{P}}n^*(z_i,h_0^*) > +\frac{1}{1+r}E_{\mathrm{F}'}V_{t+1}(z_i',h_0^*;\mathrm{F}'))$
6. Distributions ($M_t, M_{i,t}, M_{e,t}$) satisfy the law of motion.
7. All markets clear.

II.C. Functional Forms

In this section, I specify functional forms for quantitative analysis. First, the habit-adjusted consumption basket is

$$\tilde{c}_j = [\int_i^I (c_{ij} h_{i,-1}^{\theta_1})^{\frac{\rho-1}{\rho}} di]^{\frac{\rho}{\rho-1}}$$

where θ_1 represents the degree of habit that is price elastic and ρ indicates the elasticity of substitution. It gives the following demand function:

$$c_i = (\frac{p_i}{\tilde{P}})^{-\rho} \tilde{C} h_{i,-1}^{\theta_1(\rho-1)}$$

where $\tilde{P} = \int_0^1 \tilde{p}_j dj$ is aggregate price for habit adjusted consumption bundle and $\tilde{p}_j = [\int_0^{I_t} (\frac{p_{ij}}{h_{i,-1}^{\theta_1}})^{1-\rho} di]^{\frac{1}{1-\rho}}$ is the habit-adjusted price for habit adjusted consumption bundle $\tilde{c}_j$.

From the demand function, I notice that the model is equivalent to a standard real business cycle model if I turn off the habit by setting $\theta_1 = 0$. I further notice that the price elasticity of demand is fixed to ρ, unlike in many models where the price elasticity of demand is a function of market share.

To produce advertising, firms use labor and productivity using decreasing return to scale technology (Sutton 1991; Arkolakis 2010). This response can be due to media saturation or to differing tendencies to view ads among households (Grossman and Shapiro 1984).

$$y_{i,a} = e^{z_i} n_{i,a}^{\alpha_2}$$

II.D. Computation Approach

To solve the model, I first find the steady state of the model and use the first-order perturbation to analyze the aggregate dynamics (Reiter, 2009). Detailed computation approach is in the Appendix.

II.E. Calibration

The model is calibrated in three steps. First, I set certain parameters based on external information. Second, I calibrate the other parameters to match the moments at the steady state except for the standard deviations of the aggregate shocks. Last, I simulate the model to determine the standard deviations of aggregate shocks to match the aggregate business cycle moments.

TABLE I: FIXED PARAMETERS

Parameters	Explanation	Source
Households		
$\beta = 0.99$	discount factor	Annual Interest Rate $\approx 4\%$
$\sigma = 2$	intertemporal subs.	Attanasio and Weber (1995)
$\delta = 0.04$	habit depreciation	Gourio and Rudanko (2014)
$\rho = 3.3$	elasticity of subs.	Broda and Weinstein (2006)
Shocks		
$\gamma_z = 0.84$	idio productivity persistence	COMPUSTAT
$\gamma_A = 0.823$	AGG TFP persistence	Smets and Wouter (2003)
$\gamma_Q = 0.855$	Bond shock persistence	Smets and Wouter (2003)

Fixed Parameters. The set of parameters I that calibrate using external information is presented in Table 1. These parameters are related to preference and shock processes. Starting from preference parameters, I fix the discount factor (β) at 0.99 to set an annual interest rate of 4% since time is quarterly. I then fix the consumption smoothing parameter at two, which is in the mid-range of Attanasio and Weber (1993). I set the habit depreciation parameter (δ) to 0.04, which implies an approximately 15% depreciation in the customer base annually. This level is used in Gourio and Rudanko (2014) based on literature indicating that the turnover rate for the cell phone industry is approximately 11% to 26% and the turnover rate for the banking industry is approximately 10% to 20%. The estimates for the elasticity of the substitution parameter (ρ) vary significantly by the type of products. I use 3.3, which is in the mid-range of median elasticity of finely[20] separated products from

20. The estimates are from the seven- to ten-digit code level of goods.

Broda and Weinstein (2006). The estimates of Broda and Weinstein (2006) comes from the cross-elasticity of goods for newly introduced and highly disaggregated US import data.

Then, I calibrate the persistency of shocks. To calibrate persistence parameters of idiosyncratic productivity, I regress physical productivity which I identify from the COMPUSTAT data, on its lag. Then, I convert the coefficient to fit the quarterly frequency, which is 0.84. I fin dthat this value is in the range of Cooper, Haltiwanger, and Willis's (2015) estimates. For the aggregate shock persistence, I take the estimates from Smets and Wouter (2003).

Parameters Matched to the Steady State. I then calibrate the following eight parameters in Table 2 to match the moments.

TABLE II: MATCHED PARAMETERS

Parameters	Description	Value
Households		
ω	labor disutility	0.068
θ_1	degree of habit	0.310
Firms		
σ_z	idio productivity std	0.022
μ_ζ	operating cost (log mean)	-6.195
σ_ζ	operating cost (std)	4.546
α_1	advertising efficiency	0.143
α_2	advertising return to scale	0.153
Ψ	potential blueprints	0.760

The model can match the data fairly well[21]. For the labor disutility parameter, I target the value of habit-adjusted real wage to be normalized to one[22]. For the

21. Given the nonlinearity of the model, it is difficult to match the moments exactly.
22. Since I match all the moments at the same time, all parameters affect all moments.

habit parameter, I aim for the markup level from the COMPUSTAT data and match it to that of the COMPUSTAT equivalent firms in the model. The COMPUSTAT equivalent firms in the model are the top 30%[23] firms in terms of labor. For the operating cost parameters, I target the moments related to the exit rate. Two related moments are the exit rate and the 0 to 3-year survival rate. The exit rate is obtained from business dynamics statistics (BDS). The 0 to 3-year survival rate is the average of firm birth cohort data from business employment dynamics (BED). For the advertising-related parameters, I target the ratio of entrants' TFPQ estimated in Foster, Haltiwanger, and Syverson (2016) and the 0 to 2-year average employment share from BDS[24]. For the standard deviation of idiosyncratic productivity and the amount of potential goods, I target 0 to 3-year firm number share and employment share[25]. The firm number share is from BED.

TABLE III: MOMENTS USED TO MATCH PARAMETERS

Target	Data	Model
real wage	1.00[1]	1.01
markup level[2]	1.16	1.14
$\frac{\text{entrant TFPQ}}{\text{incumbent TFPQ}}$	1.02	1.00
exit rate[3]	10.73%	10.22%
0-3yr survival rate[4]	53.85%	52.80%
0-3yr firm number share[5]	31.90%	30.37%
0-3yr employment share[6]	11.24%	12.80%
0-2yr employment share[6]	8.63%	9.34%

Note: 1) Normalization
2) COMPUSTAT data to COMPUSTAT equiv. firms
3) BDS exit rate for all firms, 4) LBD average
5) BED firm numbers share, 6) BDS Employment share for firm age

Standard Deviations of the Aggregate Shock Process. To calibrate the stan-

23. The 30% estimate comes from Davis et al. (2007).
24. One can target other moments, for example, a 0-year employment share. The result is robust.
25. I chose these moments given the limited access to the data. I also tested similar moments and the results are robust.

dard deviation of the aggregate shock parameter, I simulate the model to match two moments related aggregate business cycle. To be specific, I simulate the economy for 200 quarters including 30 quarters of burn-in periods. The targets are the standard deviation of detrended log of GDP and hours worked. GDP and hours worked series are used to capture the productivity and the demand shock processes each.

TABLE IV: MOMENTS TO MATCH AGGREGATE SHOCK VOLATILITY

Parameters	Value	Target	Data	Model
σ_{TFP}	0.028	std(GDP[1])	0.033	0.022
σ_{Q}	0.021	std(N[1])	0.013	0.023

Note: 1) Detrended using quadratic time trend after log.

III. MAIN MECHANISM

In this section, I provide an analytic equation that shows the barebones of the model, the incumbent firm's markup decision. To simplify the model, I impose full depreciation of the customer base each period ($\delta = 1$)[26]. Then,the incumbent firm's markup determination equation is as follows:

$$\mu_{it} = \mu^* - \theta_1 E_t \underbrace{\frac{1}{1+r_{t,t+j}}}_{①} \underbrace{G(\zeta^*_{it})}_{②} \underbrace{\frac{mc_{it+1}}{mc_{it}}}_{③} \underbrace{\frac{c_{it+1}}{c_{it}}}_{④} \mu_{it+1} \tag{16}$$

where μ_{it} is the firm-level markup, $\mu^* = \frac{\rho}{\rho-1}$ is the markup in a standard model and $G(\zeta^*_{it+j})$ is the survival probability. The second term in the equation indicates that incumbent firms charge markups by comparing the current profit and the future value of the customer base. I explore each of these factors in detail.

26. This is akin to assuming the full depreciation of capital. In Appendix, I provide the equation without the full depreciation of the customer base.

I first document that the markup in this model is lower than in a standard model (μ^*). Since the second term is zero if I shut down the habit($\theta_1 = 0$), markup is lower than a standard model precisely due to the forward-looking term that comes from introducing habit. In other words, firms decrease their markup more than they would in the standard monopolistic competitive setting to expand the customer base. This result is slightly different from the common view that large firms charge high markup to exploit their market power. My theory shows instead that firms charge lower markup to attract more customers (Ravn, Schimitt-Grohe, and Uribe 2007). This result is due to simplifying assumption that the price elasticity of demand is constant. I find that big firms can charge higher markup than the standard markup level if the price elasticity of demand changes as a firm grows[27].

The intertemporal channel (①) is akin to that in the representative agent deep habit model (Ravn, Schimitt-Grohe, and Uribe 2006). Firms consider the future benefit of investing in a customer base when they set prices. Therefore, a change in the value of the future customer base affects the current markup decision. Relative to the standard deep habit model, this paper has an additional effect that comes from the change in firm distribution[28].

The survival probability channel (②) exists since firms' planning horizon changes as the exit probability varies (Gilchrist et al., 2017). This channel plays a key role when there is monetary policy shock. Relative to Gilchrist et al. (2017), this paper also discusses the effect of supply shock when there is customer base accumulation.

The productivity channel (③) shows that the high-productivity firms charge low markup. Since the productivity processes revert to the mean, firms want to accumulate the customer base when their marginal cost is lower than the long-run level. This condition implies that markup decreases when there is positive idiosyncratic

27. If I add an additive habit term to the demand function, which changes the price elasticity of demand, big firms can charge higher markups than the standard markup level. However, this extension requires an additional assumption that there exists a upper bound of price.

28. This is not big in the short run, however, it could be important in the medium and long run. See, for example, Sedlacek and Sterk (2017).

productivity shock[29]. This finding stands in contrast to the search-theoretic customer base models in studies such as Gourio and Rudanko (2014). In Gourio and Rudanko, the positive productivity shock increases firms' capacity to produce, but the customer base constrains the sales due to the convex adjustment cost. Thus, firms increase both prices and sales efforts. This congestion effect comes from the adjustment cost from which my model is free.

The output growth channel (④) captures the change in markup as a firm grows. At the steady state, this channel drives firm-level markup to increase since the gap between the previous period sales (the customer base) and the current sales decreases. Upon any favorable shock, the gap between the two will widens, pushing down the firm-level of markup even further.

IV. RESULTS

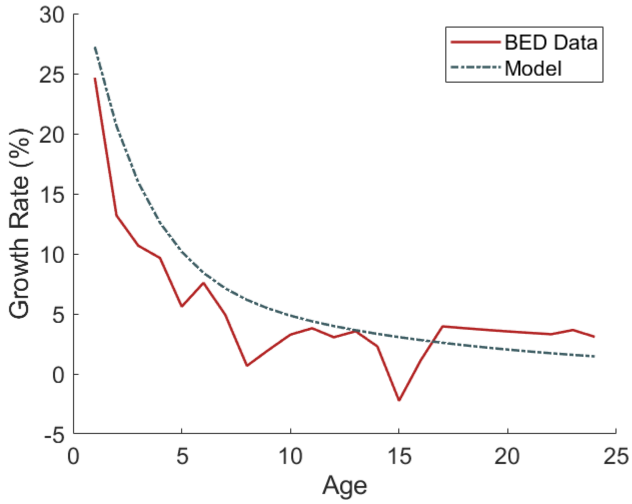
IV.A. Lifecycle Behavior of Firms

The model can match lifecycle growth speed, pricing of firms without targeting any moment related to them. Figure 6 demonstrates the model can endogenously match the labor growth rate of firms conditional on age. The green dotted line is the BED data and the red solid line is the results from the model. The growth rate of each age bin is calculated by taking an average of each cohort at the given age using all available cohorts.

To my knowledge, this is the first paper that endogenously matches the age-dependent growth speed of firms[30]. Existing literature mostly relies on supply side mechanism such as decreasing return to scale or productivity process and tends

29. For this experiment, I assume all firms in the 2nd lowest idiosyncratic productivity group to have 2nd highest productivity group. Then, I calculate the percent deviation from the initial average markup level.
30. For example, New Keynesian model with capital accumulation such as Ottonello and Winberry(2020) fails to match the growth speed of firm.

FIGURE 6: FIRM GROWTH RATE BY AGE

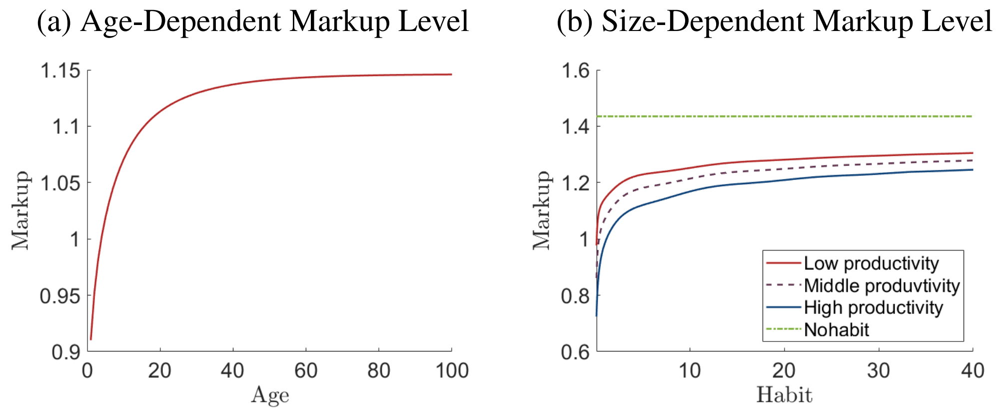


to have difficulty in endogenously matching these moments. Hence, this paper shows that demand side mechanism such as customer base accumulation is promising mechanism that can explain the age dependent growth of firms. The empirical literature such as Foster, Haltiwanger, and Syverson (2016) finds that the demand constraint is a crucial factor in slowing firm growth.

I further study the lifecycle level of markup. The left plot of figure 7 shows that the lifecycle markup is positively related to the age of the firm. Since young firms tend to be small firms, the right plot of figure 7 shows that a similar relationship exists with regard to the size of firms. Furthermore, the plot demonstrates that the more productive firms charge lower markup[31]. Foster, Haltiwanger, and Syverson (2008) show that young and high productive firms charge lower prices. Foster, Haltiwanger, and Syverson (2016) find that smaller businesses have higher productivity and lower prices than bigger firms in manufacturing industries that produce highly homogeneous goods.

31. Since high productive firms charge lower markup, high productive firms have even lower prices than low productive firms.

FIGURE 7: FIRM LEVEL MARKUP DYNAMICS



IV.B. Aggregate Dynamics

In this section, I analyze the impulse response function of the aggregate markup and GDP to one standard deviation shocks. Aggregate markup is defined as weighted average of firm-level markup using revenue as weight. Hence, I note that there is an additional distribution effect that comes from aggregation.

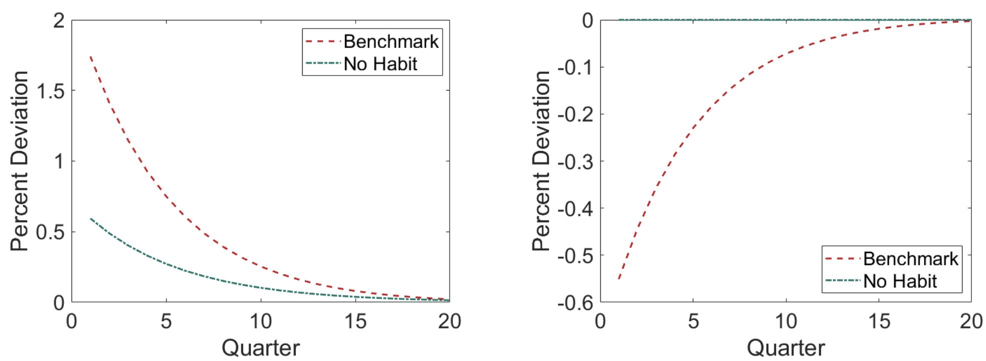
$$\mu_t = \sum_i \mu_{it} M_{it} \frac{p_{it} y_{it}}{\sum_i p_{it} y_{it}}$$

When there is an aggregate shock, distribution (M_{it}) changes and the weight ($\frac{p_{it} y_{it}}{\sum_i p_{it} y_{it}}$) changes. Due to the change in distribution, the aggregate markup varies when there is an aggregate shock even if there is no habit. To investigate the role of each channel, I present the results after shutting down habit and endogenous exit in the benchmark case. I note that there is only a distribution effect if I shut down the habit.

Aggregate Productivity Shock. I find aggregate markup is countercyclical to aggregate productivity shock. From figure 8, one can find markup goes down (panel b), and GDP goes more than the economy without habit (panel a). This is because firms think it is good time to invest in the customer base since their marginal cost is low. This mechanism is similar to Jaimovich and Floetto (2008), in which positive aggregate shock generates greater competition. However, in their model, more entry of firms causes more compeition whereas big firms charge lower markup to deter entry and push out small firms[32]. The result shows interesting contrast is to switching cost-type models that generate procyclical markup to productivity shocks (Gilbukh and Roldan 2017). In Gilbukh and Roldan, given state-contingent contracts and risk-neutral preference, markup plays only an allocative role[33]; price adjusts to transmit the effect of a shock to the customers.

FIGURE 8: AGGREGATE RESPONSE TO TFP SHOCK

(a) Change in GDP (b) Change in Markup



Monetary Policy Shock. Under the assumption that a central bank targets the real interest rate directly, I interpret bond price shock as monetary policy shock[34].

I find that the markup is countercyclical when there is a monetary policy shock. From figure 9, I find that markup is countercyclical to the monetary policy shock (panel b), and therefore, the output response is stronger (panel a). When the current demand is higher than the future demand, firms decrease markup since firms invest

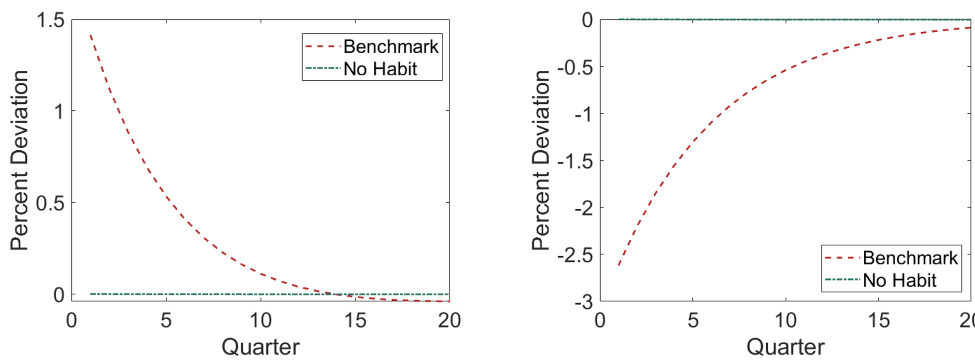
32. Exit risk dynamics are shown in Appendix.
33. In their model, the promised utility of the match, instead of markup, determines everything.
34. I note that, in the empirical section, my measure for monetary policy shock is essentially bond price shock.

in the customer base using prices. It is similar to the result in a search-theoretic customer base model such as Paciello, Pozzi, and Trachter (2019). In their work, the key mechanism that generates countercyclical markup is the incentive to increase the customer base.

FIGURE 9: AGGREGATE RESPONSE TO MONETARY POLICY SHOCK

(a) Change in GDP

(b) Change in Markup



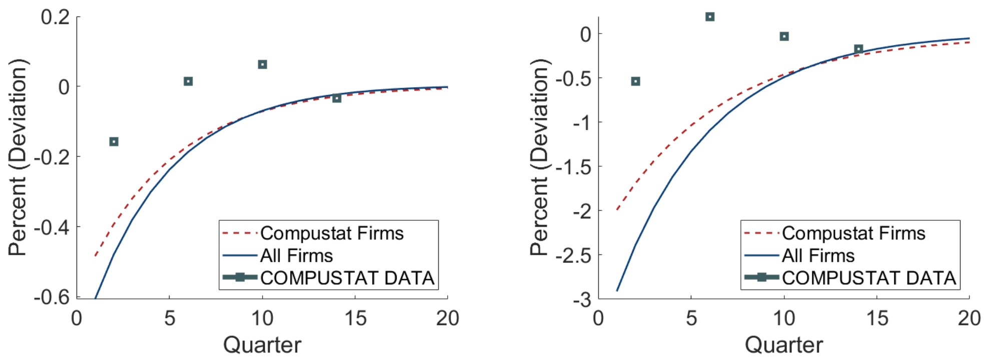
Customer Base Dependent Response. I analyze the response of markup dependent on the size of the customer base by comparing COMPUSTAT equivalent firms[35] to all firms. The aggregated markup in this part is unweighted average markup within the group to be consistent with the data.

I find that firm-level markups for COMPUSTAT firms (dotted red line) are less countercyclical to productivity and monetary policy shocks, consistent with the data. For positive productivity shock, big firms have less incentive to grow further by lowering markup. This is because the cost of lowering markup is the amount of sales quantity times the change in markup and bigger firms have larger demand base. For positive monetary policy shocks, the survival probability channel is the underlying mechanism. Since the exit risk of small firms increases more than that

35. COMPUSTAT equivalent firms are large firms in the model in terms of labor following Davis et al.'s (2007) estimates.

FIGURE 10: CUSTOMER BASE DEPENDENT MARKUP RESPONSES

(a) Response to TFP shock (b) Response to MP shock



of big firms, small firms charge even lower markup than big firms[36].

V. EXTENSION

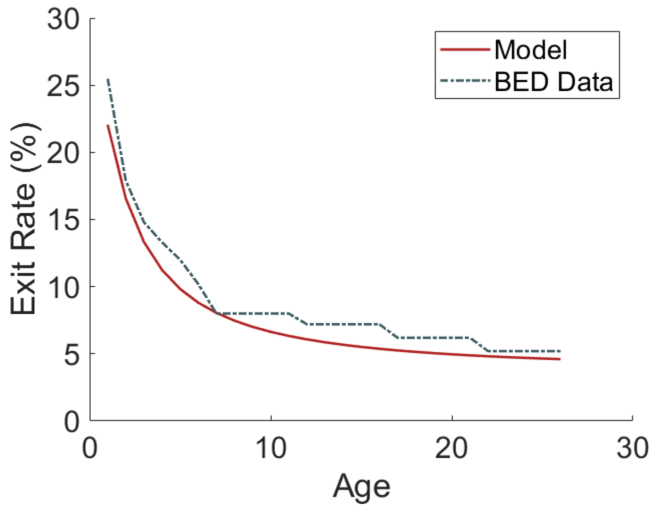
In this section, I show that the model can match the lifecycle exit rate if there is a small amount of exogenous exit. Following the empirical result of Foster, Haltiwanger, and Syverson (2008) that emphasizes the importance of the demand factor, I attempt to show how the inclusion of the demand factor can enhance my understanding of the age-dependent exit rate. For the result of this subsection, I assume that one percent of incumbent firms randomly exit from the market[37].

I find from figure 11 that the model closely matches the age-dependent exit rate of firms[38]. The green dotted line is the BED data and the red solid line is the results from the model. The model closely tracks the data including the initial high rate of exit. This experiment shows that the demand factor can significantly improve the

36. If I include exogenous exit and shut down the endogenous exit, the size-dependent markup response to monetary policy shock closes down. The result can be provided upon request.
37. This assumption is to match the exit rate of big and old firms.
38. Furthermore, all the results presented in the paper hold for the new calibration.

understanding of the lifecycle firm exit rate, especially for the initial stage of firm growth.

FIGURE 11: EXIT RISK OF FIRMS



VI. CONCLUSION

This paper provides a new angle for understanding cyclical pricing behavior and lifecycle behavior of individual firms. Exploiting the new dataset and state-of-the-art computational techniques, the paper claims that focusing on micro-level response can deliver some interesting results.

The result that firm-level markup is countercyclical to aggregate productivity and monetary shocks challenges the shock amplification mechanism of the existing models. Therefore, I propose a firm dynamics model with an emphasis on demand accumulation. The model shows that the tradeoff between invest and harvest motives in the customer base plays a crucial role in generating empirical evidence.

The paper emphasizes the importance of the underlying economic distribution. I show, from data and the model, that the markup of a firm with a lower demand base responds more strongly to aggregate shocks. This finding is due to the size-dependent sensitivity in the value of the customer base. If this is true, the amplification of the aggregate shocks would differ across economies depending on the underlying firm distribution.

The model demonstrates that the demand accumulation mechanism can endogenously match the lifecycle growth and exit rates of firms. It adds additional reason to seriously study the firm's demand accumulation mechanism. Moreover, the model can generate a positive relationship between markup and size.

More broadly, this paper emphasizes the role of intangible capital in understanding the behavior of firms. Despite progress in this paper and other literature, identifying intangible capital empirically and understanding what it does theoretically remain important future research topics.

TABLES

TABLE V: REGRESSION TABLE OF TFP SHOCK

	(1) Markup	(2) F.Markup	(3) F2.Markup	(4) F3.Markup
TFP	-0.158***	0.014	0.062***	-0.033
	(0.016)	(0.019)	(0.019)	(0.020)
L.Markup	0.601***	0.403***	0.320***	0.235***
	(0.010)	(0.013)	(0.015)	(0.017)
L.PROD	-0.138***	-0.239***	-0.136***	-0.146***
	(0.013)	(0.020)	(0.024)	(0.028)
L.SALE	-0.024***	-0.023***	-0.016***	-0.015***
	(0.002)	(0.002)	(0.002)	(0.003)
L.MS	-0.068***	-0.116***	-0.111***	-0.118***
	(0.019)	(0.029)	(0.036)	(0.039)
L.AGE	0.000***	0.001***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
L.SG&A	0.011***	0.009***	0.004*	0.004*
	(0.001)	(0.002)	(0.002)	(0.002)
L.GDP	0.027***	0.090***	0.144***	0.163***
	(0.008)	(0.012)	(0.014)	(0.015)
N	192,218	174,881	159,462	146,337

Note: * p<0.1, ** p<0.05, *** p<0.01, standard errors in parentheses

TABLE VI: REGRESSION TABLE OF MP SHOCK

	(1) Markup	(2) F.Markup	(3) F2.Markup	(4) F3.Markup
MP	-0.536***	0.198*	-0.028	-0.171*
	(0.087)	(0.103)	(0.105)	(0.102)
L.Markup	0.541***	0.342***	0.205***	0.127***
	(0.014)	(0.017)	(0.021)	(0.023)
L.PROD	-0.123***	-0.331***	-0.238***	-0.028
	(0.023)	(0.032)	(0.032)	(0.033)
L.SALE	-0.027***	-0.023***	-0.019***	-0.017***
	(0.002)	(0.003)	(0.004)	(0.004)
L.MS	-0.117***	-0.169**	-0.206**	-0.278***
	(0.043)	(0.070)	(0.088)	(0.097)
L.AGE	0.001***	0.000**	-0.000**	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
L.SG&A	0.012***	0.010***	0.008***	0.010***
	(0.002)	(0.003)	(0.003)	(0.004)
L.GDP	0.096***	0.284***	0.415***	0.317***
	(0.017)	(0.025)	(0.029)	(0.030)
N	110,226	99,783	90,571	82,973

Note: * $p<0.1$, ** $p<0.05$, *** $p<0.01$, standard errors in parentheses

REFERENCES

Ackerberg, D. A., Caves, K., and Frazer, G., ”Identification properties of recent production function estimators”, Econometrica, 83 (2015), 2411-2451.

Arkolakis, Costas, ”Market Penetration Costs and the New Consumers Margin in International Trade”, Journal of Political Economy, 118 (2010), 1151-1199

Attanasio, Orazio P. and Guglielmo Weber, ”Consumption Growth, the Interest Rate and Aggregation” The Review of Economic Studies, Vol. 60, (1993), 631-649

Autor, David, David, Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen, ”The Fall of the Labor Share and the Rise of Superstar Firms”, The Quarterly Journal of Economics, Volume 135, Issue 2, May (2020), Pages 645709

Baqaee, David Rezza and Emmanuel Farhi, ”Productivity and Misallocation in General Equilibrium”, The Quarterly Journal of Economics, Volume 135, Issue 1, February 2020, Pages 105163

Bayer, Christian and Ralph Luetticke, ”Solving heterogeneous agent models in discrete time with many idiosyncratic states by perturbation methods”, Quantitative Economics, Volume 11, Issue4, November (2020) Pages 1253-1288

Basu, Susanto, Discussion on ”Resurrecting the Role of the Product Market Wedge in Recessions” by Bils Mark, Peter Klenow, and Benjamin Malin, 2014

Bils, Mark. The Cyclical Behavior of Marginal Cost and Price. American Economic Review, 77, (1987), 83855.

Bils, Mark, and James A. Kahn, What Inventory Behavior Tells Us About Business Cycles. American Economic Review, 90, (2000), 45881.

Bils, Mark, Peter J. Klenow, and Benjamin A. Malin, ”Resurrecting the Role of the Product Market Wedge in Recessions”, American Economic Review, 108 (2018), 11181146

Blundell, Richard and Stephen Bond, ”Initial conditions and moment restrictions in dynamic panel data models”, Journal of Econometrics 87, (1998), 115143

Broda, Christian and David E. Weinstein, ”Globalization and the gains from variety”, Quarterly Journal of Economics, Volume 121, Issue 2, (2006)

Bernanke, Ben S., Mark Gertler and Simon Gilchrist, ”The financial accelerator in a quantitative business cycle framework”, Handbook of Macroeconomics, Volume 1, (1999), Edited by J.B. Taylor and M. Woodford

Campbell, Jeffrey R., Charles L. Evans, Jonas D. M. Fisher, and Alejandro Justiniano, (2012), Macroeconomic Effects of Federal Reserve Forward Guidance, Brookings Papers on Economic Activity, 180

Christiano, Lawrence J., Martin Eichenbaum and Charles L. Evans, "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy", Journal of Political Economy, Vol. 113, No. 1 (February), (2005), pp. 1-45

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.", manuscript (2014)

Collard-Wexler, Allan and Jan De Loecker, "Production Function Estimation with Measurement Error in Inputs", NBER Working Paper No. 22437, (2016)

Cooper, Russell & Haltiwanger, John & Willis, Jonathan L., "Dynamics of labor demand: Evidence from plant-level observations and aggregate implications," Research in Economics, Elsevier, vol. 69(1), (2015), pages 37-50.

Crouzet, Nicolas, and Neil R. Mehrotra. 2020. "Small and Large Firms over the Business Cycle." American Economic Review, 110 (11): 3549-3601.

Davis, S. J., J. Haltiwanger, R. Jarmin, and J. Miranda, 2007, Volatility and dispersion in business growth rates: Publicly traded versus privately held firms, in NBER Macroeconomics Annual, Volume 21, (2006), pp. 107180. MIT Press

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda., "The Role of Entrepreneurship in US Job Creation and Economic Dynamism." Journal of Economic Perspectives, 28 (3) : 3-24, (2014)

De Loecker, Jan and Jan Eeckhout, "The rise of market power and the macroeconomic implications", NBER working paper 23687, (2017)

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger, "The rise of market power and the macroeconomic implications", The Quarterly Journal of Economics, Volume 135, Issue 2, May 2020, Pages 561644

De Loecker, Jan and Warzynski, "Markups and Firm-level Export Status", The American Economic Review, Vol. 102, No. 6., (2012), 2437-2471.

Dinlersoz, Emin M. and Mehmet Yorukoglu, "Information and Industry Dynamics", American Economic Review VOL. 102, NO. 2, APRIL 2012 (pp. 884-913)

Drozd, A. Lukasz and Jaromir B. Nosal, "Understanding International Prices: Customers as Capital", American Economic Review, VOL. 102, NO. 1, FEBRUARY, 2012, (pp. 364-95)

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson, The Growth and Failure of U.S. Manufacturing Plants, The Quarterly Journal of Economics, Volume 104, Issue 4, November 1989, Pages 671698

Fernald, John G., Productivity and Potential Output Before, During, and After the Great Recession. NBER Macroeconomics Annual 2014, Vol. 29(1), pp. 1-51.

Fieldhouse, Andrew, Karel Mertens, and Morten Ravn, ”The Macroeconomic Effects of Government Asset Purchases: Evidence from Postwar US Housing Credit Policy”, The Quarterly Journal of Economics, Volume 133, Issue 3, August 2018, Pages 15031560

Fitzgerald, D., Haller, S. and Yedid-Levi, Y., ”How exporters grow.” Working Paper, Federal Reserve Bank of Minneapolis, Staff Report 524, 2017

Foster, Lucia, John Haltiwanger, and Chad Syverson, ”Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?”, American Economic Review 2008, 98:1, pp. 394425

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

Gali, Jordi and Thijs van Rens, ”The Vanishing Procyclicality of Labor Productivity”, The Economic Journal, Volume 131, Issue 633, January (2021), Pages 302326,

Gandhi, Amit, Salvador Navarro, and David A. Rivers, ”On the Identification of Gross Output Production Functions”, Journal of Political Economy Volume 128, Number 8, August, 2020

Gertler, Mark and Simon Gilchrist, ”Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms”, The Quarterly Journal of Economics, Volume 109, Issue 2, May 1994, Pages 309340,

Gertler, Mark, and Peter Karadi, Monetary Policy Surprises, Credit Costs, and Economic Activity, American Economic Journal: Macroeconomics, 7, 2015, 4476.

Gilbukh, Sonia and Pau Roldan, ”Firm dynamics and pricing under customer capital accumulation”, Journal of Monetary Economics, Journal of Monetary Economics, Volume 118, March (2021), Pages 99-119

Gilchrist, Simon, Raphael Schoenle, Jae Sim, and Egon Zakrajsek, Inflation Dynamics During the Financial Crisis, American Economic Review, 107(3): 785823, 2017

Gourio, Francois and Leena Rudanko, ”Customer Capital”, Review of Economic Studies, 81, 3, 2014

Gurkaynak, Refet S., Brian Sack, and Eric T. Swanson, Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements, International Journal of Central Banking, 1, 5593, 2005

Hall, R. E., ”Market structure and macroeconomic fluctuations.” Brookings papers on economic activity, 2, 285-322, 1986

Hall, R. E., ”What the Cyclical Response of Advertising Reveals about Markups and other Macroeconomic Wedges”, manuscript, 2014

Hong, Sungki, ”Customer Capital, Markup Cyclicality, and Amplification”, 2019 Meeting Papers 959, Society for Economic Dynamics.

Hopenhayn, Hugo A., ”Entry, Exit, and firm Dynamics in Long Run Equilibrium”, Econometrica, Vol. 60, No. 5 (Sep., 1992), pp. 1127-1150

Jaimovich, Nir and Max Floetotto, ”Firm dynamics, markup variations, and the business cycle”, Journal of Monetary Economics 55, 2008, 12381252

Jorda, O. ”Estimation and Inference of Impulse Responses by Local Projections.” American Economic Review (2005), 95.1: 161182.

Jorda, O, M. Schularick, and A. M. Taylor. ”Sovereigns Versus Banks: Credit Crises and Consequences.” Journal of the European Economic Association (2016), 14(1): 45-79.

Karabarbounis, Loukas and Brent Neiman, ”Accounting for Factorless Income.” NBER Macroeconomics Annual, 33, April 2018.

Kuttner, Kenneth N, ”Monetary policy surprises and interest rates: Evidence from the Fed funds futures market”, Journal of Monetary Economics Volume 47, Issue 3, 2001, Pages 523-544

Levinsohn, J., A and A. Petrin, Estimating production functions using inputs to control for unobservables, The Review of Economic Studies, 70(2), 2003, 317341.

Luttmer, Erzo G.J., ”Selection, Growth, and the Size Distribution of Firms,” Quarterly Journal of Economics, 2007, Vol. 122, No. 3, 1103-1144.

Luttmer, Erzo G.J., ”On the Mechanics of Firm Growth,” The Review of Economic Studies, Vol. 78, No. 3, 2011, 1042-1068

Nakamura, Emi and Jon Steinsson, ”Price Setting in Forward-Looking Customer Markets”, Journal of Monetary Economics, 58(3), 220-233, April 2011

Nakamura, Emi and Jon Steinsson, ”High Frequency Identification of Monetary Non-Neutrality: The Information Effect” Quarterly Journal of Economics, 133(3), 2018, 1283-1330.

Nekarda, Christopher J. and Valerie A. Ramey, ”The Cyclical Behavior of the Price-Cost Markup”, Journal of Money, Credit, and Banking, Volume52, Issue 2, December 2020, Pages 319-353

Olley, G. S., and A. Pakes, The Dynamics of Productivity in the Telecommunications Equipment Industry, Econometrica: Journal of the Econometric Society, 1996, pp. 12631297

Ottonello, Pablo, and Thomas Winberry, ”Financial Heterogeneity and the Investment Channel of Monetary Policy”, Econometrica, Vol. 88, No. 6, November (2020), 2473-2502

Paciello, L., A.Pozzi and N.Trachter, ”Price Dynamics with Customer Markets”, The International Economic Review, 2019, Volume 60, Issue 1

Phelps, E. S. and Winter, S. G., Optimal Price Policy under Atomistic Competition, in E. S. Phelps (ed.) Microeconomic Foundations of Employment and Inflation Theory (New York: W. W. Norton), (1970), 309337.

Pugsley, Benjamin W., Petr Sedlacek, and Vincent Sterk, ”The Nature of Firm Growth”, AMERICAN ECONOMIC REVIEW, VOL. 111, NO. 2, FEBRUARY 2021

Ravn, Morten, Stephanie Schmitt-Grohe, and Martin Uribe, ”Deep habits”, Review of Economic Studies, vol. 73(1), 2006

Ravn, Morten, Stephanie Schmitt-Grohe, and Martin Uribe, Firm Dynamics with Customer Markets, manuscript, 2007

Reiter, M., ”Solving heterogeneous-agent models by projection and perturbation.” Journal of Economic Dynamics and Control, 33(3):649-665, 2009

Romer, Christina D and David H Romer, A New Measure of Monetary Shocks: Derivation and Implications, The American Economic Review, 94 (4), 10551084, 2004

Roodman, D., ”How to do xtabond2: An introduction to difference and system GMM in Stata”. Stata Journal 9(1): 86-136, 2009.

Rotemberg, J. J. and Woordford, M., Markups and the Business Cycle, in O. J. Blanchard and S. Fischer (eds.) NBER Macroeconomics Annual (Cambridge, MA: MIT Press) 63129, 1991

Rotemberg, J. J. and Woordford, M., Dynamic General Equilibrium Models with Imperfectly Competitive Product Markets, in T. F. Cooley (ed.) Frontiers of Business Cycle Research (Princeton, NJ: Princeton University Press) 243293, 1995

Schmitt-Grohe, S. and Uribe, M., ”Solving dynamic general equilibrium models using a second-order approximation to the policy function.” Journal of Economic Dynamics and Control, 28(4):755-775, 2004

Smets, Frank and Raf Wouters, ”An estimated dynamic stochastic general equilibrium model of the EURO area”, Journal of the European Economic Association, September, 1(5):11231175, 2003

Smets, Frank and Raf Wouters, ”Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach”, American Economic Review VOL. 97, NO. 3, JUNE 2007 (pp. 586-606)

Sedlacek, Petr and Vincent Sterk, ”The Growth Potential of Startups over the Business Cycle”, American Economic Review, Vol. 107 (10), pp. 3182-3210, 2017

Winberry, T., ”A Method for Solving and Estimating Heterogeneous Agent Macro Models.”, Quantitative Economics, 9(3), 2018

A. DATA CLEANING

I follow De Loecker, Eeckout, and Unger (2020) to clean the data. Specifically, I use the firm-level financial variables of all US-listed public firms from Wharton Research Data Services (WRDS). The sample period is from 1950 to 2017, and I allow entry and exit within the period. I use the industry format and eliminate the firms that do not report the NAICS industry code. Firms without key variables to estimate the production function (sales, cost of goods, and capital) are excluded from the sample. Additionally, I eliminate the firms with higher than a 99th percentile and first percentile of labor cost share, where the percentiles are calculated for each year. I deflate all variables with GDP deflator.

B. PRODUCTION FUNCTION ASSUMPTIONS

In this section, I show how Leontief gross output production function translates to Cobb-Douglas function. I assume industry-specific Leontief gross production function that the output is proportional to the intermediate input use.

$$Q_{it} = \min\{(L_{it}^{\theta_j^L} K_{it}^{\theta_j^K})\exp(\omega_{it}), \alpha_M M_{it}\}$$

where i denotes individual firms, j denotes industry, ω_{it} is idiosyncratic productivity. At the optimum, I have

$$Q_{it} = (L_{it}^{\theta_j^L} K_{it}^{\theta_j^K})\exp(\omega_{it}) = \alpha_M M_{it}$$

Since it is hard to find appropriate intermediate input in COMPUSTAT, I use

$$Q_{it} = (L_{it}^{\theta_j^L} K_{it}^{\theta_j^K})\exp(\omega_{it})$$

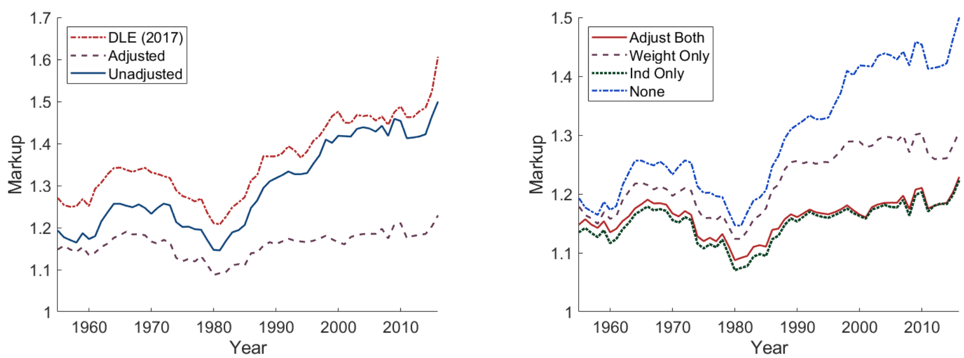
C. MARKUP TREND

In this section, I investigate the discussion related to the measurement error raised by Karabarbounis and Neiman (2018). The left panel of figure A.1 shows that I have difficulty in replicating their results leveraging on the code of De Loecker and Warzynski (2012). However, I do not take stance in the markup trend since I find the rise of markup with dynamic panel approach (Blundell and Bond, 1998)[39].

A.1: MARKUP TREND

(a) Markup Trend

(b) Decomposition



Note: The left panel uses two moment conditions whereas the right panel uses the condition related to labor only.

The measurement error affects aggregate markup in two aspects. First, the measurement error distorts the inverse of the cost-share of productive firms. Given the industry-specific production function, the inverse of the cost share of a high-measurement-error firm is larger in the first stage. Therefore, the markup of high-measurement-error firms is inflated ($\mu_{it} \equiv \frac{P_{it}}{\Lambda_{it}} = \theta_{it}^{v}\frac{P_{it}Q_{it}}{P_{it}^{V}V_{it}\exp(\varepsilon_{it})}$). However, the green dotted line in the right panel of figure A1 shows that the first channel itself is not important since low-measurement-error firms cancel each other. Second, the weights on high-measurement-error firms are different. Since I use the sales-weighted average to aggregate markup, the sales of high-measurement-error firms become larger

39. This result is available upon request.

if I do not adjust for the measurement error.

$$\mu_t = \sum_i^N \mu_{it} \frac{P_{it} Q_{it}}{\exp}(\varepsilon_{it}) / (\sum_i^N \frac{P_{it} Q_{it}}{\exp}(\varepsilon_{it}))$$

The purple dashed line in the right figure of figure A.1 illustrates that this weight channel accounts for half of the increase. The other half of the increase falls on the multiplicative effect of two channels.

D. Robustness Tests

This subsection provides the robustness of my results. The robustness tests are executed in four aspects as illustrated below. The results are generally robust.

Detrending. To test the effect of the detrending method, I use the first difference. By using the first difference, I can take out firm-specific trend; hence I can test the potential bias both from a quadratic trend and from an industry-specific time trend. To test, I take the first difference of the data and use ordinary least squares (OLS) to estimate the following model.

$$\triangle \text{Markup}_{i,t+h} = \gamma_1 \text{Shock}_t + \gamma_2 \triangle \text{GDP}_{i,t-1} + \gamma_3 \text{Control}_{i,t-1} + \gamma_i + e_{it}$$

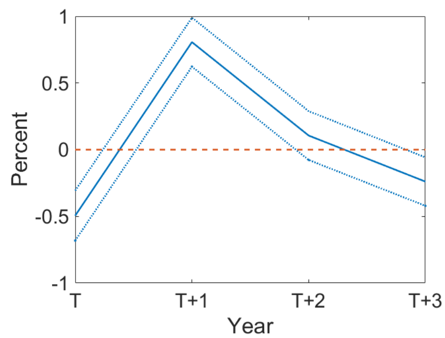
As in the main regression, the controls are sales, market share, productivity, sales effort, and age. Figure A.2 shows that the results are robust, although the size dependent response to aggregate TFP shock is similar across all sizes.

To test the detrending method further, I include the industry-specific time trend rather than detrending each variable directly. Specifically, I include a quadratic industry specific trend term instead of detrending each variable, and the control variables are the same as above. In this approach, variables share the common trend whereas each variable has its own industry trend in the main results.

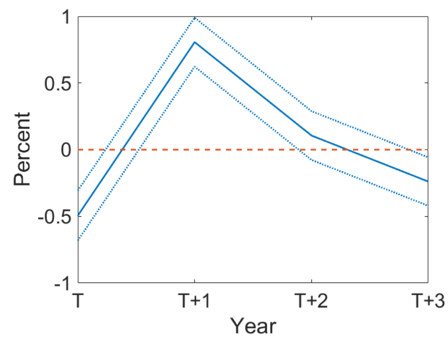
A.2: MARKUP RESPONSE TO AGGREGATE SHOCKS (FIRST DIFFERENCE)

Average Response[1]

(a) TFP shock

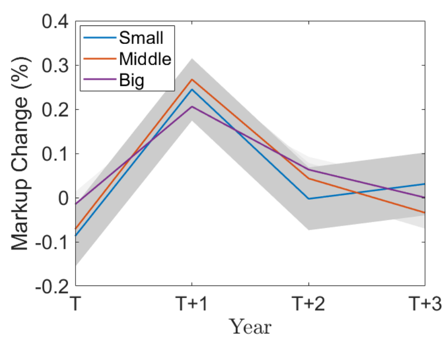


(b) Monetary Policy shock

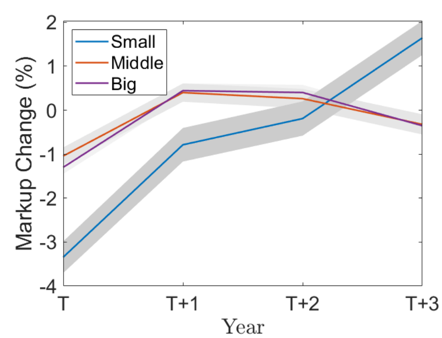


Size Dependent Response[2]

(a) TFP shock



(b) Monetary Policy shock



Note: 1) Solid lines are firm-level responses and dotted lines are 95th percentile confidence band.
2) Solid lines are firm-level responses and shaded areas are 68th percentile confidence band.

$$\text{Markup}_{i,t+h} = \gamma_1^h \text{Shock}_t + \gamma_2^h \text{Markup}_{i,t-1} + \gamma_3^h \text{Control}_{i,t-1} + \gamma_i^h \text{F(trend)} + e_{it+h}$$

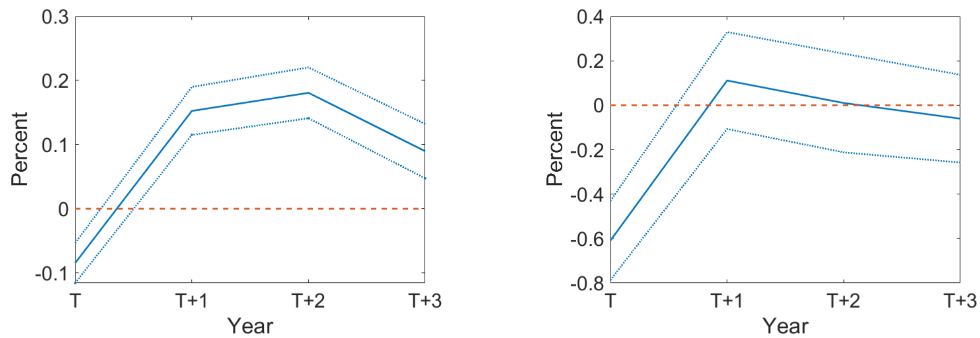
Figure A.3 shows that the results are generally robust for both average and size dependent responses. I note that the results are robust to cubic time trend.

Production Function. I test the robustness to specification and estimation of the production function. To test the specification of the production function, I set

A.3: MARKUP RESPONSE TO AGGREGATE SHOCKS (INDUSTRY TREND)

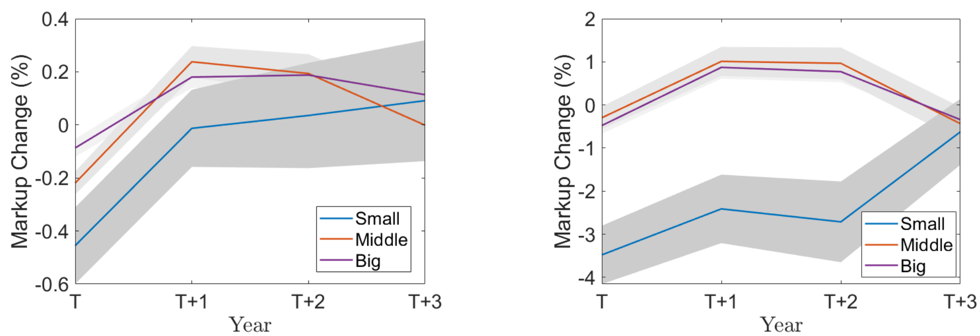
Average Response[1]

(a) TFP shock (b) Monetary Policy shock



Size Dependent Response[2]

(a) TFP shock (b) Monetary Policy shock



Note: 1) Solid lines are firm-level responses and dotted lines are 95th percentile confidence band.
2) Solid lines are firm-level responses and shaded areas are 68th percentile confidence band.

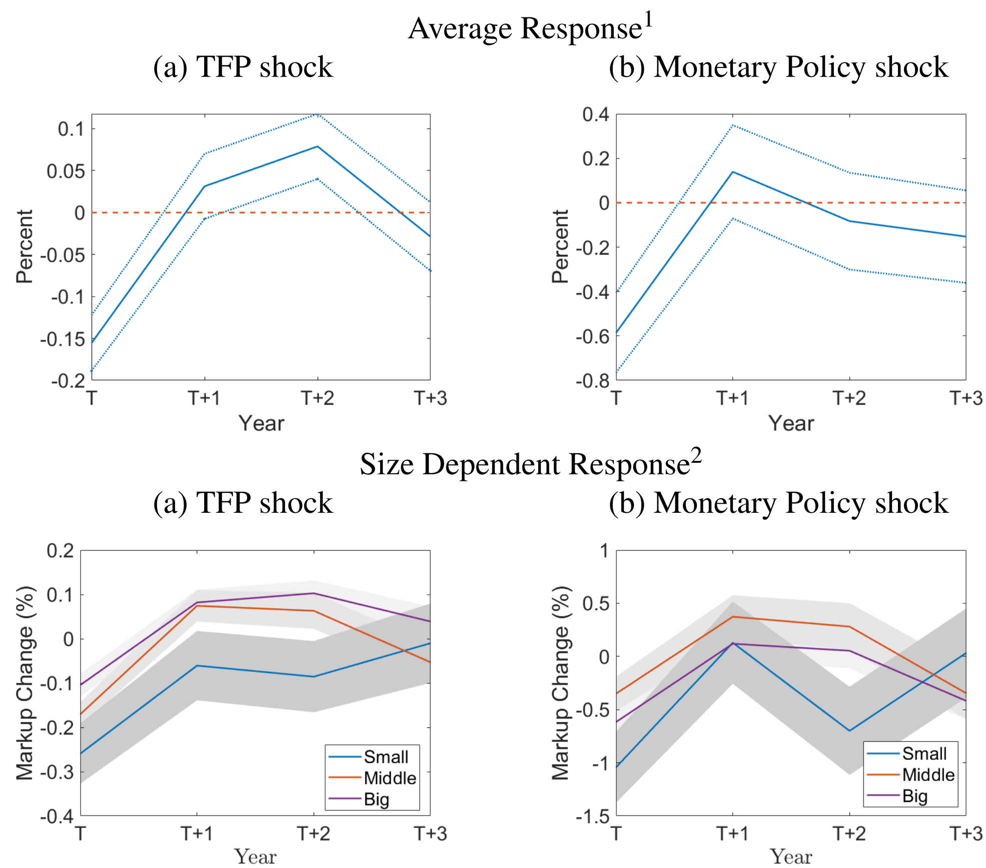
a flexible production function, i.e., the translog production function. I approximate the function with second-order, and I do not include the interaction term of labor and capital due to the possible measurement error of the capital. A detailed discussion can be found in Collard-Wexler and De Loecker (2016). In sum, I regress the following equation using local projection.

$$Q_{it} = F(L_{it}, K_{it})\exp(\omega_{it})$$

$$\tilde{q}_{it} = \theta_{jt}^{v1} l_{it} + \theta_{jt}^{v2} l_{it}^2 + \theta_{jt}^{k1} k_{it} + \theta_{jt}^{k2} k_{it}^2 + \omega_{it} + \varepsilon_{it}$$

Figure A.4 illustrates that the impulse responses are very similar to those in the main results.

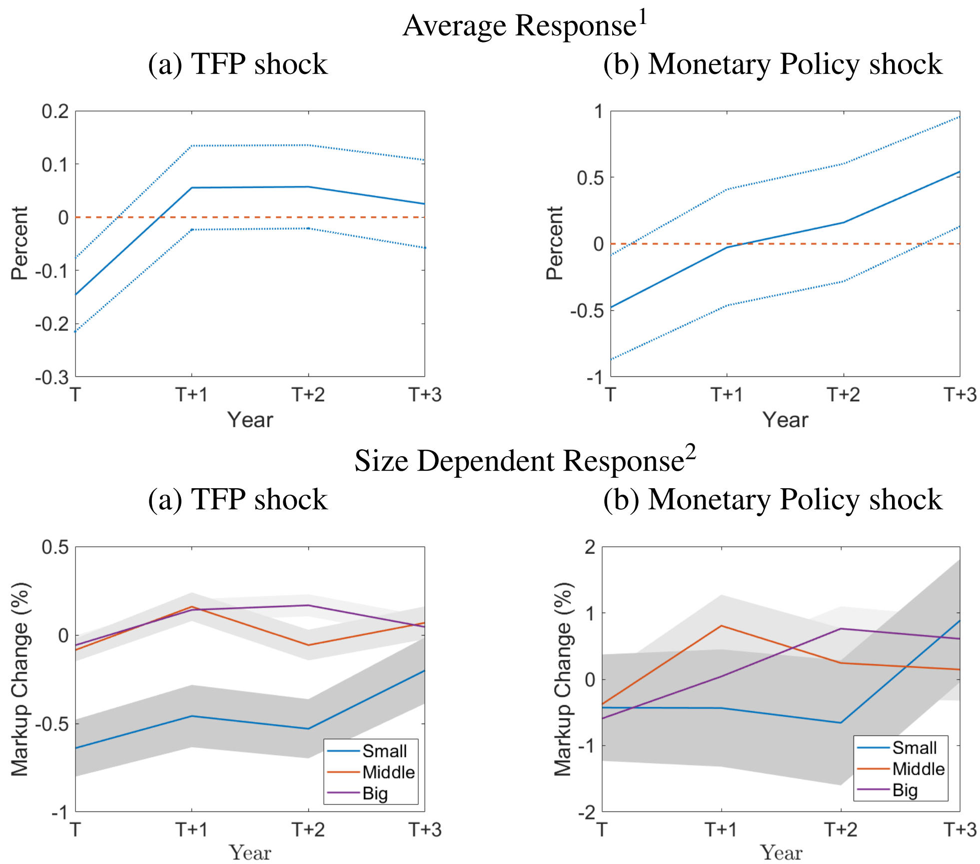
A.4: MARKUP RESPONSE TO AGGREGATE SHOCKS (TRANSLOG)



Note: 1) Solid lines are firm-level responses and dotted lines are 95$^{\text{th}}$ percentile confidence band.
2) Solid lines are firm-level responses and shaded areas are 68$^{\text{th}}$ percentile confidence band.

To deal with the identification issue raised in Gandhi, Navarro, and Rivers (2020), I estimate the production function with a dynamic panel data method (Blundell and Bond, 1998) that allows autocorrelation in the error term. Figure A.5 shows that most impulse responses are similar to the main responses. However, there is no significant difference among different size groups in the response to aggregate monetary policy shocks.

A.5: MARKUP RESPONSE TO AGGREGATE SHOCKS (BLUNDELL-BOND)



Note: 1) Solid lines are firm-level responses and dotted lines are 95th percentile confidence band.
2) Solid lines are firm-level responses and shaded areas are 68th percentile confidence band.

Production Approach. I further execute two tests related to the production approach. The first regression is related to the concern raised in Karabarbounis and Neyman (2018) that not adjusting the measurement error can generate a significant difference in studying the markup trend[40].

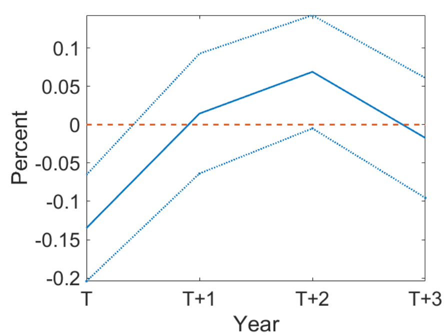
40. I further suspect that any estimation method would give robust results as long as the coefficients are fixed over time since the output elasticity is simply a scaling of the labor cost expenditure ratio.

Since I detrend variables, the difference in the trend may not be an issue, and indeed, figure A.6 illustrates that the results are robust to this margin. However, there is no significant difference among size groups in response to the aggregate monetary policy shocks.

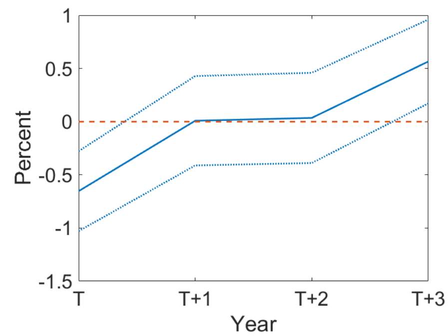
A.6: MARKUP RESPONSE TO AGGREGATE SHOCKS (NO MEASUREMENT ERROR)

Average Response[1]

(a) TFP shock

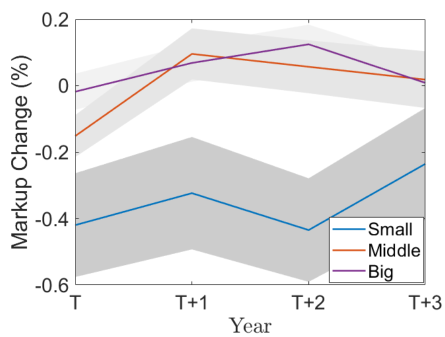


(b) Monetary Policy shock

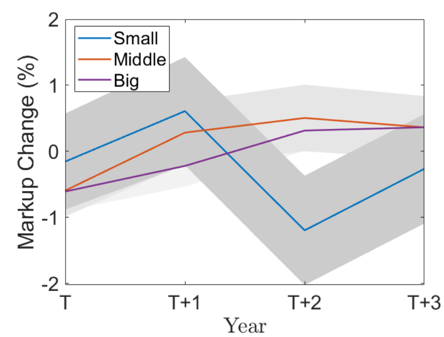


Size Dependent Response[2]

(a) TFP shock



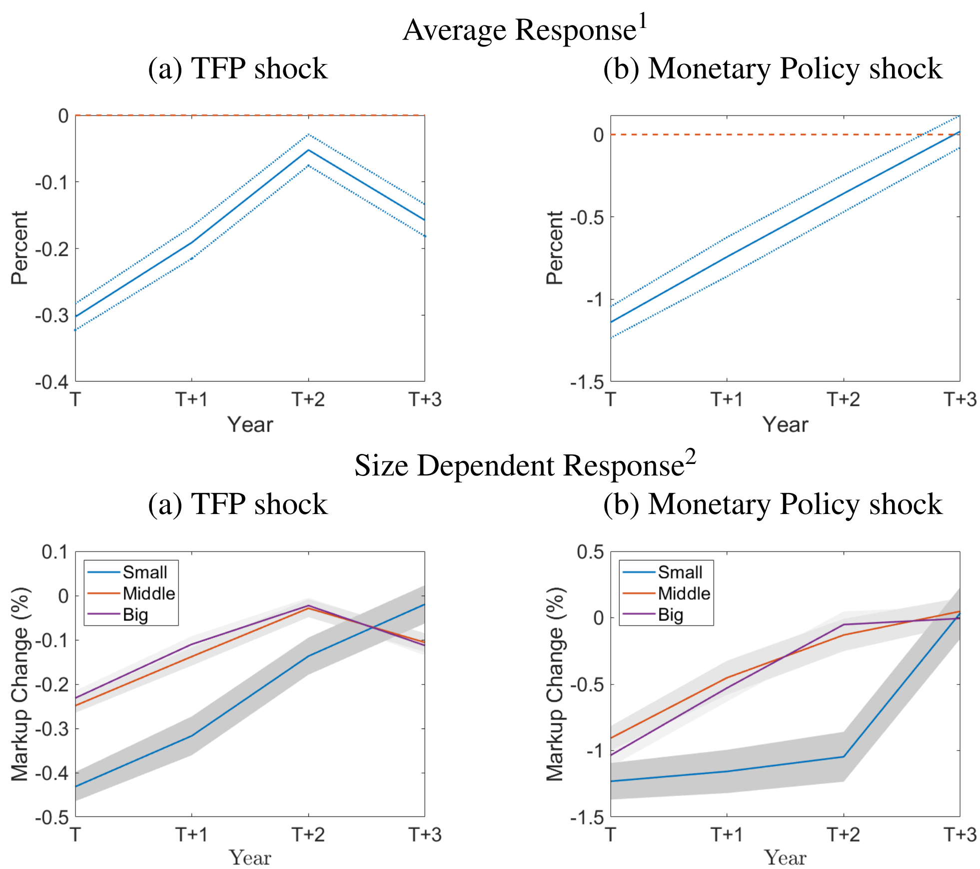
(b) Monetary Policy shock



Note: 1) Solid lines are firm-level responses and dotted lines are 95$^{\text{th}}$ percentile confidence band.
2) Solid lines are firm-level responses and shaded areas are 68$^{\text{th}}$ percentile confidence band.

Another concern regarding the production approach is the data choice for the variable cost. Traina (2018) claims that it is important to include "Selling, General and Administrative (SG&A)" costs in the variable cost for the markup trend. If I set variable cost as the sum of SG&A and cost of goods sold, there is no significant change in the markup trend. As I show above, the change in the trend does not affect the response at the business cycle frequency.

A.7: MARKUP RESPONSES TO DIFFERENT VARIABLE COST



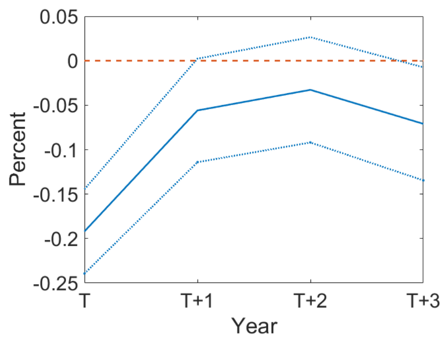
Note: 1) Solid lines are firm-level responses and dotted lines are 95th percentile confidence band.
2) Solid lines are firm-level responses and shaded areas are 68th percentile confidence band.

Local Projection. Last, I check whether the shock is exogenous from other shocks. Specifically, I include the two shocks in one regression and check the coefficients. Figure A8 shows that the results are almost identical to the main impulse responses. Therefore, the shocks I use are independent of each other.

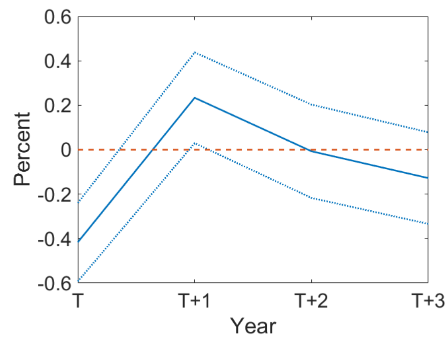
A.8: Markup Responses (All Shocks)

Average Response[1]

(a) TFP shock

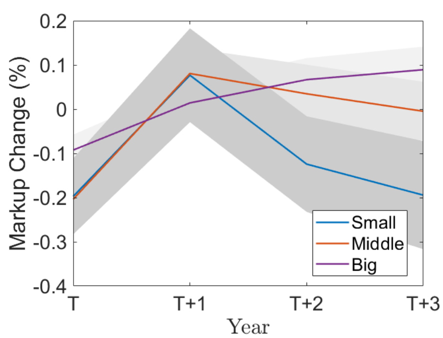


(b) Monetary Policy shock

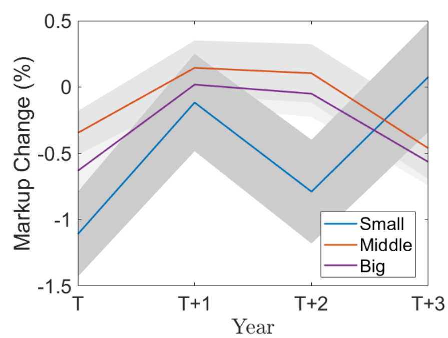


Size Dependent Response[2]

(a) TFP shock



(b) Monetary Policy shock



Note: 1) Solid lines are firm-level responses and dotted lines are 95th percentile confidence band.

2) Solid lines are firm-level responses and shaded areas are 68th percentile confidence band.

E. Computation Approach

To solve the model, I first find the steady state of the model and use the first-order perturbation to analyze the aggregate dynamics (Reiter, 2009).

To solve the model at the steady state, I discretize the state space. I choose seven grids for productivity and 100 grids for habit. The productivity grids are chosen to be equally distanced within the range of three standard deviations to both sides. For habit, I ensure that firms' choice is within the bound, and the grids are chosen to be exponentially distanced.

I use the following procedure to solve the model[41]. First, I guess the aggregate habit stock ($\tilde{C}$). Second, I solve the incumbent firm's problem. To solve the problem, I first approximate the value function by using the Chebyshev polynomial for computational efficiency.

$$V(z_i, h_{i,-1}; F) = \sum_{a=1}^{n_z} \sum_{b=1}^{n_h} \theta^v_{a,b} T_a(z) T_b(h_{-1})$$

With the approximated value function, I find the habit choice.

$$h^* = \arg\max_h \{\mu(\frac{p_i y_i - W n_i}{\tilde{P}}) + \max_{\text{exit,stay}} [0, -\frac{\zeta}{\tilde{P}} + \beta e^Q EV'(z_i', h_i; \mathrm{F}')]\}$$

Then, I iterate the policy function many times to find the value function. I iterate the obtained value function until it converges. Then, I approximate the value function and the habit choice function by using the Chebyshev polynomials. With the approximated value function, I solve the entrant's problem by using the approximated value function. I update the distribution and iterate until the aggregate habit-adjusted consumption ($\tilde{C}$) converges. I use collocation to approximate the Bellman equation and Gauss-Hermite quadrature to evaluate the expectation concerning idiosyncratic shocks.

To find the dynamics to aggregate shocks, I use a projection and perturbation approach (Reiter, 2009)[42]. Let

$$V(z_i, h_{i,-1}; F) = \sum_{a=1}^{n_z} \sum_{b=1}^{n_h} \theta^v_{a,b} T_a(z) T_b(h_{-1})$$

41. I leverage some of the routine from Winberry (2016) to compute the steady state.

42. I modify some of the routine from Bayer and Luetticke (2020) to compute aggregate dynamics.

$$V(z_i', h_i; F') = \sum_{a=1}^{n_z} \sum_{b=1}^{n_h} \theta_{a,b}^{v'} T_a(z') T_b(h)$$

I can then write the system of equations in a Schmitt-Grohe and Uribe (2004) form:

$$E_{F_t}[f(X_t, X_{t-1}, Y_t, Y_{t-1})] = 0$$

where $X = \{V, \tilde{C}\}, Y = \{M, A, Q, \varphi\}$ and $f()$ are the equations that subtract the left-hand side from the right-hand side at the system of equations in the next section. I numerically differentiate the system around the steady state to study the impulse response with respect to the aggregate shocks.

F. INCUMBENT FIRM'S MARKUP DETERMINATION

I can derive the full analytic equation for incumbents' markup determination. I first set Lagrangian function for incumbents' problem.

$$\begin{aligned} L = & \, p_i y_i - w n_i + \max_{\text{exit,stay}} [0, -e^{\zeta} + \frac{1}{1+r} EV(z, h; \mathrm{F})]\} \\ & + \lambda_n (e^z e^A n - y) + \lambda_h [(1-\delta) h_{-1} + \delta y - h] + \lambda_c [p^{-\rho} h_{-1}^{\theta_1 (1-\rho)} \tilde{c} - y] \end{aligned}$$

FOCs are following:

$$[n] : -w + \lambda_n e_{it}^z e^A = 0 \tag{17}$$

$$[y] : p + \lambda_n + \lambda_h \delta - \lambda_c = 0 \tag{18}$$

$$[h] : \beta EV_h - \lambda_h = 0 \tag{19}$$

$$[p] : y - \lambda_c p_{it}^{-\rho-1} \rho h_{-1,it}^{\theta_1(1-\rho)} \tilde{C}_t = 0 \tag{20}$$

I notice that λ_n is marginal cost and $\lambda_c = \frac{1}{\rho} p_{it}$ from equation (20). Plug those into equation (18) and rearrange to get the below.

$$\delta\lambda_h = (\frac{1-\rho}{\rho})p_{it} + mc_{it} \tag{21}$$

Using envelope condition, I obtain

$$\lambda_h = \frac{1}{1+r}G(\zeta_{it}^*)[\lambda_h'(1-\delta) + \lambda_c'\theta_1(\rho-1)\frac{y'}{h}] \tag{22}$$

Forward iterate equation (22) and use equation (21) to obtain,

$$(\frac{1-\rho}{\rho})p_{it} + mc_{it} = (\frac{\rho-1}{\rho})\delta\theta_1\frac{1}{1+r}G(\zeta_{it}^*)[\sum_{j=1}^{\infty}(1-\delta)^{j-1}\Pi_{j=1}^{\infty}[\frac{1}{1+r_{t,t+j}}G(\zeta_{it+j}^*)p'\frac{y_{it+j}}{h_{it+j-1}}] \tag{23}$$

Divide equation (23) by mc_{it} and multiply $\frac{\rho}{1-\rho}$ on both sides to obtain

$$\mu_{it} = \mu^* - E_t A_{it}$$
$$A_{it} = \delta\theta_1\sum_{j=1}^{\infty}(1-\delta)^{j-1}\Pi_{j=1}^{\infty}[\frac{1}{1+r_{t,t+j}}G(\zeta_{it+j}^*)]\frac{mc_{it+j}}{mc_{it}}\frac{y_{it+j}}{h_{it+j-1}}\mu_{it+j} > 0$$

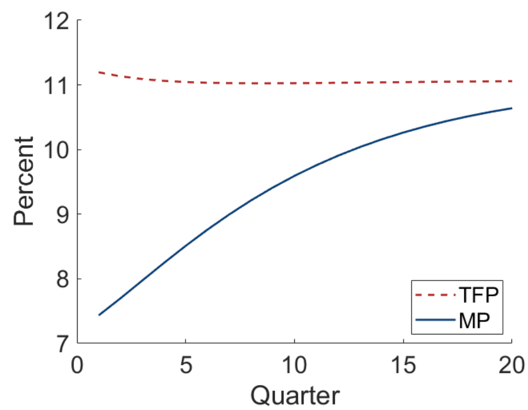
G. EXIT RISK DYNAMICS

This section shows how exit risk changes when there is aggregate shocks. I find that exit risk is stable with respect to positive technology shock. It implies that the value of a firm is stable when there is positive aggregate productivity shock since markup goes down. By contrast, the exit rate decreases for expansionary monetary policy shock since demand increases.

H. VALUE FUNCTION AND POLICY FUNCTION

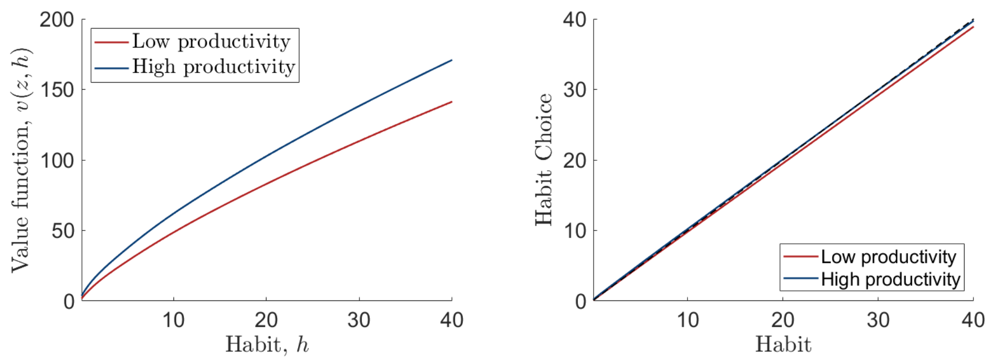
Figure A.10 shows that there exists a value function that is a fixed point of the incumbent firms' problem.

A.9: CHANGE IN EXIT RISK UPON AGGREGATE SHOCKS



A.10: VALUE FUNCTION AND POLICY FUNCTION

(a) Value Function (b) Policy Function



I. ADDITIVE HABIT

In this section, I depart from the constant elasticity of substitution (CES) demand assumption by adding additive habit. I generalize the Ravn, Schimitt-Grohe, and Uribe (2006) preference by including additive habit (maniacs) and multiplicative habit (loyal customers) at the same time. Additive part of habit represents maniacs to the given product and multiplicative part of habit denotes loyal customers. Additive habit represents demand that is completely price inelastic. Functional form for habit-adjusted consumption basket is now

$$\tilde{c}_j = [\int_i^I (c_{ij} h_i^{\theta_1} - \theta_2 h_i)^{\frac{\rho-1}{\rho}} di]^{\frac{\rho}{\rho-1}}$$

which gives following demand function:

$$c_i = (\frac{p_i}{\tilde{P}})^{-\rho} \tilde{C} h_i^{\theta_1(\rho-1)} + \theta_2 h_i^{1-\theta_1}$$

where θ_2 represents the degree of habit that is fully price inelastic. Then the incumbent's problem is the following.

$$V(\mathrm{S}_{-1};\mathrm{F}_{-1}) = \max_{p_i,h_i,n_i,y_i} \{\frac{p_i}{\tilde{P}} y_i - \frac{W}{\tilde{P}} n_i + \max_{\text{exit,stay}} [0, -\frac{\zeta}{\tilde{P}} + \Lambda E V(\mathrm{S};\mathrm{F})]\}$$

subject to

$$y_i = (\frac{p_i}{\tilde{P}})^{-\rho} \tilde{C} h_{i,-1}^{\theta_1(\rho-1)} + \theta_2 h_{i,-1}^{1-\theta_1}, \qquad \frac{p_i}{\tilde{P} \leq \bar{p}}$$

and production function, operating cost distribution, and eight laws of motion for the state variables from the main text. Additionally, I need to assume that a maximum price exits[43] since there is completely price inelastic demand. The maximum level of price is set to be high enough[44]. I calibrate θ_2 using an additional moment[45].

In this specification, firms price elasticity is a weighted sum of the price elastic habit part and price inelastic habit part. Therefore, firms price elasticity changes as a firm grows. This channel adds additional effect of channel. However, the result is similar[46].

43. One may relax this assumption slightly using a Logit function that the probability of dropping habit increases as relative price increases. In this case, I need more parameters to match.

44. I set the relative price cannot exceed 2.5.

45. For the results in this section, I use 0-2 year firm number share. I repeat the entire calibration process, and the model can match data fairly well.

46. I provide the result upon request.