Market Concentration and Business Cycles

Ke Chao
Washington University in St. Louis

Click here for the latest version

October 2021

Abstract

How does market concentration affects business cycles? We build a model featuring the dynamic competition between a forward-looking large firm and a continuum of heterogeneous entrepreneurs who are financially constrained. The elasticity of demand of the large firm is endogenously determined by the strategic competition. We find that the effect of the market concentration is non-monotonic. It is conditional on how a shock alters the market power of the large firm and, hence, how her markup responds to the shock. We show that although the endogenous markup mitigates shocks on large firms, it significantly amplifies shocks biased to entrepreneurs such as credit crunches. Furthermore, for homogeneous productivity shocks, due to financial frictions, entrepreneurs are more cyclically sensitive. Consequently, the strategic competition amplifies the aggregate decline in output while accelerates the recovery. The increasing in concentration dampens homogeneous productivity shocks because large firms are more productive. Yet, the markup of large firms is more responsive, which offsets the effect of the higher productivity by 81%. We show that the strategic competition distorts the optimal policy intervention during crisis by dampening the subsidization on entrepreneurs. The predictions of the model are consistent with the Compustat data.

1 Introduction

With the increasing trend of the market concentration in the United States, many research have explored the implications of the concentration on the economic dynamics (see, e.g. Dorn et al. (2017), Autor et al. (2020), Wang and Werning (2020)), and highlighted the differentiated exposures of the small and the large firms to aggregate shocks (e.g. Crouzet and Mehrotra (2020)). Yet, the researches studying the relation between market concentration and business cycle are largely missing. What are the implications of the concentration for economic fluctuations? How does the concentration distorts the optimal stabilization policy? To answer the questions, we study business cycles and the optimal policy interventions with a model featuring the strategic competition between a large firm and a continuum of competitive entrepreneurs who are financially constrained.

We find that the implications of the market concentration on business cycles are non-monotonic: it is conditional on how does the shock change the market power and, therefore, how does the markup of large firms respond to the shock. We find that the subsidization policy is by its nature benefiting the large firm, since her marginal cost is more elastic to factor prices. As a result, the welfare cost to implement the policy is amplified since the raised markup suppresses the wage. The study suggests that the more detailed and agent-specific policies should be applied to minimize the the distortions on the labor demand. The predictions of the model are consistent with the empirical evidence from the Compustat data.

The economy we study is composed of a continuum of sectors. Within each sector, based on the timing of the dominant firm model by Gowrisankaran and Holmes (2004), we assume that there exists a dynamic Stackelberg competition between a large firm who plays as a leader, and a continuum of competitive entrepreneurs behaving like followers. We generalize their framework by assuming that the competitive entrepreneurs are financially constrained and are heterogeneous in wealth and productivity as in Buera and Moll (2015). Taken prices as given, the competitive entrepreneurs not only choose the amount to produce, but also the amount to save in order to finance their capital investment in future. Anticipating the saving and the supply of the competitive entrepreneurs, the forward-looking large firm optimizes the pricing strategy to maximize her lifetime value based on the current states and independent with the histories.

In the model, the elasticity of the demand of the leader is endogenously determined by the strategic competition. Similar to the demand system imposed by the Kimball aggregator as in Boar and Midrigan (2019), the model has the property that the static elasticity of demand of the leader is a decreasing function with respective to her market share. Additionally, because of financial frictions, the forward-looking leader internalizes that her pricing strategy has dynamic effects on the wealth accumulation of the followers, which affects the demand of the leader in future. Therefore, in aggregate, the elasticity of demand for the leader is compound. It consists of the static and the dynamic elasticity of demand, where the latter is the discounted present value of the influence of the current pricing strategy on her demand in future. By proposing a tractable framework, the model provides a specific micro-foundation for the demand system, which gives a new insight into how does the firm's markup endogenously evolve along with business cycles.

To elucidate the mechanism through which the endogenous markup influences the fluctuations of the economy, we set a control group with the isomorphic framework except that the large firm is consisted of a continuum of homogeneous unconstrained firms that behave monopolistic competitively. The elasticity of substitution is calibrated such that the representative unconstrained firm charges the constant markup that is identical to the steady state markup of the strategic large firm. We evaluate the dynamics of the economy of the both models after the same productivity shocks. By comparing the differences in resource allocations, pricing strategies and responses in optimal policies across the two models, we illuminate the implications of the strategic competition for business cycles.

We find that the effect of the strategic competition is non-monotonic and depends on how does the shock alters the market power of the leaders and, therefore, how does the markup respond to the shock. Specifically, in transition dynamics, we isolate the two channels, the markup-output channel and the wealth-accumulation channel, through which the endogenous markup affects the shock. Because of financial frictions, a transitory shock could have long-lasting effect by undermining the strength of the balance sheet of the constrained firms. Subsequently, the wealth-accumulation channel represents the implication of the endogenous markup by altering the process to accumulate wealth. The closed-form analysis reveals that the two channels work in opposite directions. By raising the efficiency wedge, the increase in markup amplifies the shock through the markup-output channel. Yet, the raised markup promotes wealth accumulation by increasing the relative demand and thus the profits of entrepreneurs. The quantitative analysis shows that the markup-output channel always dominates. More specifically, although the endogenous markup mitigates shocks on large firms, it significantly amplifies shocks that are biased to entrepreneurs.

We generalize our quantitative analysis by evaluating the dynamics of the economy with standard shocks such as credit crunches and homogeneous productivity shocks. Throughout the dynamic transitions, because of financial frictions, entrepreneurs are more cyclically sensitive than large firms. Therefore, the endogenous markup amplifies the aggregate decline of the economy yet accelerates the speed of convergence. In particular, the effect of credit crunches would be largely underestimated if we do not take into account the strategic competition.

In further, by separately calibrating the productivity of large firms, we explore the implications of the increasing in concentration after homogeneous productivity shocks. The calibration captures the feature that when the economies are becoming more concentrated, the productivity of large firms would be higher. In aggregate, we find the increasing of concentration mitigates the shocks because the resources are re-allocated to the more productive leaders. However, because the markup is more elastic to the elasticity of demand, the effect of the endogenous markup is more significant too. Consequently, the more responsive markup offsets the effect of the higher productivity by 81%.

To investigate the policy implications of the strategic competition, we study the optimal stabilization policy to cut interest rate after a negative shock on the wealth of entrepreneurs as Itskhoki and Moll (2019)[1]. We draw two main lessons from the experiments of the uniform interest cut.

[1]Itskhoki and Moll (2019) discusses the optimal development policy to subsidize labor supply within a competitive

First, consistent with Ottonello and Winberry (2020), we find that large firms benefit more from the uniform subsidization, since they are unconstrained and, therefore, their marginal cost is more elastic to factor prices. Second, the strategic competition distorts the optimal policy by suppressing the equilibrium wage and thus increasing the welfare cost on households. The experiment implies that the government should apply more-detailed and agent-specific interventions that is targeted on the constrained. We in further evaluate the optimal interest rate cut that is biased to entrepreneurs. Compared to the uniform interest cut, the biased policy exhibits a higher welfare improvement. Yet, because of the strategic competition, the policy maker should be conservative on the interest rate cut during the crisis. Because the endogenous markup dampens the subsidization to entrepreneurs by distorting the resource allocations and raising the equilibrium wage.

We test the three main predictions of the model: the shock is amplified when it increases the market share of large firms; yet in the long run, the same type of the shock accelerates the recovery and the wealth accumulation of entrepreneurs. We test the predictions with the Compustat data. More specifically, we define the large firm as the top 2 firms of each sector in NAICS 6-digit level to match the market share of the large firms defined by Crouzet and Mehrotra (2020). We confirm the effect in the short run by validating that there exists a statistically significant negative relation between the increase of the market share of the large firms and the demeaned growth rate of the sales of the sector. We validate the long run implication from the two dimensions: between-crisis and within-crisis. We first compare the dynamics between the recent three crisis after 2000. We find that the crisis where large firms are relatively more impacted features a slower recovery process, accompanied with a more persistent decline in the liability and the income of small firms. Furthermore, we test the relation within the period of the financial crisis. We confirm the long run effect by showing the significant positive relation between the change of the market share of large firms and the future growth rate of the equity and the income of small firms. We confirm the implication of increasing in concentration by documenting the significant positive effect of the leader's market share on the sector's demeaned growth rate when the shock is biased to followers.

1.1 Literature Review

Our study is related to the series of papers exploring the macroeconomic implications of concentration in the product market. On the side of the macroeconomics, the mainstream modeling of market concentration is based on the combination of the Kimball aggregator and monopolistic competitions to guarantee that the elasticity of demand is a decreasing function of the market share. Therefore the aggregator imposes the specific demand system such that the firms with the greater market shares endogenously charge higher markup. Boar and Midrigan (2019) evaluates the aggregate and distributional impact of product market interventions in a model featuring the concentration in wealth and firm ownership. Baqaee and Farhi (2020) and Edmond et al. (2018) investigate how does the increase in the size of the market affects welfare and GDP. Although the aggregator is flexible to the

framework where the economy starting from the initial state such that the wealth of entrepreneurs are way below than steady state level; while our analysis of business cycle focuses on the minor deviations of the wealth of followers.

market structure and therefore has the great capability to fit the data, we lose the insights behind the different pricing strategies across the firms exactly because of the convenience the aggregator brings. In contrast, our paper provides a particular micro-foundation in a tractable framework that features the dynamic strategic competition, which helps us to establish the intuitions and to bridge the gap between the model and the real world.

Conversely, the papers modeling the concentration with strategic competitions are mostly established on the structure of the duopoly game. Chen et al. (2020) explores the implications of competition-distress feedback effect on asset prices and financial contagion with a model where two firms compete over customer bases. Mongey et al. (2017) combines nominal rigidity with duopoly competitions and finds that the output responses to monetary shocks are much higher compared to the monopolistic competition. Olmstead-Rumsey (2019) builds a model based on the structure of duopoly game in creative destruction to explain the raising of the concentration in United States around 2000. Although the deep insights they bring to the macro dynamics, the papers omit the other important component of the economy, the small competitive firms. By contrast, we are focusing on the competition between the dominant firms and the competitive small firms, which are quantitatively important by constituting the other half of the economy and have the distinct pricing strategies compared to the market leaders.

Furthermore, because of financial frictions, we think that it is also qualitatively important to highlight the strategic competition between the small and the large firms is of great importance in exploring the implications of concentration on business cycle, especially in studying the propagation of the shocks. Starting by Gertler and Bernanke (1989), it is widely studied that a transitory shock could have long-lasting effects by undermining the strength of the balance sheet of the firms (see else, e.g. Bernanke et al. (1999), Brunnermeier and Sannikov (2014), Brunnermeier et al. (2012)). Given that the small firms are more financially constrained and more sensitive to the monetary shocks, (e.g. Chodorow-Reich et al. (2021), Caglio et al. (2021)), the strategic competitions between the small and the large firms can have serious implications on the process of rebuilding the balance sheet after crisis and thus altering the macro dynamics in the recovery process. Deviating from the traditional competitive framework in the literature of financial frictions, the paper contributes to the literature on exploring the macroeconomic implications of financial frictions on business cycle by evaluating the effects of strategic pricing on the process of the wealth accumulations of the constrained entrepreneurs.

Our work is related to the strand of literature that highlights the differences in the fluctuations between the small and large firms around the business cycle. Starting from Gertler and Gilchrist (1994), many has explored the firms' responses to shocks are heterogeneous in size. Kudlyak and Sanchez (2016) uses public QFR data to show that the large firms' short-term debt and sales contracts relatively more than that of the small firms in the Great Recession. Crouzet and Mehrotra (2020) documents that the cyclicality of sales and investment are declining with firm size with the confidential QFR data. The strand of the research implies that the business cycle has heterogeneous effects on the firms with different sizes. Therefore, the market power of the dominant firms, or the market concentration itself is evolving over the business cycle. By contrast, our work contributes to the literature by exploring the implication of the cyclical evolution of market concentration on business cycle and the propagation of crisis.

The paper is also related to the literature studies the optimal policy intervention with financial frictions. Our work is closely related to Itskhoki and Moll (2019) where they studies optimal development policy to subsidize the constrained entrepreneurs in the transition dynamics where the economy starts from an initial state with the wealth of the entrepreneurs below its steady state value. They find that it is optimal to increase labor supply in the initial phase of transition when the financial wealth low enough. By contrast, we are focusing on the optimal intervention with minor deviation of the wealth around the steady state. Boar and Midrigan (2019) evaluate various steady state policy and find that optimal regulation improves allocative efficiency, thereby increasing product market concentration. Conversely, our analysis investigates the stabilization policy and focuses on how does the optimal policy response is distorted by the concentration. In further, different from their work, we emphasizes the cost to implement the policy is increasing with concentration, since the raised markup of dominant firms suppresses the wage. Bianchi (2016) analyzes the trade-off of bailout policies between the ex-post welfare improvement and ex-ante risk-taking. In contrast to their work, we study the ex-post improvement while focusing on the distortion from the concentration.

In terms of methodology, following Atkeson and Burstein (2008) and Wang and Werning (2020) we formulate a model where the economy is consisted of a continuum of sectors, while within each sector, by applying the idea of the dominant firm model by Gowrisankaran and Holmes (2004), we introduce the dynamics strategic competition over prices between the dominant firm and competitive fringes. Corbae and D'Erasmo (2019) applies the same structure of the strategic competition to the banking sector to study the financial regulation. Different from their work, we focus on the implications of the concentration of the product market on business cycle. Incorporating the idea of the intrinsic differences in the outputs of the large and small firms by Holmes and Stevens (2014), we assume that the elasticity of substitution between the large and small firms is lower than that within the products of the small firms, which is a crucial assumption for the closed form characterization. Finally, given the documented fact that the small firms are more constrained (e.g. Chodorow-Reich et al. (2021), Caglio et al. (2021)), following Buera and Moll (2015) and Itskhoki and Moll (2019), the financial friction is introduced as the leverage constraint applied only to the competitive entrepreneurs.

The remaining part of the paper proceeds as follows. Section 2 illustrates the basic statistics related to the market structure of the United States and the correlation between the concentration and business cycles. Section 3 describes the model. Section 4 analytically characterizes the dynamics of the economy after impulse shocks. Section 5 quantifies the model and displays numerical experiments. Section 6 explores the policy implications and section 7 contains empirical studies. Finally, section 8 concludes the paper.

2 Basic statistics

In this section we show three main facts that motivate our research and modeling choices, including the concentrated product market of the United States, the differentiated financial conditions across firm size, and the evolution of market concentration along with business cycles.

Table 1 from Crouzet and Mehrotra (2020) documents the high degree of skewness in the distribution of sales and growth of firms across the whole economy using the confidential QFR data. The top 0.5% of firms have assets of 6 billion and sales of 1.5 billion which are roughly 10 times greater than the firms with size between 99% to 99.5%. Meanwhile, the average asset and sales of firms within the bottom 90% of the size distribution are less than thousandth of that of market leaders. Similarly, there exists a significant trend of concentration within sectors. Table 2 displays the within-sector share of sales by size group where the sector is defined in 6-digit level. According to the Census 2012, on average, there are more than 700 firms within each sector while 42% of the market is taken by top 4 firms. By contrast, firms of top 5 to 50 own another 40% and the last 20% of the market belongs to the left 650 firms. The statistics suggest that ordered by assets, the marginal market share of firms declines very quickly even within-sector, implying a market structure with the coexistence of very few dominant firms and a large amount of small firms.

Table 3 presents an overview of the financial statistics provided by Crouzet and Mehrotra (2020). Compared to the top 0.5% firms, the firms within the bottom 90% of the assets distribution are more reliant on debt finance. The fraction of bank debt of the top 0.5% of firms is 0.28 while that of the bottom 90% of firms is close to 0.5. Furthermore, the difference in the fraction with zero leverage is significant. 20% of small firms are not leveraged. The financial statistics illustrates the differences of the financial methods and conditions between small and large firms. Together with the theory of financial frictions that shows the persistence of shocks is related to the wealth accumulation process of the economy, the paper is motivated to explore how concentration alters business cycles through the effect of strategic competitions on the wealth accumulation process of small firms.

The firms with different sizes are heterogeneously exposed to aggregate shocks. In general, small firms are more cyclically sensitive to aggregate fluctuations. Figure 1 illustrates the comparison of the mean of the sales growth rate between small and large firms across sectors over business cycle with the COMPUSTAT data. The large firms are classified as the top 2 firms measured in asset within each sector defined in NACIS-6[2]. The variance of the growth rate of small firms is 13.69%, nearly half larger that of large firms (9.63%). Crouzet and Mehrotra (2020) documents the similar trend by the confidential QFR data in figure 2. The heterogeneous cyclicality implies that during crisis, the market power of large firms are increasing. Consistent with the implication, the second plot shows that during the Financial Crisis and the pandemic of Covid, the detrended market share of large firms is increasing. An exception is the Recession 2001 where the market share of large firms is decreasing initially then increasing, implying that the effect of the Recession 2001 is biased

[2]The choice of the top 2 firms is to match the market share (around 50% in the Census 2012) of large firms defined as the top 1% firms by Crouzet and Mehrotra (2020)

to large firms. The heterogeneous exposure to aggregate shocks alters the strategic competition between large and small firms. The third plot shows the demeaned weighted markup of firms, where the markup is calculated following De Loecker et al. (2020)[3]. In general, the evolution of the markup of large firms is countercyclical and consistent with the evolution of their market power. Furthermore, compared to large firms, the volatility of markup of small firms are much less sensitive. The evolution of markup implies that aggregate shocks differentially alters the pricing strategies among firms with different market power. Hence, inversely, the paper is motivated to explore how the strategic competition affects business cycles.

The statistical facts imply that to explore the relation between market concentration and business cycle, it is both quantitatively and qualitatively important to simultaneously model the large and small firms. Quantitatively, although the product market of the United State is highly concentrated, the market share of small firms is sizable when aggregated. Qualitatively, the market leader and the competitive fringes are not only different in their pricing strategies, but also distinguished in their borrowing capacities. Moreover, firms are heterogeneous exposed to aggregate shocks. Hence aggregate shocks would alter the strategic competition among firms with different sizes. Therefore, to investigate the implications of market concentration on business cycle, and to emphasize the interactions between small and large firms, we consider a stylized market structure with dynamic price competition between a dominant firm and a continuum financial constrained entrepreneurs. Consistent with figure 1, the model provides a mechanism through which the markup of large firms evolves along with business cycles corresponding to the fluctuations of their market power. The analysis of the strategic competition between small and large firms sheds light upon the way in which market concentration impacts the dynamics of the economy.

3 Model

To evaluate the relation between boom-bust cycles and market concentration, this section presents the model with the strategic competition between a large firm and a continuum of small firms, where the markup of the large firm is endogenously determined by the strategic competition. In order to capture the heterogeneous pricing strategies of the firms, we assume that the firms are playing a dynamic Stackelberg game. The small firms are assumed to be the price-takers and thus reacting competitively as followers. Whereas the dominant firm anticipates the reactions of competitive fringes and behave like leaders. To explore the mechanism through which the particular concentrated market structure influences business cycle, we first consider the stylized impulse shocks on technology that are biased to the leader or the followers. We compare the dynamics of the economy of the aggregate output and wealth, between the case of the Stackelberg competition and the competitive benchmark where the leader is a price-taker.

[3]For simplicity, the output-input elasticity is chosen as the constant 0.85 consistent with the benchmark in their paper.

3.1 Households

We assume that there exists a representative household who supplies labor l_t inelastically and maximizes the discounted utility of the consumption of the final good c_t

$$\max_{c_t, b_t, l_t} \sum_{t=0}^{\infty} \beta^t u(c_t), \tag{1}$$

subject to

$$\begin{aligned} c_t + b_t &= w_t l_t + (1+r)\, b_{t-1}, & (2) \\ b_t &\geq 0, & (3) \end{aligned}$$

where b_t is the risk-free bond. By $b_t \geq 0$, it is assumed that the household cannot borrow. The interest rate r is assumed to be a exogenous constant, which implies that the model depicts an open economy with a fixed real rate. Note that if $\beta(1+r) < 1$, given the assumption that the household cannot borrow, it implies that the household is hand-to-mouth at the steady state.

3.2 Firms

The production side of the economy consists two layers: the final good sector and the intermediate good sectors. The final good y is produced with the a continuum of intermediate good y_i according to the CES aggregator:

$$y = \left[\int_0^1 (\rho_i y_i)^{\frac{\sigma-1}{\sigma}} \, \mathrm{d}i \right]^{\frac{\sigma}{\sigma-1}}, \tag{4}$$

where y_i is the output of the intermediate good sector i and ρ_i is exogenous and represents the sector specific productivity. The CES aggregator implies that, after normalizing the price of the final good to 1, the demand function for y_i is

$$y_i = p_i^{-\sigma} \rho_i^{\sigma-1} y. \tag{5}$$

Within each intermediate good sector i, there exists a dominant firm (leader) with some constant productivity $z_{i,l}$ and one unit of a continuum of competitive entrepreneurs (followers) indexed by $j \in (0,1)$. The entrepreneur j of the sector i is endowed with some productivity $z_{i,j}$ and produces a particular variety $y_{i,j}$ if she chooses to be active. To keep the analytical tractability of the model, the varieties produced by the entrepreneurs within the sector i are assumed to be perfectly substitutable. The output of sector i is produced by combining the output of the leader, i.e. $y_{i,L}$ and the aggregate output of the followers $y_{i,F}$ according to the CES technology

$$y_i = \left[(\rho_{i,L}y_{i,L})^{\frac{\epsilon-1}{\epsilon}} + (\rho_{i,F}y_{i,F})^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}}, \tag{6}$$

$$y_{i,F} = \int y_{i,j}\mathrm{d}j. \tag{7}$$

where $\rho_{i,L}$ and $\rho_{i,F}$ represent the agent-specific productivity.

The assumption of the imperfect substitution between the dominant firm and the competitive entrepreneurs within a sector follows the idea of Holmes and Stevens (2014), that small firms are facilitated by face-to-face contact and targeted to specialty goods, while large firms focus on standardization with mass production technology. Thereafter, the elasticity of substitution between the products of the entrepreneurs and the dominant firm should be lower than the elasticity within the entrepreneurs. In particular, we assume that the outputs of competitive entrepreneurs are perfectly substitutable. Furthermore, without the loss of generality, the elasticity of substitution σ across sectors is assumed to be less than the elasticity ϵ within a sector.

The standard CES aggregator implies that the demand for the leader and the aggregate demand for the followers within the sector i take the form that

$$y_{i,L} = \left(\frac{p_{i,L}}{p_i}\right)^{-\epsilon}\rho_{i,L}^{\epsilon-1}y_i, \tag{8}$$

$$y_{i,F} = \left(\frac{p_{i,F}}{p_i}\right)^{-\epsilon}\rho_{i,F}^{\epsilon-1}y_i, \tag{9}$$

where the price index p_i equals to

$$p_i = \left[\left(\frac{p_{i,L}}{\rho_{i,L}}\right)^{1-\epsilon} + \left(\frac{p_{i,F}}{\rho_{i,F}}\right)^{1-\epsilon}\right]^{\frac{1}{1-\epsilon}}. \tag{10}$$

3.2.1 Timing

To characterize the heterogeneity in pricing strategies across firm size, we assume that the large and small firms are playing a dynamic Stackelberg game over price: with the market power, the large firm anticipate the policy functions of the entrepreneurs and announce the pricing rule as a leader. Whereas observing the announcement of the large firm, the competitive small firms are price-takers. The game is dynamic therefore leaders are forward-looking. The timing of the game across all the sector $i \in (0,1)$ is identical and shown as follows.

1. At time t, given cash-on-hand $m_{i,j}$, the entrepreneur j in the sector i draws a new productivity $z'_{i,j}$ from an i.i.d. Pareto distribution $g_i(z) = \gamma_i z^{-\gamma_i - 1}$ with a Pareto parameter $\gamma_i > 1$;

2. the productivity shocks ρ'_i, $\rho'_{i,L}$, $\rho'_{i,F}$ of the next period are realized;

3. the leader of the sector i announces the price of the next period $p'_{i,L}$;

4. taken $p'_{i,L}$, p'_i and $p'_{i,F}$ as given, the followers choose saving $a'_{i,j}$ and output $y'_{i,j}$ for the next period.

5. at time $t+1$, the leaders and followers produce according to the plan.

The following diagram summarizes the timeline of the game.

3.2.2 Followers

Now consider the entrepreneur j of the sector i, after drawing the productivity $z_{i,j}$ of the next period, the Bellman equation of that entrepreneur can be written as

$$v_{i,j}\left(m_{i,j}, z_{i,j}\right) = \max_{y'_{i,j}, a'_{i,j}, l'_{i,j}, k'_{i,j}} \log\left(m_{ij} - a'_{ij}\right) + \beta E v_{i,j}\left(m'_{i,j}, z'_{i,j}\right), \tag{11}$$

subject to

$$m'_{i,j} = (1+r)\, a'_{i,j} + \pi'_{i,j}, \tag{12}$$

$$\pi'_{i,j} = p'_{i,F} y'_{i,j} - w' l'_{i,j} - r k'_{i,j}, \tag{13}$$

$$y'_{i,j} \leq \left(z_{i,j} k'_{i,j}\right)^{\alpha} l_{i,j}^{1-\alpha}, \tag{14}$$

$$k'_{i,j} \leq \lambda a'_{i,j}. \tag{15}$$

Here $m_{i,j}$ denotes an entrepreneur's "cash-on-hand" and $a'_{i,j}$ represents the net wealth of the entrepreneur saving for next period. Due to financial frictions, the borrowing capacity of the entrepreneur is constrained by its own net wealth: the leverage ratio of the entrepreneur is bounded by λ. The specific formulation of capital market imperfection favors analytical convenience and is isomorphic to the model where entrepreneurs own capital and issue debt to finance the investment as discussed in Itskhoki and Moll (2019). With $\lambda > 1$, the entrepreneur has to have "skin in the game": the fraction of externally financed capital investment is bounded up-to $1 - \frac{1}{\lambda}$. Therefore, entrepreneurs choose saving $a'_{i,j}$ not only to smooth consumption but also to satisfies the internal finance in next period.

Lemma 1 is a straight-forward generalization of the characterization of the entrepreneurs in Itskhoki and Moll (2019).

Lemma 1 *There exists a productivity cutoff:* $\underline{z}_i' = p_{i,F}'^{-\frac{1}{\alpha}} \frac{r}{\alpha} \left(\frac{w'}{1-\alpha}\right)^{\frac{1}{\alpha}-1}$*, such that*

$$y_{i,j}' = \begin{cases} p_{i,F}'^{-1} \frac{r}{\alpha} \frac{z_{i,j}'}{\underline{z}_i'} \lambda a_{i,j}', & z_{i,j} \geq \underline{z}_i \\ 0, & otherwise \end{cases}, \tag{16}$$

$$\pi_{i,j}' = \begin{cases} \left(\frac{z_{i,j}}{\underline{z}_i} - 1\right) r\lambda a_{i,j}', & z_{i,j} \geq \underline{z}_i \\ 0, & otherwise \end{cases}, \tag{17}$$

$$l_{i,j}' = \begin{cases} \frac{r}{\alpha} \frac{1-\alpha}{w} \frac{z_{i,j}}{\underline{z}_i} \lambda a_{i,j}, & z_{i,j} \geq \underline{z}_i \\ 0, & otherwise \end{cases}. \tag{18}$$

Furthermore, the saving of entrepreneur j is a constant fraction to cash-on-hand

$$a_{i,j}' = \beta m_{i,j}. \tag{19}$$

As shown by Lemma 1, constant return to scale together with the linear collateral constraint guarantee that the output, profit and labor demand of entrepreneur j are linear functions of savings, i.e. $a_{i,j}'$. The linearity is the key property to keep the analytical tractability of the model. In particular, since the profit is a linear function of wealth, the rate of return on saving of the entrepreneur j is constant. Thereafter, the entrepreneur would save a constant fraction of cash-on-hand, which exactly equals to the discount factor β in the case of logarithmic utility.

Lemma 2 *The aggregate output, profit and savings of entrepreneurs in sector i take the form*

$$y_{i,F}' = \left(p_{i,F}' \frac{1-\alpha}{w}\right)^{\frac{1}{\alpha}-1} \lambda\beta m_{i,F} \int_{\underline{z}_i}^{\infty} z dG_i(z), \tag{20}$$

$$\pi_{i,F}' = \frac{\alpha}{\gamma} p_{i,F}' y_{i,F}', \tag{21}$$

$$m_{i,F}' = \beta(1+r) m_{i,F} + \pi_{i,F}', \tag{22}$$

where $G_i(z)$ is the cumulative probability function of z following Pareto distribution.

Since the productivity shock is *i.i.d.*, at each period, the distributions of productivity and wealth are independent. Therefore, by aggregating (16) and (19) over all entrepreneurs within sector i, in Lemma 2 we can derive aggregate output and the law of motion of aggregate cash-on-hand of entrepreneurs are both linear functions of the aggregate cash-on-hand of the sector $m_{i,f}$. Lemma 3 significantly simplifies the dynamics game. Due to the perfect substitution of the varieties across entrepreneurs, the demand of the leader is a function of the aggregation of the output of followers. Meanwhile, because both the aggregate output and the law of motion of cash-on-hand are only functions of the current aggregate cash-on-hand, $m_{i,f}$ could serve as a sufficient state variable in the dynamic game and we do not need to characterize the joint distribution of wealth and productivity of followers, significantly reducing the state space.

To summarize, the four key assumptions that contribute to the tractability of the model are: competitive entrepreneurs, constant return to scale, linear borrowing constraint and *i.i.d.* productivity shock. The assumption of competitive entrepreneurs ensures that the followers do not behave strategically. The assumptions of constant return and linear borrowing constraint together guarantee the linear policy function for the entrepreneurs. Finally the assumption of *i.i.d.* productivity shock reduces the state space of the dynamic game to the aggregate cash-on-hand of followers. However, it worth to mention that the only assumption required to maintain the numerically tractability of the model is the competitive entrepreneurs. Without linearity, the dynamics are numerically achievable except that one has to trace the dynamics of follower's joint distribution of wealth and productivity.

3.2.3 Leaders

In each sector i, there exists a dominant firm whose the shares are traded in a global market. The dominant firm is assumed to be different from competitive fringes in two dimensions. First of all, compared to small firms, the dominant firm is assumed to be unconstrained. More importantly, the dominant firm in the model is assumed to behave strategically as a leader, that is, she anticipates the best response of competitive fringes and chooses the optimal pricing strategy to maximize the discounted profits. In particular, the dynamic equilibrium we are looking for is the feedback Stackelberg equilibrium. Alternatively speaking, the equilibrium is sub-game and Markov perfect. Therefore, the state space is independent of histories and only depends on the current states.

To describe the leader's dynamic problem, it is convenient to start with the static cost minimization problem. We assume that the leader in sector i produces output $y_{i,l}$ with labor $l_{i,l}$ and capital $k_{i,l}$ using a constant return to scale technology. Therefore, the cost minimization problem of the leader i implies that the leader's marginal cost of production is a constant given by

$$\phi_i \quad = \quad z_{i,L}^{-\alpha} \left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{w}{1-\alpha}\right)^{1-\alpha}. \tag{23}$$

Since the leaders are traded in a global capital market and are not financially constrained, the leaders choose price $p_{i,l}$ to maximize the present value of the profits. Furthermore, since the marginal cost of the leaders is a constant, it is equivalently to characterize the pricing strategy by the markup κ_i they charges. Subsequently, anticipating the aggregate output and law of motion of wealth of followers, the Bellman equation of the leader of sector i is given by

$$v_{i,L}\left(m_{i,F}; w\right) \quad = \quad \max_{\kappa_i'} \left(\kappa_i' - 1\right) \phi_i' y_{i,L}' + \frac{1}{1+r} v_{i,L}\left(m_{i,F}'; w'\right) \tag{24}$$

subject to

$$
\begin{aligned}
p'_{i,L} &= \kappa'_i \phi'_i, && (25)\\
y'_{i,L} &= \left(\frac{p'_{i,L}}{p'_i}\right)^{-\epsilon} \left(\frac{p'_i}{p'}\right)^{-\sigma} \rho'^{\epsilon-1}_{i,L} \rho'^{\sigma-1}_{i} y, && (26)\\
p'_i &= \left[\left(\frac{p'_{i,L}}{\rho'_{i,L}}\right)^{1-\epsilon} + \left(\frac{p'_{i,F}}{\rho'_{i,F}}\right)^{1-\epsilon}\right]^{\frac{1}{1-\epsilon}}, && (27)\\
y'_{i,F} &= \left(\frac{p'_{i,F}}{p'_i}\right)^{-\epsilon} \left(\frac{p'_i}{p'}\right)^{-\sigma} \rho'^{\epsilon-1}_{i,F} \rho'^{\sigma-1}_{i} y, && (28)\\
&= \left(p'_{i,F}\frac{1-\alpha}{w}\right)^{\frac{1}{\alpha}-1} \lambda\beta m_{i,F} \int_{\underline{z}_i}^{\infty} z \mathrm{d}G_i(z), && (29)\\
m'_{i,F} &= \beta(1+r) m_{i,F} + \frac{\alpha}{\gamma} p'_{i,F} y'_{i,F}. && (30)
\end{aligned}
$$

The leader anticipates the responses of followers. First, she internalizes that by increasing her markup, it raises the price index of the sector p'_i, and, therefore it alters the demand and thus the equilibrium price of the followers $p'_{i,F}$, which in further decreases the demand of the leader. While dynamically, she anticipates the effect of $p'_{i,L}$ on the savings of followers. Consequently, the aggregate saving $m'_{i,F}$ influences the followers capacity to supply and thus the demand for the leader in future.

The FOC of the Bellman equation is

$$
\underbrace{y'_{i,L}\left[1 + \left(1 - \kappa'^{-1}_i\right) \frac{\frac{\partial y'_{i,L}}{y'_{i,L}}}{\frac{\partial p'_{i,L}}{p'_{i,L}}}\right]}_{\text{static}} + \underbrace{\frac{1}{1+r} \frac{\partial v_{i,L}}{\partial m'_{i,F}} \frac{\partial m'_{i,F}}{\partial p'_{i,L}}}_{\text{dynamic}} = 0, \tag{31}
$$

where the Envelope Theorem gives the partial derivative of value function:

$$
\frac{\partial v_{i,L}}{\partial m_{i,F}} = \left(\kappa'_i - 1\right) \phi'_i \frac{\partial y'_{i,L}}{\partial m_{i,F}} + \frac{1}{1+r} \frac{\partial v_{i,L}}{\partial m'_{i,F}} \frac{\partial m'_{i,F}}{\partial m_{i,F}}. \tag{32}
$$

The FOC of the Bellman equation consists of a static and a dynamic part. Note that by ignoring the dynamic part, one can find the standard relation between markup κ'_i and elasticity of demand. However, in the dynamic game, different from the classical New-Keynesian or international trade models, the leader is facing a discounted elasticity of demand: the pricing strategy today not only alters the current demand but it also impacts the demand in future by influencing the savings of the followers. Denote by $\tilde{v}_{t+k}$ the elasticity of demand in the period $t+k$ with respective to the price in period t, and by $\mu_{i,L,t}$ the market share of the leader, i.e. $\mu_{i,L,t} \equiv \frac{p_{i,L,t} y_{i,L,t}}{p_i y_i}$. Lemma 4 provides the characterization of the optimal markup of the leader by combining the FOC and the Envelope

Theorem iteratively.

Lemma 3 *The markup of leader is only a function of the discounted elasticity of demand $\Upsilon_{i,t}$ such that*

$$\kappa_{i,t} = \frac{1}{1-\Upsilon_{i,t}^{-1}}, \tag{33}$$

where $\Upsilon_{i,t}$ is the sum of the static elasticity $\upsilon_{i,t}$ and the discounted present value of the forward-looking elasticity $\hat{\upsilon}_{i,t+k}$ given by

$$\Upsilon_{i,t} = \upsilon_{i,t} + \sum_{k=1}^{\infty}\left(\frac{1}{1+r}\right)^k \frac{\pi_{i,L,t+k}}{\pi_{i,L,t}}\hat{\upsilon}_{i,t+k}, \tag{34}$$

$$\upsilon_{i,t} = \epsilon - (\epsilon-\sigma)\left[1+\left(\mu_{i,L,t}^{-1}-1\right)\frac{\sigma+\frac{\gamma}{\alpha}-1}{\epsilon+\frac{\gamma}{\alpha}-1}\right]^{-1}, \tag{35}$$

$$\hat{\upsilon}_{i,t+k} = \frac{\frac{\partial y_{L,t+k}}{y_{L,t+k}}}{\frac{\partial m_{F,t+k}}{m_{F,t+k}}}\left(\prod_{i=1}^{k-1}\frac{\frac{\partial m_{F,t+i+1}}{m_{F,t+i+1}}}{\frac{\partial m_{F,t+i}}{m_{F,t+i}}}\right)\frac{\frac{\partial m_{F,t+1}}{m_{F,t+1}}}{\frac{\partial p_{F,t}}{p_{F,t}}}\frac{\frac{\partial p_{F,t}}{p_{F,t}}}{\frac{\partial p_{L,t}}{p_{L,t}}}, k \geq 1. \tag{36}$$

In particular, $\upsilon_{i,t}$ is a decreasing function of $\mu_{i,L,t}$ such that

$$\lim_{\mu_{i,L,t}\to 1}\upsilon_{i,t} = \sigma, \tag{37}$$

$$\lim_{\mu_{i,L,t}\to 0}\upsilon_{i,t} = \epsilon. \tag{38}$$

Lemma 3 reveals that the optimal markup still is only a function of the elasticity of demand, except that different from the standard case, the elasticity of demand of the leader is a summation of the static elasticity and the discounted present value of the forward-looking elasticity. Furthermore, the static elasticity, i.e. $\upsilon_{i,t}$, is a function of market share of the leader, while in the standard case with CES aggregator, the elasticity of demand is a constant given by the within-sector elasticity of substitution.

Equation (34) highlights the difference between the dynamic game and standard CES models. First of all, $\upsilon_{i,t}$ is a decreasing function of the leader's market share and bounded between $[\sigma, \epsilon]$. Intuitively, when the market share of the leader is approaching to 1, the leader is a monopoly in the intermediate good sector i. However, because the final output is CES aggregated by a continuum of intermediate good, the leader is then monopolistic competitive in the upper-layer of the economy. Thereafter the static elasticity of demand equals to the between-sector elasticity of substitution σ. On the contrary, when the market share of the leader approaches to 0, the leader is approximately behaving monopolistic competitively within sector i. Therefore the static elasticity equals to the within-sector elasticity of substitution ϵ. In equilibrium, the static elasticity of demand, as a function of aggregate wealth of followers and productivity shocks, is endogenously oscillating between that of a monopolistic competitive firm and a monopoly. In addition, the second part of equation (34) explains the channel through which the current pricing affects the future demand. The leader internalizes that the current change of price would influences the profits and the savings of followers. Moreover, the leader also internalizes the law of motion of $m_{i,F}$, therefore the followers' abilities to borrowing and to supplying in the every future period. Subsequently, as Lemma 4 shows, the discounted elasticity of demand would be a life-time discounted elasticity that is the summation of the static elasticity and the discounted present value of the forward-looking elasticity of demand.

3.3 Definition of an Equilibrium

Normalize aggregate price index $p = 1$, and denote by ω the exogenous fraction of the population that are entrepreneurs. Given sequences of productivity and leverage shocks, a feedback Stackelberg equilibrium is defined by sequences of allocations and prices such that:

1. Given sequences of factor prices $\{r, w_t\}_{t=0}^{\infty}$, households choose $\{c_t, l_t, b_t\}_{t=0}^{\infty}$ to the maximize life-time utility in (1) subject to the budget constraint by (2).

2. For all $i, j \in (0, 1)$, given the prices sequences $\{p_{i,L,t}, p_{i,F,t}, p_{i,t}, r, w_t\}_{t=0}^{\infty}$, the sequences of allocations $\{c_{i,j,t}, a_{i,j,t}, y_{i,j,t}, l_{i,j,t}, k_{i,j,t}\}_{t=0}^{\infty}$ satisfies the competitive entrepreneur's Bellman equation (11) subject to the constraints (12)-(15).

3. Anticipating entrepreneurs' output in (20) and savings (22), taken the sequences of aggregate states $\{w_t, y_t\}$ as given, the sequence of markups $\{\kappa_{i,t}\}_{t=0}^{\infty}$ satisfies the Bellman equation (24) of the market leader.

4. The final goods market, the intermediate goods market, and the labor market clear:

$$y = \left[\int_0^1 (\rho_i y_i)^{\frac{\sigma-1}{\sigma}} \mathrm{d}i\right]^{\frac{\sigma}{\sigma-1}}, \tag{39}$$

$$y_i = \left[(\rho_{i,L} y_{i,L})^{\frac{\epsilon-1}{\epsilon}} + (\rho_{i,F} y_{i,F})^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}}, \forall i \in (0, 1) \tag{40}$$

$$(1-\omega)L = \omega \int \left(l_{i,L} + \int l_{i,j} \mathrm{d}j\right) \mathrm{d}i. \tag{41}$$

4 Closed-form characterizations

To analytically explore the effect of market concentration on business cycles, we characterize the evolution of the economy after productivity shocks with log-linearization. To highlight the role of the strategic behaviors of the leaders, we compare the dynamics of the economy with that of the monopolistic competitive (hence MC) benchmark. The structure of the benchmark is isomorphic to that of the model with strategic competition, except that the leaders in the benchmarks are behaving non-strategically.

The analysis emphasizes the importance of the strategic behavior of leaders. We observe two important results. First, the effect of concentration on business cycle is non-monotonic and shock-specific: conditional on how does the shock alter the market power of leaders, the endogenous change of markups might amplify or mitigate the accumulated effects of shocks. Furthermore, the direct effect of endogenous markup and the effect of it along with transition dynamics are contradictory. For the shocks increase the market power of leaders, the raised markup amplifies the impact of the shock. Yet, the higher markup dampens the decline of the wealth of the constrained entrepreneurs and, therefore, accelerates the recovery of the economy. The opposite is same for the shocks that declines the market power of leaders.

The section would proceed in the following order. First it would introduce the shocks, followed by the definition of the monopolistic competitive benchmark. Afterwards, it compares the dynamics of the economy between the benchmark with the model of strategic competition.

4.1 Formulating the shocks

We assume that economy starts at an steady state with $\rho_{i,j,1} = 1$, where $i \in (0, 1)$ and $j \in \{L, F\}$. To reveal the mechanism of the model, we impose agent-specific and unanticipated impulse productivity shocks that are symmetric across sectors.

Define by $\mathrm{d}\log x_t$ the deviation of x_t away from its steady state value, i.e. $\mathrm{d}\log x_t \equiv \mathrm{d}x_t/x$. The system starts at the steady state. At $t = 1$, we assume there is an unanticipated impulse shock on leaders or on followers that is symmetric across sectors. After the period of the shock, the productivity permanently reverts to the initial value from period $t = 2$ onward. In particular,

$$\mathrm{d}\log \rho_{L,t} = \begin{cases} 0, & t = 0, \\ \mathrm{d}\log \rho_{L,1} < 0, & t = 1, \\ 0, & t \geq 2. \end{cases}$$

or

$$\mathrm{d}\log \rho_{F,t} = \begin{cases} 0, & t = 0, \\ \mathrm{d}\log \rho_{F,1} < 0, & t = 1, \\ 0, & t \geq 2. \end{cases}$$

Note that since the shock is symmetric across sectors, I drop the subscript i. The parameter $\rho_{i,j,t}$ represents the productivity of the good $j \in \{L, F\}$ in the aggregation process of the intermediate good i. Therefore $\mathrm{d}\log \rho_{j,t} < 0$ captures a negative demand shock on agent j.

The main reason that we impose the impulse shocks is that it perfectly separates the impact and the propagation of the shock with financial frictions. The economy starts at steady state, which implies that $m_{F,1} = m_F^*$. Therefore, the deviation of the output at $t = 1$ is a direct outcome of the productivity shock. Although the productivity reverts to steady state for $t \geq 2$, the effect of the shock is persistent. The shock impacts the strength of the balance sheets of followers and, therefore, their borrowing capacity. Thus the only reason that $\mathrm{d}\log y_t \neq 0$ for $t \geq 2$ is the propagation of the shock due to financial frictions. Furthermore, by separately comparing the deviation of output during and away from the period of the shock between the benchmark model and the model with strategic competition, we can perfectly isolate the influence of market concentration on the instantaneous effect and the propagation of the shock. The analytical tractability brought by the impulse shocks helps us to interpret the dynamics of the economy in the numerical sections where we consider persistent shocks.

4.2 Monopolistic Competitive Benchmark

To reveal the effect of market concentration, we consider the following monopolistic competitive benchmark. We denote all variables of the monopolistic competitive (MC) benchmark with tilde, i.e. $\tilde{x}_t$. The main structure of the MC is consistent with the model presented in section 3. In the production side, there exists two layers, the final good and the intermediate good sectors. Yet, in the MC, the intermediate good $\tilde{y}_i$ is an aggregation of the product of a continuum of homogeneous unconstrained firms and a continuum of constrained heterogeneous entrepreneurs:

$$\tilde{y}_i = \left[(\rho_{i,L}\tilde{y}_{i,L})^{\frac{\epsilon-1}{\epsilon}} + (\rho_{i,F}\tilde{y}_{i,F})^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}}, \forall i \in (0,1)\,, \tag{42}$$

$$\tilde{y}_{i,F} = \int \tilde{y}_{i,j}\mathrm{d}j, \tag{43}$$

$$\tilde{y}_{i,L} = \left(\int \tilde{y}_{i,L,j}^{\frac{\eta-1}{\eta}}\mathrm{d}j\right)^{\frac{\eta}{\eta-1}}. \tag{44}$$

Note that the within-sector market structure is isomorphic to the model of strategic competition. Furthermore, the technology of unconstrained firms is homogeneous and assumed to be identical to the leaders in the model of strategic competition. Hence in the MC, the representative unconstrained firm corresponds to the leader in the model of strategic competition. Yet, the difference is that the output of the unconstrained is a CES aggregation across a continuum of homogeneous firms with elasticity of substitution η. In particular, we set $\eta = \Upsilon^*$, where Υ^* represents the steady state elasticity of demand of the strategic leaders. Other than that, the benchmark is assumed to be identical to the model of concentration.

Different from the strategic leaders, the unconstrained firms behave monopolistic competitively. Given that $\eta = \Upsilon^*$, they charge a constant markup $\tilde{\kappa}$ that is identical to the steady state markup of the strategic leader. By contrast, the markup of the strategic leader is endogenously evolving along with transition dynamics. Furthermore, because the representative unconstrained firm shares the same technology and productivity with the strategic leader, the steady state market share and real allocations across the two models are identical. By comparison to the monopolistic competitive benchmark, we hence isolate the effect of the strategic competition, so as to answer the question that given market concentration, how does the strategic behavior affect business cycles.

4.3 Log-linearization Analysis

In this section we characterize and explore the effect of endogenous markup using log-linerization analysis. We disaggregate the accumulated effect of the shock into the impact effect and the propagational effect, and separately evaluate the implications of endogenous markup.

Specifically, we are comparing the three deviations: $\mathrm{d}\log y_1$, $\sum_{t=2}^{\infty}\mathrm{d}\log y_t$ and $\sum_{t=2}^{\infty}\mathrm{d}\log m_{F,t}$ with respective to the MC. Because the shock is assumed to be a impulse, the deviation of the output during the period of the shock is a direct outcome of the shock, while the aggregate deviation of output in the following periods is a result of the fact that the shock impacts the strength of the balance sheet of followers. Therefore, $\mathrm{d}\log y_1$ and $\sum_{t=2}^{\infty}\mathrm{d}\log y_t$ isolate the impact and the propagation of the shock respectively. Related to the propagation of the shock, given that $m_{F,t}$ is the only state variable of the economy, $\sum_{t=2}^{\infty}\mathrm{d}\log m_{F,t}$ represents the aggregate deviation of the system along with the transition dynamics and $\frac{\mathrm{d}\log m_F'}{\mathrm{d}\log m_F}$ characterizes the speed of convergence. By separately comparing the three deviations to that of the MC, we can disaggregate the influence of endogenous markup throughout the whole transitions after the shock.

4.3.1 Analysis of a Myopic-Leader

We first consider the case with a myopic leader, by assuming that the leader ignores the forward-looking elasticity defined in Lemma 3, i.e. $\frac{\partial y_{L,t+k}}{y_{L,t+k}}/\frac{\partial m_{F,t+k}}{m_{F,t+k}} = 0$ for all $k \geq 1$. Different from a forward-looking leader, the discounted elasticity Υ_t of the myopic leader is static and only a function of its market share. It worth mentioning that the assumption of a myopic leader does not implies that the pricing strategy of the leader has no influence upon the dynamics of the system. The change of the markup the still affects the wealth of followers and, therefore, alters the dynamics of the economy.

With the assumption of a myopic leader, we obtain the analytical solution for the dynamics of the system. For the model with a forward-looking leader, we will provide a sufficient condition such that the effect of the strategic competition remains consistent.

In particular, for mathematical simplicity, we calibrate the productivity of the leader such that $\mu_L = \mu_F$, which is consistent with the market share of the large firms in the United States[4]. Given the simplification, the steady state markup is given by

$$\kappa \quad = \quad \frac{\Upsilon^*}{\Upsilon^* - 1}, \tag{45}$$

where the steady state elasticity of demand of the myopic leader is

$$\Upsilon^* \quad = \quad \epsilon - \frac{(\epsilon - \sigma)}{2}\frac{\epsilon + \frac{\gamma}{\alpha} - 1}{\frac{\epsilon+\sigma}{2} + \frac{\gamma}{\alpha} - 1}. \tag{46}$$

[4]We use the definition of the large firms by Crouzet and Mehrotra (2020) as the 1% firms. According to the census 2012, the mean of number of firms within each sector is 732, while the accumulated revenue share of the top 8 firms is 54.9%.

Because the MC and the model of strategic competition share the identical resource allocations and prices in the steady state, the difference of the steady state log-linearization across the two models only consists of the effect from the response of the markup of strategic leaders. Define by Δ_{x_t} the difference in the log-deviation across the two models, i.e. $\Delta_{x_t} = \mathrm{d}\log x_t - \mathrm{d}\log \tilde{x}_t$. Following the log-linearizating the system we obtain that, for $j \in \{L, F\}$

$$\Delta_{y_t} = \begin{cases} \underbrace{\frac{\partial \log y}{\partial \log \kappa} \frac{\mathrm{d}\log \kappa}{\mathrm{d}\log \rho_{j,1}} \mathrm{d}\log \rho_{j,1}}_{\text{direct impact effect}}, & t = 1 \\ \underbrace{\frac{\partial \log y}{\partial \log \kappa} \frac{\mathrm{d}\log \kappa}{\mathrm{d}\log m_F} \mathrm{d}\log m_{F,t}}_{\text{wealth-markup channel}} + \underbrace{\frac{\partial \log y}{\partial \log m_F} \Delta_{m_{F,t}}}_{\text{wealth-accumulation channel}}, & t \geq 2. \end{cases} \tag{47}$$

Equation (47) reveals that endogenous markup has two effects: the direct impact and the transitional effects on output. The transitional effect, characterized by $\sum_{t=2}^{\infty} \Delta_{y_t}$, refers the effect of the endogenous markup along with transition dynamics. Particularly, for the transitional effect, we can decompose two channels: wealth-markup channel and the wealth-accumulation channel, where the latter represents how does the endogenous markup influences the dynamics through altering the wealth accumulation of followers. Lemma 4 explores the implication of the endogenous markup in the short run (the impact effect).

Lemma 4 *Depending on how does the shock alter the market power of the dominant firm, the endogenous markup has differentiated effects in the short run. Particularly,*

1. *for negative shocks on followers, i.e.* $d\log \rho_{F,1} < 0$*,*

$$\Delta_{y_1}(\rho_{F,1}) \quad < \quad 0,$$

2. *for negative shocks on leaders, i.e.* $d\log \rho_{L,1} < 0$*,*

$$\Delta_{y_1}(\rho_{L,1}) \quad > \quad 0.$$

The lemma reveals that the impact effect depends on how does the shock alter the market power of the leader. In the short run, the endogenous markup amplifies the shocks to the followers, while mitigates the shocks biased to the leader. Intuitively, as shown in equation (34), the elasticity of demand of the leader is a decreasing function of her market share. Therefore, if the shock is on the followers, the market share of the leader will increase in equilibrium. Consequently, her elasticity of demand drops and, therefore, she would increase markup endogenously. Since $\frac{\partial \log y_t}{\partial \log \kappa_t} < 0$, the increase of the markup suppresses the aggregate output by raising the efficiency wedge. Thereafter, compared to the MC, the aggregate output decreases by more. Conversely, the shock biased to the leader is eventually dampened by the decreased markup.

As shown by equation (47), the transitional effect is consisted of the wealth-markup channel and the wealth-accumulation channel. We first focus on the wealth-accumulation channel. Similarly, the difference in the first order effect of the shocks on the deviations of wealth across the two models is only because of the endogenous markup. Define by $\Delta_{m_{F,t}}$ the difference in log-deviation of wealth, i.e. $\Delta_{m_{F,t}} = \mathrm{d}\log m_{F,t} - \mathrm{d}\log \tilde{m}_{F,t}$. By log-linearization of equation (22) around steady state, we obtain that

$$\Delta_{m_{F,t}} = \left(\underbrace{\frac{\partial \log m_F'}{\partial \log m_F} + \frac{\partial \log m_F'}{\partial \log \kappa}\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log m_F}}_{\text{through wealth}}\right)^{t-1}\left(\underbrace{\frac{\partial \log m_F'}{\partial \log \rho_j} + \frac{\partial \log m_F'}{\partial \log \kappa}\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log \rho_j}}_{\text{through initial shock}}\right)\mathrm{d}\log\rho_{j,1} \tag{48}$$

$$-\left(\frac{\partial \log m_F'}{\partial \log m_F}\right)^{t-1}\frac{\partial \log m_F'}{\partial \log \rho_j}\mathrm{d}\log\rho_{j,1} \tag{49}$$

Note that $m_{F,t}$ is the only state variable for the economy. Therefore aggregate deviation of the system throughout the transition dynamics can be characterized by $\sum_{t=2}^{\infty}\Delta_{m_{F,t}}$. The higher $\sum_{t=2}^{\infty}\Delta_{m_{F,t}}$ implies the wealthier followers. Lemma 6 explores the effect through the wealth-accumulation channel.

Lemma 5 *If the within-sector elasticity of substitution is large enough such that*

$$\epsilon > \frac{\kappa+1}{2}\frac{1}{1-\alpha}+1, \tag{50}$$

depending on how does the shock alter the market power of the dominant firm, the endogenous markup has differentiated effects on the accumulation of the wealth of the followers. Particularly,

1. *for negative shocks on followers, i.e.* $\mathrm{d}\log\rho_{F,1} < 0$, $\forall t \geq 2$,

$$\Delta_{m_{F,t}}(\rho_{F,1}) > 0,$$

and, therefore,

$$\sum_{t=2}^{\infty}\Delta_{m_{F,t}}(\rho_{F,1}) > 0;$$

2. *for negative shocks on leaders, i.e.* $\mathrm{d}\log\rho_{L,1} < 0$, $\forall t \geq 2$,

$$\Delta_{m_{F,t}}(\rho_{L,1}) < 0,$$

and, therefore,

$$\sum_{t=2}^{\infty}\Delta_{m_{F,t}}(\rho_{L,1}) < 0.$$

Condition (50) is a sufficient condition such that $\frac{\partial \log m'_F}{\partial \log \kappa} > 0$ and $\frac{\partial \log m'_F}{\partial \log \rho_L} < 0$. Intuitively, if the elasticity of substitution between the output of followers and leaders is high enough, the raise of the markup would benefit the followers by increasing the relative demand and thus raising the profits of followers. Likewise, with the elasticity of substitution high enough, after a positive shock on leaders, more resources would be allocated away from followers, which decreases the followers' profit and wealth.

Lemma 5 explores the influence of market concentration on the wealth of the followers. Specifically, for the shocks on the followers, the market concentration mitigates the deviation of the wealth. Intuitively, the negative shock on the followers increases the market power of the leader. So that the leader endogenously raise her markup. When the output of the leader and the followers are highly substitutable, i.e. equation (50) satisfies, the raise of the markup of the leader increases the profits of the followers. Therefore, the economy is less deviated along with the transition dynamics. Furthermore, one can show that $\left|\frac{d \log m'_F}{d \log m_F}\right| < \left|\frac{d \log \tilde{m}'_F}{d \log \tilde{m}_F}\right|$. The increased markup accelerates the speed of convergence. Conversely, with the negative shocks on the leaders, the wealth of the followers is expanding in this extreme case. Due to the loss of market power, the leader declines her markup, which dampens the expansion of the wealth of the followers[5].

Note that the transitional effect of the endogenous markup is composed by the wealth-markup channel and the wealth-accumulation channel:

$$\Delta_{y_t} = \underbrace{\underset{(+)}{\frac{\partial \log y}{\partial \log \kappa} \frac{\mathrm{d} \log \kappa}{\mathrm{d} \log m_F}} \mathrm{d} \log m_{F,t}}_{\text{wealth-markup channel}} + \underbrace{\underset{(+)}{\frac{\partial \log y}{\partial \log m_F}} \Delta_{m_{F,t}}}_{\text{wealth-accumulation channel}} . \tag{51}$$

Combining with lemma 5, we can show that the two channels work in the opposite directions. [6]. To be specific, suppose there exists a negative shock on the followers, i.e. $\mathrm{d} \log \rho_{F,1} < 0$. Following lemma 5, the increased markup dampens the contraction of $m_{F,t}$ by raising followers' profits, which implies that $\Delta_{m_{F,t}} > 0$. The stronger balance sheets enables the higher borrowing capacity and, therefore, the wealth-accumulation channel mitigates the decline on output. Yet, note that the shock on the followers means that $\mathrm{d} \log m_{F,t} < 0$. The leader still has greater market power in the transitional process so that throughout the transition dynamics, her markup is higher than that of the representative unconstrained firm in the MC. Therefore, the wealth-markup channel amplifies the shock in the propagation process. In aggregate, the transitional effect depends on the trade-off and is not analytically tractable. The quantitative analysis shows that the wealth-markup channel always dominates.

To summarize, by comparing the dynamics of the model of strategic competition to the mo-

[5]The impulse shock on leaders here serves an extreme example to help us interpret the implication of strategic competition after more generalized shocks, that although is biased to leaders, instead of the wealth expansion in the example, the wealth of the followers also declines. The example implies that the decreased markup dampens the rebuild of the balance sheet of the followers, which decelerates the speed of convergence.

[6]Lemma 5 gives the sign of $\Delta_{m_{F,t}}$. To show that the the two channel has opposite effects, one has to show the sign of the elasticity $\frac{\partial \log y}{\partial \log \kappa}$, $\frac{\mathrm{d} \log \kappa}{\mathrm{d} \log m_F}$ and $\frac{\partial \log y}{\partial \log m_F}$, which are given in the derivations of lemma 4.

nopolistic competitive benchmark, we answer the question what is the implication of the strategic competition between large and small firms on business cycles. We disaggregate the effect of the endogenous markup into the impact effect and the transitional effect. Particularly, we find that the implications of the endogenous markup are non-monotonic in the following two dimensions. First, the implications of the increase and the decrease of the markup are exactly opposite. Therefore, it implies that, the effect of the strategic competition is conditional on how does the shock alter the market power of the leader. Furthermore, we find that for the transitional effect, the wealth-markup channel and the wealth-accumulation channel work in opposite direction. Although quantitative analysis show that the wealth-markup channel dominates, given that the wealth-accumulation channel characterizes the speed of convergence, it implies that the endogenous markup has opposite effects on the decline of output and the speed of recovery.

4.3.2 Analysis of a Forward-looking Leader

In the previous section, with the assumption of a myopic leader, we derive the specific solution of the linearized system and analyze the dynamics after impulse shocks on technology. In particular, we have the explicit solution for the elasticity of markup with respective to state variables, which is intractable when the leader is forward-looking. However, we derive a sufficient condition such that the effect through the endogenous markup remains consistent by characterizing the sign of the elasticity of markup with respective to aggregate states.

From equations (24)-(30), we can find the Eular equation of the leader:

$$\left(1-\kappa_t^{-1}\right)^{-1} = \upsilon_t + \frac{1}{1+r}\frac{\pi_{L,t+1}}{\pi_{L,t}}\left(\hat{\upsilon}_{t+1}-\upsilon_{t+1}\right) + \frac{1}{1+r}\frac{\pi_{L,t+1}}{\pi_{L,t}}\left(1-\kappa_{t+1}^{-1}\right)^{-1}, \tag{52}$$

where υ_{t+1} and $\hat{\upsilon}_{t+1}$ are the static and forward-looking elasticity respectively at time $t+1$. Lemma 6 characterizes the steady state markup charged by a forward-looking leader.

Lemma 6 *The steady state elasticity of forward-looking leader is given by*

$$\Upsilon = \upsilon + \frac{\hat{\upsilon}_1}{r}, \tag{53}$$

$$\upsilon = \epsilon - (\epsilon-\sigma)\left[1+\left(\mu_L^{-1}-1\right)\frac{\sigma+\frac{\gamma}{\alpha}-1}{\epsilon+\frac{\gamma}{\alpha}-1}\right]^{-1}, \tag{54}$$

$$\hat{\upsilon}_1 = \frac{\frac{\gamma}{\alpha}\left[1-\beta(1+r)\right](\epsilon-\sigma)^2(1-\mu_L)\mu_L}{\left(\sigma+\frac{\gamma}{\alpha}-1+(\epsilon-\sigma)\mu_L\right)^2} > 0. \tag{55}$$

It turns out that when the leader is forward-looking, same as the model of a myopic leader, the steady state discounted elasticity is a function of her market share. Furthermore, since $\hat{\upsilon}_1 > 0$, by internalizing the effect of raising prices on the wealth accumulation of followers, the forward-looking leader has higher elasticity of demand compared to the myopic leader. Therefore, the markup charged by the forward-looking leader is lower.

Lemma 7 *There exists $\underline{\epsilon}$ such that for all $\epsilon \geq \underline{\epsilon}$, $\frac{d\log\kappa}{d\log m_F} < 0$, $\frac{d\log\kappa}{d\log\rho_{F,1}} > 0$, $\frac{d\log\kappa}{d\log\rho_{L,1}} < 0$.*

Lemma 7 provides a sufficient condition such that the effect through the channel of endogenous markup remains consistent with the case where the leader is myopic. Lemma 7 is derived from the two dimensional system of difference equations given by the Eular equation (52) and the law of motion (30). It provides a sufficient condition that if the within-sector elasticity of substitution is high enough, the leader would increase markup when its market share is increasing and decrease markup vice-versa. Specifically, the $\underline{\epsilon}$ is pinned down by $\frac{\mathrm{d}\log\hat{\upsilon}_1(\epsilon)}{\mathrm{d}\log\mu_L} = 0$. For all $\epsilon > \underline{\epsilon}$ the one-period forward-looking elasticity is a decreasing function of leaders' market share, i.e. $\frac{\mathrm{d}\log\hat{\upsilon}_1(\epsilon)}{\mathrm{d}\log\mu_L} < 0$.

Note that at steady state, the discounted present value of the forward-looking elasticity is a geometric summation of the one-period forward-looking elasticity. The sufficient condition provided by lemma 8 could be interpreted as the condition to guarantee the discounted present value of the forward-looking elasticity is a decreasing function with respective to the market share of the leader. Intuitively, conditional on the within-sector elasticity of substitution high enough, the market leader has more monopoly power so that she is less sensitive to the growth of its potential competitors[7] and more inclined to raise markup. The sufficient condition guarantees the discounted elasticity of demand for the leader declines with the increase of the leader's market share. Thus, the effect of the endogenous markup as we discussed in the previous section remains consistent.

To conclude, we derive a sufficient condition such that the main mechanism of the endogenous markup remains same in the model of a forward-looking leader. Although we cannot derive the explicit solutions for the elasticity of markup with aggregate states, the sufficient condition, which requires that the within-sector elasticity high enough, guarantees the sign of the effects of endogenous markup remains consistent.

5 Quantifying the model

In this section, we calibrate the model. We first quantify the log-analysis in section 4. We find that the endogenous markup amplifies the shocks on followers while dampens the shocks on leaders. Furthermore, because the leaders are in general more productive and less constrained, compared to the shocks biased to the leaders, the effect of the strategic competitions are more significant when the shocks are on the followers.

Given that credit crunches are isomorphic to the productivity shocks that are biased to the constrained, we compare the effect of the endogenous markup on credit crunches as in Buera and Moll (2015). We find that the effect of the credit crunch would be largely underestimated if we do not take into consideration the strategic competition. More specifically, because the resources are re-allocated to the less constrained while productive firms, we show that without the strategic

[7] $\underline{\epsilon}$ is only a function of μ_L, within-sector elasticity and Pareto parameters: $\underline{\epsilon} = \left(\mu_L^{-1} - 1\right)\left(\sigma + \frac{\gamma}{\alpha} - 1\right) - \left(\frac{\gamma}{\alpha} - 1\right)$. Note that $\underline{\epsilon}$ is a increasing function of σ. Intuitively, when the between-sector substitution is more elastic, the intermediate firms are more inclined to substitute the output of the leader with the product of other sectors. Subsequently, the market leader is less dominated within the sector: it simultaneously faces more fierce competitions both within and between sector substitution. Therefore, it requires a higher lower bound of within-sector elasticity to guarantee that the leader is less sensitive to the wealth accumulation of followers.

competition, the effect of the shocks on collateral constraints is minor. Yet, the effect of the resource allocation is overturned by the increasing of the markup. With the strategic competition, the economy features a greater decline in output, TFP and the demand for capital.

We generalize our analysis by considering the standard TFP shocks that homogeneously decreases the demand for the whole economy. Due to financial frictions, the small firms are more cyclically sensitive throughout the transition dynamics. Consequently, the quantitative experiment show that for the homogeneous shock, the strategic competition amplifies the aggregate decline on output yet accelerates the recovery.

Heretofore, the paper focuses on the question that for a given level of concentration, how does the strategic competition impacts business cycles. We enrich the analysis by exploring what is the implications of the deepening of concentration, where we compares the dynamics of the economy between the models that the leaders are heterogeneous in productivity and, therefore, the economies are differentiated in the steady state concentration. The analysis features the trade-off between the effect from the change in the endogenous markup and the effect from the change in the productivity distribution. It shows that in aggregate, the effect from productivity dominates. The concentration mitigates the homogeneous productivity shock because resources are reallocated to the more productive leaders. Yet, the more responsive endogenous markup offsets the effect of the raised productivity by 81%.

5.1 Parameterization

We start by explaining the parameterization strategy of the model.

Assigned parameters The paper normalizes the lower bound of the Pareto distribution of the productivity of entrepreneurs by 1. Technology parameters ρ across the economy are normalized to 1 at steady state. Following Buera and Moll (2015), the discount factor β is set to 0.95 , the interest rate is set as $r = 0.02$ and the borrowing capacity is set to be $\lambda = 2$. The capital share of production function is set at $\alpha = 1/3$. The within-industry and the between-industry elasticities of substitution are set at $\epsilon = 10$ and $\sigma = 1.01$ which is consistent with the values of Atkeson and Burstein (2008).

Calibrated parameters The Pareto parameter γ is calibrated to 1.149 to match the distribution of incomes of the firms by COMPUSTAT. It implies the steady state productivity cutoff $\underline{z}$ is 6.269. Productivity of the forward-looking leaders is calibrated as $z_{i,L} = 18.245$ [8], which implies that the average markup of the economy at steady state is 1.255. ω is set be 0.2 to match the population share of entrepreneurs in the United States.

[8] The productivity of the myopic leader is calibrated to $\tilde{z}_{i,L} = 21.547$ to match the revenue share (50%) of large firms in the United States. The productivity of representative unconstrained firm in the MC is set to be same as that of the strategic leader, i.e. $z_L = \tilde{z}_L$. Together with the assumption that they charges the identical markup in steady state, it implies that the steady state allocations and prices across benchmark and the strategic models are identical. Therefore the difference in output highlights the effect from endogenous markup.

Table 4 summarizes the assigned and calibrated parameters together with the related moments to match.

5.2 Quantifying the Log-analysis

The log-analysis in section 4 implies that the effect of the strategic competition is non-monotonic and depends on how does the shock changes the market power of leaders. We first quantify the log-analysis. To be more specific, we are considering the agent-biased shocks on the followers and the leaders respectively, i.e. $\mathrm{d}\log \rho_{F,t} < 0$ or $\mathrm{d}\log \rho_{L,t} < 0$, that are symmetric to all the sectors in the economy. In particular, the shock recovers according to the following AR(1) process:

$$\rho_{i,t} = (1-\delta)\rho_i + \delta\rho_{i,t-1}, \quad \delta \in (0,1) \tag{56}$$

where δ represents the persistence of the shock. To shed lights on the endogenous markup, following the same logic as the previous section, we compare the dynamic paths of the economy between the MC benchmark and the model with the strategic competition.

Figure 3 plots the evolution of resources and prices along with the biased productivity shock. Consistent with the prediction of the log-analysis, endogenous markup amplifies the shock to followers while mitigates the shock to leaders. More specifically, for the shock biased to followers, the market power of the leader is enhanced. Consequently, the raised markup amplifies the shock throughout the transition dynamics, while dampens the contraction of the wealth of followers by promoting their profits. Contrariwise, the effect of the endogenous markup is reversed after the shock on leaders.

Furthermore, the quantitative results shows that when the shock is biased to followers, the decline of the output in the MC is minor compared to that after the shock biased to leaders. Hence, the effect of strategic competition is more significant when the shock is biased to the followers:

$$\sum_{t=1}^{\infty}\left|\frac{\mathrm{d}\log y_t\left(\rho_{F,t}\right)-\mathrm{d}\log \tilde{y}_t\left(\rho_{F,t}\right)}{\mathrm{d}\log \tilde{y}_t\left(\rho_{F,t}\right)}\right| > \sum_{t=1}^{\infty}\left|\frac{\mathrm{d}\log y_t\left(\rho_{L,t}\right)-\mathrm{d}\log \tilde{y}_t\left(\rho_{L,t}\right)}{\mathrm{d}\log \tilde{y}_t\left(\rho_{L,t}\right)}\right|. \tag{57}$$

The shock biased to followers reallocates resources from the followers to the leaders, who are generally less constrained and more productive. Subsequently, when the leader does not behave strategically, the shock is largely mitigated[9]. Yet, the positive effect of the resource allocation is overturned by the strategic behaviors. The accumulated effect of the increased markup significantly amplifies the shock. By contrast, for the productivity shock that is biased to leaders, the effect of from the endogenous markup is less significant.

The significance of the effect of the strategic competition on the shocks is conditional on the within-sector elasticity. When the within-sector elasticity is higher, the more resources would be reallocated to the leaders and, hence, the more significant the effect of the increased markup would

[9]Intuitively, when within-sector elasticity high enough, or between-sector elasticity low enough, the shock on followers could raises the output in the MC. Because the higher elasticity of substitution between leaders and followers promotes the resource re-allocations. Yet, strategic behaviors still would overturn the effect of reallocation

be. The between-sector and the within-sector elasticity of substitution are set as $\sigma = 1.5, \epsilon = 10$ to be consistent with Atkeson and Burstein (2008). In particular, the survey of Anderson and Van Wincoop (2004) shows that the within-sector elasticity is likely in the range of 5 to 10. In appendix, we re-calibrate $\epsilon = 5$ and replicate the exercises of the biased productivity shock, where the strategic competition's effect on the shocks biased to the followers is less significant, yet still remarkably greater than that on the shocks biased to leaders.

5.3 Credit Crunches

In this section, we explore the implication of the strategic competition on credit crunches. Buera and Moll (2015) shows that by distorting the investment decisions of entrepreneurs, credit crunches are isomorphic to the productivity shocks on the constrained. Consequently, the productivity shock biased to followers shown in section 5.2 can be interpreted as a credit crunch. To compare with credit crunches and the shocks biased to followers, following Buera and Moll (2015), we impose a persistent shock on collateral constraints following the process given by equation (56).

Figure 4 plots the evolution of the real allocations along with the after a shock on collateral constraints. The credit crunch distorts the investment decisions and, therefore, lowers the aggregate productivity of followers. Additionally, the experiment shows that the distortion is amplified by the endogenous markup. The raised markup of the leaders suppresses the equilibrium wage by decreasing labor demand, which in further lowers the productivity cutoffs of entrepreneurs. Subsequently, consistent with the case where the shock is biased to followers, the accumulated effect of the strategic competition is significant. In particular, in the initial periods, the final output of the MC is even slightly increasing, compared to the significant decline in the model of strategic competition.

The increased capital in the MC explains the mechanism behind. In the MC, given that the shock constrains the borrowing capacity of the followers, it reallocates the resources to the more productive representative leader. Because the representative leader is unconstrained, the reallocation increases the capital demand of the economy. The increased output of the unconstrained firms offsets the negative effect from the contraction of collateral constraints. However, this is reversed by the strategic competition. With higher market power, the strategic leader exploits the demand by increasing markup and decreasing supply. Contrarily, the aggregate demand of capital is declining in the strategic model. The accumulated effect the endogenous markup overturns the positive effect from the resource allocations and hereby, amplifies the effect of credit crunch. The experiment emphasizes that, in a concentrated economy where firms are heterogeneous in market power and borrowing capacities, the effect of the credit crunch on marginal firms would be largely underestimated if we do not take firms' strategic behaviors into considerations.

5.4 Homogeneous productivity shocks

To highlight the mechanism behind the strategic competition, we impose the shocks that are agent-specific. In this section, we generalize our analysis by exploring the dynamics after a homogeneous shock on the productivity that is persistent and symmetric to all the sectors. In particular, we assume that the economy stays at steady state at $t = 0$. At time $t = 1$, there is a exogenous shock on productivity $\rho_{i,1} < 1$ that is symmetric across intermediate sectors and following the process in equation (56).

The shock is homogeneous since it decreases the aggregate demand for intermediate good sectors. Yet, due to financial frictions, the effect of the shock is still biased. Figure 5 plots the evolution of the economy along with the shock. At the beginning, the market share of the leader is lower than the steady state value. Consistent with Ottonello and Winberry (2020), the relative marginal cost of unconstrained firms is more elastic compared to the constrained entrepreneurs[10]. The increased relative marginal cost decreases demand for leaders. So that at the beginning, the effect of the shock is biased to leaders. Consequently, the decreased markup mitigates the shock initially. However, along with the transition dynamics, the effect of the shock is gradually transferred to followers because of financial frictions. The contraction of the balance sheets of followers gradually increases the market power of leaders and eventually overturns the initial negative effect of the shock. Along with the transition dynamics, the leader is increasing markup which amplifies the shock. Yet, corresponding to the effect on aggregate output, the effect on the wealth of followers is reversed. Initially, the wealth of followers declines more because of the decreased markup. Afterwards, when the shock is transferred to followers, the increased markup accelerates the speed of convergence. As shown in the figure, the recovery of the wealth in the model of strategic competition exceeds that in the MC.

In aggregate, the amplification effect dominates. Figure 6 plots the accumulated effect of endogenous markup on output and wealth respectively as a function of the initial shock on productivity. Note that the the accumulated effect of the endogenous markup is even a convex function of the shock. Also, consistent with the analysis before, the endogenous markup raises the wealth of followers and hence accelerating the speed of convergence. More specifically, for the impulse shock $\rho_{i,1} = 0.7$ on productivity[11], the accumulated decline of the output is amplified by 6.7%, while the time taken for the recovery, measured by half-life, is reduced by 29.4%.

The exercise highlights the importance of the interactions between the strategic competition and financial frictions. Even if the shock is homogeneous and initially biased to leaders, because of financial frictions, the effect of the shock could be transferred to followers. Overall, because of financial frictions, the small firms are more cyclically sensitive. Although the endogenous markup mitigates the shock initially, throughout the transition dynamics, the strategic competition amplifies the decline of output yet accelerates the recovery.

[10] Note that, if we ignore the change of wage, the marginal cost of the leader is constant. While the demand shock crowds out the less productive followers, which increases the aggregate productivity and lowers the marginal cost of the followers, i.e. $E(z_i \mid \underline{z}_i)$. Hence the relative marginal cost of the leader is increasing.

[11] With $\rho_{i,1} = 0.7$, the initial drop of the output is roughly 10% in the model of strategic competition.

5.5 Increasing in concentration

Heretofore, we answer the question that given a magnitude of concentration, how does the strategic competition affect business cycles. In this section, we expand our analysis to explore the implications of the increasing in concentration. In particular, by separately calibrating the productivity of the leaders, we are comparing the dynamics between the economies with different steady state market shares after a homogeneous productivity shock[12]. The calibration captures a important feature highlighted by Autor et al. (2020) that the more concentrated the economy is, the higher the productivities of large firms are.

To investigate the implication of market concentration, for the given shock, we compare the aggregate deviation in output across the economies with the differentiated steady state market share of leaders, denoted by $\Delta_y^{\mu}(\mu_L) \equiv \sum_{t=1}^{\infty} \mathrm{d}\log y_t(\mu_L) - \sum_{t=1}^{\infty} \mathrm{d}\log y_t(\mu_L^*)$, where μ_L^* is the baseline magnitude of concentration calibrated to the United State. Note that $\mathrm{d}\log \tilde{y}_t$ denotes the deviation of the output in the MC, we decompose $\Delta_y^{\mu}(\mu_L)$ into the following two effects:

$$\Delta_y^{\mu}(\mu_L) = \underbrace{\Delta_y(\mu_L) - \Delta_y(\mu_L^*)}_{\text{difference in endogenous markup}} + \underbrace{\Delta_{\tilde{y}}^{\mu}(\mu_L)}_{\text{difference in productivity dist.}}, \tag{58}$$

where

$$\Delta_{\tilde{y}}^{\mu}(\mu_L) \equiv \sum_{t=1}^{\infty} \mathrm{d}\log \tilde{y}_t(\mu_L) - \sum_{t=1}^{\infty} \mathrm{d}\log \tilde{y}_t(\mu_L^*), \tag{59}$$

$$\Delta_y(\mu_L) \equiv \sum_{t=1}^{\infty} \mathrm{d}\log y_t(\mu_L) - \sum_{t=1}^{\infty} \mathrm{d}\log \tilde{y}_t(\mu_L). \tag{60}$$

To be specific, the effect of strategic competition, i.e. $\Delta_y(\mu_L)$, is differentiated across the economies with different magnitudes of concentration. Hence the first effect of concentration comes from the heterogeneous effect of the endogenous markup: $\Delta_y(\mu_L) - \Delta_y(\mu_L^*)$. Moreover, the productivity distributions are heterogeneous across the economies that are differentiated in concentration. Note that the representative unconstrained firm in the MC always charges the constant markup. We isolate the effect from productivity distribution through the difference of the deviations in output across the MCs with different magnitudes of concentration.

Figure 7 plots the effect of the increasing in concentration. It shows that the increased concentration mitigates the homogeneous productivity shock. Furthermore, by decomposition, the figure shows that the effects from the difference in the endogenous markup and the difference in productivity distribution are opposite. Note that in figure 6, we show that due to financial frictions, followers are more sensitive to the homogeneous shock. Throughout the transition dynamics, the shock reallocates the resources from followers to leaders. Given that the productivity of the leader

[12] The homogeneous shock is identical to the shock described in the section 5.3. The reason we use the homogeneous shock here is because we are comparing the economies that are heterogeneous in the steady state market shares of the leaders and the followers. Therefore, the biased shock has heterogeneous effects across the economies with different magnitudes of concentration.

is increasing in concentration, consequently, with the more resources re-allocated to leaders, the raised productivity mitigates the shock[13].

On the contrary, the decomposition shows that the effect from the difference in the endogenous markup offsets that of the increased productivity. As plotted in figure 8, when the economy is more concentrated, the markup of the strategic leader is more sensitive to the shock. Note that the markup of the strategic leader is determined by $\kappa = \Upsilon/(\Upsilon - 1)$. It implies that $\mathrm{d}\log\kappa/\mathrm{d}\log\Upsilon = -(\kappa - 1)$: the higher markup charged by leaders is, or equivalently, the more concentrated the economy is, the more elastic the markup to the change of elasticity of demand would be. Therefore, when the shock alters the elasticity of demand, the strategic leader in the highly concentrated economy responses more, hence the effect of the endogenous markup is more significant. Given that the accumulated effect of the endogenous markup amplifies the homogeneous productivity shock, the amplification effect would be more significant when the markup is more responsive.

In aggregate, figure 7 shows that the mitigation effect from the improved productivity distribution dominates. Yet, the results re-emphasize the importance of strategic competitions: if we do not take into consideration the effect of endogenous markup, we would overestimate the implication of the increasing of concentration on homogeneous shocks by 81.49%.

6 Policy implications

Because of financial frictions, the economy is inefficient. In the perfect market, the borrowing capacity of the entrepreneurs is not constrained, where the most productive entrepreneur gathers all the resources from the less productive firms while the latter choose to save since the market rate of the bond is higher than their rate of return if choose to be active. Therefore, Buera and Moll (2015) argued that the credit crunch is isomorphic to a TFP shock since it distorts the entrepreneurs' investment decisions. Conversely, as discussed by Itskhoki and Moll (2019), the government could optimize the allocations of resources by subsidizing the constrained entrepreneurs. In this section, we are exploring how does the optimal subsidization is twisted by the strategic competition.

To be more specific, we are focusing on the implications of endogenous markup on the optimal permanent interest rate subsidization. In particular, we investigate the optimal policy in two scenarios. First, we explore how does the strategic competition alter the steady state optimal interest rate cut. Furthermore, by imposing shocks on the wealth of the entrepreneurs, we investigate what is the response of the optimal policy responds to the shocks and in further how does the strategic competition distorts the optimal policy responses after the shock.

6.1 The uniform interest rate subsidization

We start with the standard policy of the uniform interest rate cut that is applied to all the active firms within the economy. We assume that, financed by the lum-sum tax on households, the

[13] The homogeneous productivity shock re-allocates resources to the more productive leaders. Contrariwise, if the shock is biased to leaders, in the more concentrated economy, the resources are re-allocated away from the leaders with higher productivity. As a result, the shock would be amplified.

utilitarian government maximizes the discounted welfare of the households and entrepreneurs by subsidizing the interest rate payment of the active entrepreneurs and large firms. The government cannot directly re-allocate wealth across different individuals while it internalizes the pricing strategies of the leaders and the saving functions of each individual entrepreneurs.

Given some initial distribution $G_0(a,z)$ of entrepreneurs, the utilitarian government chooses the subsidy rate τ_r to maximize

$$\sum_{t=0}^{\infty}\beta^t\left[(1-\omega)\,u_h\,(c_t)+\omega\int u_f\,(c_f(a,z))\,\mathrm{d}G_t\,(a,z)\right], \tag{61}$$

subject to

$$c_{h,t}+b_{t+1} = w_t l_{h,t}+(1+r)\,b_t+T_t, \tag{62}$$

$$T_t = r\tau_r\left(k_{L,t}+\int_{\underline{z}} k_{j,t}dG_t\,(a,z)\right), \tag{63}$$

together with all the market clearing conditions and the policy functions provided by Lemma 1 to Lemma 4. With the subsidization, the interest rate paid by the active firms is given by

$$r^* = r\,(1+\tau_r)\,. \tag{64}$$

For simplicity, equation (63) implies that we assume the government is clearing its budget constraint each period.

6.1.1 The optimal interest subsidization at steady state

We start by the analysis of the optimal steady state policy. By assuming the economy starts from the steady state without policy intervention, i.e. $G_0(a,z)=G^*(a,z;0)$, we explore what is the optimal subsidization τ_r throughout the transitions from the initial state to a new steady state, i.e. $G^*(a,z;\tau_r)$. Consistent the analysis before, to highlight the effect of endogenous markup, we compare the optimal intervention with strategic competition to the case where the leader has the identical technology yet behaves monopolistic competitively.

Figure 9 displays the optimal subsidization of the utilitarian government. The disaggregation of the welfare of entrepreneurs and households are plotted as in figure 10. The optimal interest rate implied by the two models are distinct. For the benchmark, the general interest rate cut is favored by the government. In particular, even with the negative transfer, the welfare of the households is improving with the subsidization. Yet, for the entrepreneurs, in aggregate they are worse off even if being subsidized. While the government prefers raising the interest rate in the model with strategic competitions.

We start by explaining the result of the benchmark. The interest rate subsidization has distinct implications on the unconstrained firms and the constrained firms. Since the constrained firms always borrow up-to-limit, the interest rate cut does not enhance the borrowing capacity of the constrained firms directly. If we do not take into account the effect on the equilibrium prices, the policy is isomorphic to a direct transfer from the households to the constrained firms. In particular, the marginal cost of the constrained firms is inelastic to the interest rate cut. However, for the unconstrained firms, the decrease of the interest rate distorts the their demand of capital which lowers their marginal cost. Therefore, consistent with Ottonello and Winberry (2020), the large firms are more responsive to the interest rate cut because they are unconstrained and have a flatter marginal cost curve. It implies that the uniform interest rate cut is biased to the large firms in our model.

Moreover, since the marginal cost of the large firm is more elastic, the policy distorts the demand of the output within the intermediate good sectors and reallocates resources from the small firms to the large firms. In particular, when the elasticity of substitution within-sector is high enough[14], the interest rate cut could even decreases the wealth of the followers. Intuitively, the more substitutable the outputs are, the more sensitive the relative demand is to the change of the relative marginal cost. When the substitutability is high enough, the aggregate wealth of the followers is decreasing because of the decline in the demand. Consistent with the analysis, figure 10 shows that the wealth of the followers in the benchmark is declining. However, the decline of the entrepreneurs' welfare is offset by the welfare improvement of the households. With more resources reallocated to the more productive unconstrained firms, the labor demand is increasing which benefits the households. In aggregate, the economy prefers the uniform subsidy in the benchmark.

Yet, everything is overturned by the strategic competition. As shown by figure (11), since the policy is by its nature benefiting the large firms, they are raising markups, which reverses the welfare of both the entrepreneurs and the households. First, the increased markup raises the relative demand for the followers. Therefore, different from the benchmark, the wealth and the welfare of the entrepreneurs are improving along with the subsidization. However, the raised markup decreases the welfare of the households by suppressing the labor demand and, therefore, the equilibrium wage. With endogenous markup, instead of the subsidy, the government prefers the taxing on the interest rate.

We draw two main lessons from the analysis of the steady state policy. First, because the marginal cost of unconstrained firms are more elastic, the uniform interest rate cut might not benefit the targeted constrained firms. Second, the exercise emphasizes the importance of endogenous markup in the welfare analysis. We would largely underestimate the fiscal cost to implement the policy, if we do not internalize the effect of endogenous markup. More specifically, endogenous markup has differentiated implications on the welfare of the entrepreneurs and the households: (i) because of the raising markups, the followers could still benefit from the policy that is biased to the leaders; (ii) on the contrary, the raised markup largely amplifies the welfare cost of the households to implement the policy.

[14] The lower bound of the elasticity of substitution is given by $\underline{\epsilon} = \frac{(\kappa+1)}{2}\frac{(2\gamma-1)}{1-\alpha} + 1$.

6.2 The biased interest rate subsidization

The analysis of the uniform interest rate cut implies that the large firms actually benefit more from the policy, because their marginal cost is more elastic to the factor prices. When the within-sector elasticity of substitution is high enough, the constrained entrepreneurs might even be worse off because the policy re-allocates resources to the unconstrained large firms. Therefore, it motivates us to consider the policy that biased to the constrained entrepreneurs.

To be more specific, similarly, we assume that given some initial distribution of entrepreneurs, i.e. $G_0(a, z)$, there exists an utilitarian government who maximizes the social welfare function (61). Yet, we assume that the government can only subsidize the interest rate payment of the entrepreneurs. Therefore, the government follows the budget constraint

$$T_t = r\tau_r \int_{\underline{z}} k_{j,t} dG_t(a, z). \tag{65}$$

The other constraints remains consistent, except that the leader is not subsidized while the rent of the capital for the active followers is given by

$$r^* = r(1 + \tau_r). \tag{66}$$

6.2.1 The optimal steady state policy

We first investigate how does endogenous markup alters the optimal steady state policy. Similarly, we are considering the transition dynamics from the initial steady state distribution with $\tau_r = 0$, i.e. $G_0(a, z) = G^*(a, z; 0)$, to the new steady state $G(a, z; \tau_r)$ Figure 12 plots the discounted utility of the utilitarian government and the optimal biased interest rate cut. Different from the uniform interest rate subsidization, both in the benchmark and in the model of strategic competition, the optimal policy is to subsidize the entrepreneurs. Furthermore, compared to the benchmark, the optimal rate of subsidization in the model of strategic competition is much higher. Additionally, the disaggregation of the welfare plotted by figure 13 shows that endogenous markup minimizes the welfare cost to implement the policy: even if there exists the negative transfer from the households to the followers, the welfare of the households could still be improved.

By plotting the prices and allocations in the new steady state, figure 14 explains the effect of endogenous markup. Because the policy is biased to entrepreneurs, the market share of the leaders is declining. Consequently, the leaders are decreasing their markup. The declined markup mitigates the cost to implement the policy in two dimensions. First, it decreases the welfare cost to implement the policy. The decline of the markup raises the labor demand and, therefore, the equilibrium wage, which minimizes the welfare cost of the households. In particular, even with the negative net transfers, the welfare of households could still be improved through the increase of the wage. Second, the declined markup offsets the negative effect from the resources allocations. The biased subsidization has the negative effect on the aggregate productivity since it reallocating resources from the more productive leaders to the less productive followers. Yet, the declined markup offsets the negative effect from the resource allocations.

To conclude, compared to the benchmark, endogenous markup mitigates the welfare cost of the biased subsidization, suggesting a more aggressive subsidization to the entrepreneurs. Furthermore, compared to the uniform interest rate cut, the economy is favoring the biased policy. In particular, the welfare improvement provided by the biased policy is 3 times higher than that given by the uniform subsidization.

6.2.2 The optimal stabilization policy

Deviating from the steady state policy, in this section, we investigate how does the concentration distorts the optimal response of the policy to the initial shock on the wealth of the followers. Figure 15 compares the response of the optimal policy to the benchmark. It turns out in both cases, after the negative shock on the entrepreneurs, the optimal response of the utilitarian government is to raise the subsidization. Define by $\left|\frac{\tau_r(x_0)-\tau_r^*}{\tau_r^*}\right|$ the optimal policy response after the initial shocks x_0. The figure shows that the optimal subsidization in the model with strategic is less sensitive to the shocks compared to the benchmark, i.e. $\left|\frac{\tau_r^*(x_0)-\tau_r^*(1)}{\tau_r^*(1)}\right| < \left|\frac{\tilde{\tau}_r^*(x_0)-\tilde{\tau}_r^*(1)}{\tilde{\tau}_r^*(1)}\right|$, where τ_r^* represents the optimal rate of subsidy and $x_0 < 1$ is the initial shock on wealth.

In further, by the disaggregation of the social welfare, figure 16 reveals that the change of the welfare of the entrepreneurs contributes to the insensitivity in the policy response. Compared to the benchmark, it features a minor difference in the welfare improvement of the economy with the initial shock, compared to the economy starting from the steady state, i.e., for $x_0 < 1$,

$$\frac{\frac{\Delta U^e(\tau_r;x_0)}{U_0^e(x_0)} - \frac{\Delta U^e(\tau_r;1)}{U_0^e(1)}}{\frac{\Delta U^e(\tau_r;1)}{U_0^e(1)}} < \frac{\frac{\Delta \tilde{U}^e(\tau_r;x_0)}{\tilde{U}_0^e(x_0)} - \frac{\Delta \tilde{U}^e(\tau_r;1)}{\tilde{U}_0^e(1)}}{\frac{\Delta \tilde{U}^e(\tau_r;1)}{\tilde{U}_0^e(1)}}, \tag{67}$$

where U_0 denotes the welfare of the entrepreneurs with $\tau_r = 0$. Given the less significant welfare improvement on entrepreneurs, the optimal policy intervention responses by less.

Figure 17 explains how does endogenous markup distorts the policy response. Note that the markup of the strategic leader is given by $\kappa = \frac{\Upsilon}{\Upsilon-1}$. It implies that $\left|\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}\right| = \kappa - 1$: the elasticity between the markup and the elasticity of demand is a increasing function in the markup. Because of the negative wealth shock on the followers, the market power and the markup of the leader is higher, which implies that her markup is more elastic to her elasticity of demand compared to that at steady state. When the government subsidies the entrepreneurs, it relatively decreases the market power, and thus, the elasticity of the demand of the leader. With higher elastic markup, the leader declines the markup by more, compared to the steady state. Consistent with the analysis figure 17 plots the reaction of markup to the subsidy rate at the first period of the shock. Compared to the steady state, with the negative shock on followers, the leader's markup reacts by more, while in the benchmark, the markup is constant. Consequently, the lowering markup distorts the profits of the entrepreneurs by more, which constrains the effect of the policy invention.

To summarize, by evaluating the optimal interest rate subsidization, we draw three main findings.

First, the uniform interest rate subsidy is by its nature biased to the large firms, since they are less constrained so that their marginal cost is more elastic to the factor prices. Consequently, the increasing in markup distorts the uniform intervention by depressing the equilibrium wage and raising the welfare cost. Second, the distortion of the welfare cost suggests that the economy is preferring the more detailed and agent-biased policy. The biased policy not only direct improves wealth of the constrained firms, but also minimizes the cost of implementation by lowering the markup of the leaders and increasing the equilibrium wage. Finally, endogenous markup diminishes the subsidization to the entrepreneurs during the crisis, suggesting that the government should be more conservative in the interest rate cut to smooth the shocks on the entrepreneurs.

7 Empirical studies

The log-linear analysis and the numerical experiments provide the following three main implications that we empirically test in this section. The model implies that, given a magnitude of concentration,

1. in the short run, the shock is amplified when it increases the market share of the leader, while it is mitigated if decreases the market share of the leader;

2. in the long run, by raising the profits of the followers the recovery is accelerated when the shocks are biased to followers, while it is slowed down when the shocks are biased to leaders.

3. Furthermore, the increase in concentration mitigates the shocks that are biased to followers, while amplifies the shocks biased to leaders[15].

7.1 Data sources and measurement

This section summarizes the data sources and measurements of the empirical studies.

7.1.1 Data sources

To grasp a sketch of the dynamics of business cycles, the empirical analysis draws on the balance sheet reports of North-America public-traded firms by the Compustat for the years 1980-2018, which includes the historical data of annual fundamentals for all traded firms across the economy. We first introduce our definitions of the sectors and the large firms.

Our model based on the assumption that outputs are highly substitutable within-sector, which implies the definition of the sector should be narrow enough. We define the sector in six-digit North American Industry Classification System (NAICS) level covering more than 1200 different sectors. Furthermore, the median number of firms within a sector by the Census 2012 is 328. The Compustat covers much less firms than the Census. To distinguish the large firms from the relatively smaller

[15]In the quantitative part, we focus on the homogeneous productivity shock whose effect is biased to the followers because of financial frictions. The concentration mitigates the shock because the shock reallocates the resources to the more productive leaders in the more concentrated economy. Contrariwise, when the shock reallocates the resources away from leaders, the concentration amplifies the shock.

firms in the Compustat, and to trace the cyclicality of the market share of large firms along with business cycles, the exercise excludes the sectors with the number of firms less than 10. The refined sample includes 312 sectors. The median number of firms within each sector is 32.

The empirical study requires a time-consistent definition of large and small firms. We define large firms as the top-2 firms within each sector measured by total asset. On average, the defined large firms take up 54% of market share within each sector, which is in general consistent with the market share of large firms in US[16]. The other firms are defined as small firms.

7.1.2 Measurement

The effect of the endogenous markup depends on how does the shock locally changes the market power of the leaders. A crucial variable connecting empirical works and the model is the measure of the change of the market share of the large firms. Denote by $s_{i,j,t}$[17] as the annual sales of the firm with jth amount of total asset in the sector i during the year t, the market share and the change of the market share are then defined as

$$\mu_{i,L,t} = \frac{\sum_{j=1}^{2} s_{i,j,t}}{\sum_{j=1}^{N_{i,t}} s_{i,j,t}}, \tag{68}$$

$$\Delta\mu_{i,L,t} = \mu_{i,L,t} - \mu_{i,L,t-1}, \tag{69}$$

where by $N_{i,t}$ we denote the total number of firms within the sector i.

It worth to mention that since the Compustat only contains the balance sheet of the active public-traded firms, the market share measured might not be that of the actual leaders of the sector: leaders could be non-publicly traded. More importantly, the change of the market share we defined are susceptible to the market operations including while not limited to, list and delist, or merger and acquisition. Therefore, to reduce the measurement error, the sample precludes the sectors with different top two firms for consecutive years, which precludes the influence of leader's market operations such as list or delist.

Another key variable to link the model is the growth rate of output. Consistent with the measurement of market share, the growth rate of the sector i at year t is measured by the log difference of the aggregate sales:

$$g_{i,t} = \ln s_{i,t} - \ln s_{i,t-1}, \tag{70}$$

$$s_{i,t} = \sum_{j=1}^{N_{i,t}} s_{i,j,t}, \tag{71}$$

[16] Following Crouzet and Mehrotra (2020), the large firms are defined to be 1% firms in assets. According to the Census, the aggregate market share of the top 1% firms is 54.9%

[17] The reports of balance sheet provided by Compustat are based on firms' fiscal years. All the values are the linearly adjusted to calendar years according to the fiscal-year of the firm.

Finally, the sales, together with all the other subsequent control variables, are normalized to the real values of the year 1980 by Producer Price Index (PPI) of all commodities. The PPI is modified to annual value by linear average.

7.2 Empirical test of the short-run effect

In this section we test the short run implication of endogenous markup on business cycles. The model implies that the influence of endogenous markup depends on how does the shock alters the market share of the leaders. Specifically, since the elasticity of demand of the strategic leader is decreasing with respective to the leader's market share, the leader would charge a higher markup when her market share is increasing, which raises the efficiency wedge and amplifies the shock. Conversely, if the market share of the leader declines after the shocks, the decreased markup mitigates the contraction of the economy. Therefore, the model implies a negative relationship between the difference in the market share of the leader $\Delta\mu_{i,t}$ and the demeaned growth rate of the sector $\hat{g}_{i,t}$, which are given by

$$\hat{g}_{i,t} = g_{i,t} - \bar{g}_t, \tag{72}$$

$$\bar{g}_t = \frac{1}{M_t}\sum_{i=1}^{M_t} g_{i,t}, \tag{73}$$

where denote by M_t the number of the sectors of the year t.

Figure 18 plots the relationship between $\Delta\mu_{i,L,t}$ and $\hat{g}_{i,t}$ over the periods 1980-2018. Many outliers though it has, in general the plot reveals a negative relation. By precluding the data with different leaderships for the consecutive two years, the sample excludes the effect from the list or delist of large firms. Yet, the sample does not exclude the effects of the list or delist of small firms, which would mechanically imply the negative relation. Consider a case where the output of the existing firms holds exactly same as the previous year, the new listed small firms mechanically increases the aggregate output while decreases the market share of leaders. Thereafter, further detailed analysis is required and in particularly we should control the size of the sector.

We estimate the model predicting sector growth rate of the form

$$\hat{g}_{i,t} = \delta_{i,t} + \beta_0\Delta\mu_{i,L,t} + \beta_1\mu_{i,L} + \beta_2 \times X_{i,t} + e_{i,t}, \tag{74}$$

where the dependent variable is the demeaned growth rate of the sector i at time t. $\mu_{i,L}$ is the mean of the market share of the sector i. Note that our analysis of the effect of endogenous markup is based on the fixed magnitude of concentration. Therefore, we control the concentration by the mean of market share of each sector. $X_{i,t}$ is the vectors of controls related to the size effect and standard error $e_{i,t}$ are clustered at the sector level. Since the dependent variable is annually demeaned, the independent variables does not contain the time fixed effect. We first estimate the model by pooling the data in 1980-2018, followed by the estimation using the two sub-samples of 1980-2000 and 2001-2018.

Table 5 confirms that for all three samples, there exists a robust negative relation between the increasing in market share of leaders and the demeaned growth rate of the sector. The positive coefficient in front of the difference in number of firms confirms the mechanical relation between the sector growth rate and the number of small firms. By controlling the difference in number of firms, the regression excludes the influence of the list or delist behaviors of small firms. The negative coefficient of total asset can be interpreted as decreasing return. Finally, the model implies that the effect of the endogenous markup is more significant when the market leader has a higher steady state market share. In our regression, the weighted mean of leaders' market share in the period of 2001-2018 is 54.82%, higher than that in the period 1980-2000 (43.13%). Consistent with the prediction of the model, the regression shows that, with the more concentrated economy, during the period 2001-2018, the growth rate responses more to the change of leaders' market share.

7.3 Empirical test of the long-run effect

In this section we test the transitional effect of endogenous markup through the wealth-accumulation channel. The model predicts that, given a magnitude of concentration, if the shock increases the market share of leaders, the raised markup promotes the profits of followers, which smooths the spread of the shock. Vice-versa for the shock that decreases the market share of leaders. We test the prediction from the two dimensions, across-crisis and within-crisis.

First of all, between different crisis, for the crisis that is biased to leaders (characterized by the local decrease of the market share of leaders), the model implies that the crisis is more persistent and has prolonged recovery process. Contrariwise, the recovery process of crisis is less persistent when the market power of the leader is enhanced during the crisis.

Figure 19 plots the recent three main crisis: the 2001 recession, the financial crisis and the crisis of the COVID-19. The three crisis are heterogeneous in the detrended market share of large firms. During the 2001 recession, the market share of leaders declines, while in the crisis of COVID-19, there exists a significant sign of increasing in market share of leaders, implying that the shock during COVID-19 is biased to followers. The financial crisis is measured in between the other two crisis. Initially, the market share of the leader is below the trend, but is constantly increasing during the crisis. The trend of market share measured by Compustat is in general consistent with the measured change documented by Crouzet and Mehrotra (2020) in figure 2 with the confidential QFR data: the market concentration increased during the financial crisis while decreased in the recession 2001. Consistent with the model's prediction, with the increase of the market share, the real GDP and sales after the crisis of COVID is bouncing back very quickly compared to the other two crisis. Furthermore, compared to the financial crisis, although the recession 2001 is interpreted as a shorter crisis, it takes roughly similar period of time for the full recovery. In particular, the sales of the followers declines even further and takes more time to recover. The evolution of the sales of the followers is consistent with that of liability. Although we do not observe the vast bankruptcy and the short of liquidity as the financial crisis, the liability of the followers during the recession 2001 declines by a lot, compared to the minor change of that of the leaders. The trend of the liability is consistent with the transitional effect of endogenous markup. As the shock is biased to leaders, the declined markup suppresses the rebuilding of the balance sheet of followers which constrains their borrowing capacities.

In further, we complement our analysis by testing the relation between the change of the leaders' market share and followers' future growth rate of equity and income within the period of financial crisis.

Figure 20 plots the relation between the future demeaned growth rate of the followers i and the current change of market share of the leaders. The first and the second row confirm the predictions of the transitional effect, by showing the positive relation between the current change of the leaders' market share and the followers' growth rate of equity and income in next period. In further, the regression shown in table 6 the positive relation is statistically significant. Except the standard controls as in previous sections, the regression additionally controls a term of lags to rule out the possibility that the equity and income of followers are mean-reverting. In particular, we find that there exists a significant mean-reverting process of the demeaned growth rate of the sales, but not of followers' equity and income. With the p-values around 5%, the positive relation plotted in figure 20 cannot be rejected. Finally, note the the table shows that the relation between the change of market share and the future growth of the sector is insignificant, which is also consistent with our analysis. Note that the wealth-markup channel and the wealth-accumulation channel are opposite. If the shock is biased to the leader, the raised markup although mitigates the spread of the shock by raising the profits of the followers, the direct effect of raising markup increases the efficiency wedge and amplifies the shock. Therefore, the relation between the future growth rate and the change of the market share depends on the trade-off. Consistent with the prediction, despite the significantly negative relation between $\Delta_{m_{i,L,t}}$ and $\hat{g}_{i,t}$ in the short run empirical analysis, the relation between $\Delta_{m_{i,L,t}}$ and the future growth rate $\hat{g}_{i,t}$ is insignificant.

7.4 Empirical test of the effect of concentration

The model predicts that increasing in concentration mitigates the shock that biased to followers. Because the resources are re-allocated to leaders and the productivity of leaders is a increasing function of the degree of concentration. Contrariwise, the shock biased to leaders are amplified by concentration, because that the shock re-allocates resources away from leaders. Therefore, the model predicts the relation between $\hat{g}_{i,t}$ and $\mu_{i,L,t}$ should conditional on the sign of $\Delta\mu_{i,L,t}$. Consequently, we estimate the model of the form

$$\hat{g}_{i,t} \quad = \quad \delta_{i,t} + \beta_0 \mathbb{1}_{\Delta\mu_{i,L,t}} + \beta_1 \mu_{i,L,t} + \beta_2 \mathbb{1}_{\Delta\mu_{i,L,t}} \cdot \mu_{i,L,t} + \beta_3 \times X_{i,t} + e_{i,t}, \tag{75}$$

where $X_{i,t}$ contains other controls and $\mathbb{1}_{\Delta\mu_{i,L,t}}$ is a indicator function:

$$\mathbb{1}_{\Delta\mu_{i,L,t}} \quad = \quad \begin{cases} 1, & \Delta\mu_{i,L,t} > 0, \\ 0, & \Delta\mu_{i,L,t} \leq 0. \end{cases} \tag{76}$$

Therefore, $\beta_1 + \beta_2$ represents the effect of concentration when the shocks are biased to followers; β_1 shows the situation when it biased to leaders. Table 7 confirms the predictions of the concentration. First, $\mu_{i,L,t}$, $\mathbb{1}_{\Delta\mu_{i,L,t}}$ and their product are jointly significant. Furthermore, $\beta_1 + \beta_2 > 0$ implies that increasing in concentration mitigates the shocks that are biased to followers. While $\beta_1 < 0$ confirms the prediction that increasing in concentration amplifies the shocks biased to leaders.

8 Conclusion

The paper analyzes the effect of market concentration on business cycles, by proposing a model featuring the dynamic Stackelberg game between a large firm who plays as a leader, and a continuum of heterogeneous entrepreneurs who are financially constrained and behaving as followers. In particular, we first answer the question that given a degree of concentration, how does the strategic competition between large and small firms affects business cycles. We find that the effect of the strategic competition is non-monotonic. It is conditional on how does the shock alters the market power of the large firm, and, therefore, how does her markup respond to the shock. Although the strategic competition mitigates the shock biased to leaders, it significantly amplifies the shocks such as credit crunches that are biased to followers because of the increased markup of large firms. For the homogeneous productivity shocks, due to financial frictions, followers are more cyclically sensitive. Consequently, the strategic competition amplifies the aggregate decline of output yet accelerates the recovery by increasing the profits of followers.

Furthermore, we explore the implication of the increasing in concentration on business cycles, where the analysis focuses on the trade-off between the increased productivity and the more responsive markup of leaders. We find that the effect from the raised productivity dominates. For homogeneous productivity shocks, the increasing in concentration mitigates the output decline because the shock reallocates resources towards the more productive leaders. However, the markup of the leader is more responsive to the shock, which offsets the effect of the raised productivity by 81%.

We explore how does the strategic competition alters the optimal stabilization policy of interest rate cut. We find that because the marginal cost of large firms are more elastic, the uniform interest rate cut is by its nature benefiting large firms. Hence the raised markup increases the welfare cost of complementing the policy by suppressing the equilibrium wage. For the interest rate cut that is biased to followers, we show that the government should be more conservative in the cut because the strategic competition distorts the demand and lowers the profits of followers.

We test the three main implications of the model with the Compustat data. We first confirm the short run implication of the strategic competition by showing the negative relation between the demeaned growth rate of a sector and the change of market share of the sector's leader. We test the long-run implication of the strategic competition by showing the significant positive relation between the change of leader's market share and the future growth rate of followers' equity and income. Finally, we confirm the effect of the increasing in concentration by showing that the effect of the increasing in market share of leaders on the demeaned growth rate is positive when the shock is biased to followers; the effect is negative when it is biased to leaders.

References

Anderson, J. E. and Van Wincoop, E. (2004). Trade costs. *Journal of Economic literature*, 42(3):691–751.

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

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2):645–709.

Baqaee, D. and Farhi, E. (2020). The darwinian returns to scale. Technical report, National Bureau of Economic Research.

Bernanke, B. S., Gertler, M., and Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. *Handbook of macroeconomics*, 1:1341–1393.

Bianchi, J. (2016). Efficient bailouts? *American Economic Review*, 106(12):3607–59.

Boar, C. and Midrigan, V. (2019). Markups and inequality. Technical report, National Bureau of Economic Research.

Brunnermeier, M. K., Eisenbach, T. M., and Sannikov, Y. (2012). Macroeconomics with financial frictions: A survey.

Brunnermeier, M. K. and Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, 104(2):379–421.

Buera, F. J. and Moll, B. (2015). Aggregate implications of a credit crunch: The importance of heterogeneity. *American Economic Journal: Macroeconomics*, 7(3):1–42.

Caglio, C. R., Darst, R. M., and Kalemli-Özcan, á. (2021). Risk-taking and monetary policy transmission: Evidence from loans to smes and large firms. Technical report, National Bureau of Economic Research.

Chen, H., Dou, W., Guo, H., and Ji, Y. (2020). Feedback and contagion through distressed competition. *The Rodney L. White Center Working Papers Series at the Wharton School, Jacobs Levy Equity Management Center for Quantitative Financial Research Paper.*

Chodorow-Reich, G., Darmouni, O., Luck, S., and Plosser, M. (2021). Bank liquidity provision across the firm size distribution. *Journal of Financial Economics.*

Corbae, D. and D'Erasmo, P. (2019). Capital requirements in a quantitative model of banking industry dynamics. Technical report, National Bureau of Economic Research.

Crouzet, N. and Mehrotra, N. R. (2020). Small and large firms over the business cycle. *American Economic Review*, 110(11):3549–3601.

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

Dorn, D., Katz, L. F., Patterson, C., Van Reenen, J., et al. (2017). Concentrating on the fall of the labor share. *American Economic Review*, 107(5):180–85.

Edmond, C., Midrigan, V., and Xu, D. Y. (2018). How costly are markups? Technical report, National Bureau of Economic Research.

Gertler, M. and Bernanke, B. (1989). Agency costs, net worth and business fluctuations. In *Business cycle theory*. Edward Elgar Publishing Ltd.

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

Gowrisankaran, G. and Holmes, T. J. (2004). Mergers and the evolution of industry concentration: results from the dominant-firm model. *RAND Journal of Economics*, pages 561–582.

Holmes, T. J. and Stevens, J. J. (2014). An alternative theory of the plant size distribution, with geography and intra-and international trade. *Journal of Political Economy*, 122(2):369–421.

Itskhoki, O. and Moll, B. (2019). Optimal development policies with financial frictions. *Econometrica*, 87(1):139–173.

Kudlyak, M. and Sanchez, J. M. (2016). Revisiting gertler-gilchrist evidence on the behavior of small and large firms. *FRB St. Louis Working Paper*, 2016-5.

Mongey, S. et al. (2017). Market structure and monetary non-neutrality. *Job market paper.*

Olmstead-Rumsey, J. (2019). Market concentration and the productivity slowdown. *Job market paper.*

Ottonello, P. and Winberry, T. (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica*, 88(6):2473–2502.

Wang, O. and Werning, I. (2020). Dynamic oligopoly and price stickiness. Technical report, National Bureau of Economic Research.

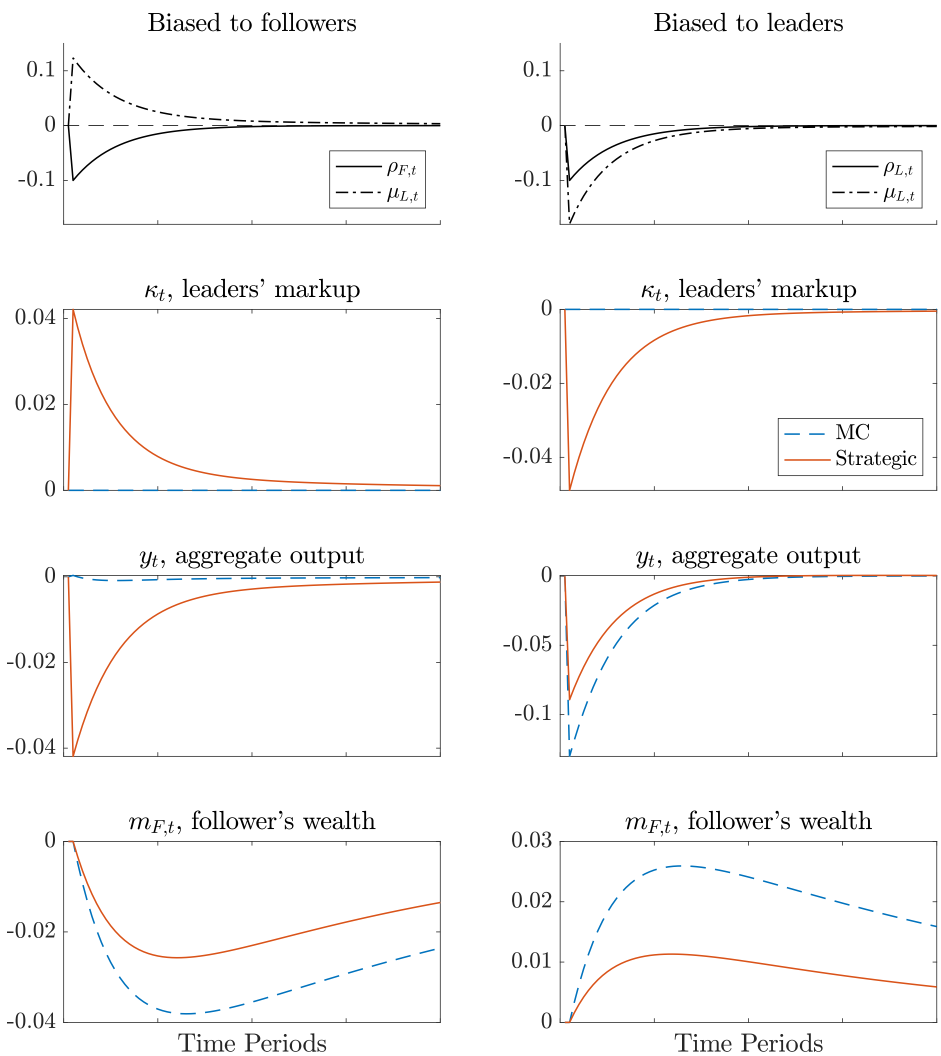
List of Figures

List of Tables

Note: The large firms are defined as the top 2 firms by assets within-industry to match the market share of 1% firms with the QFR data, while small firms are the rest of firms. Weight is given by the share of aggregate sales of each sector.

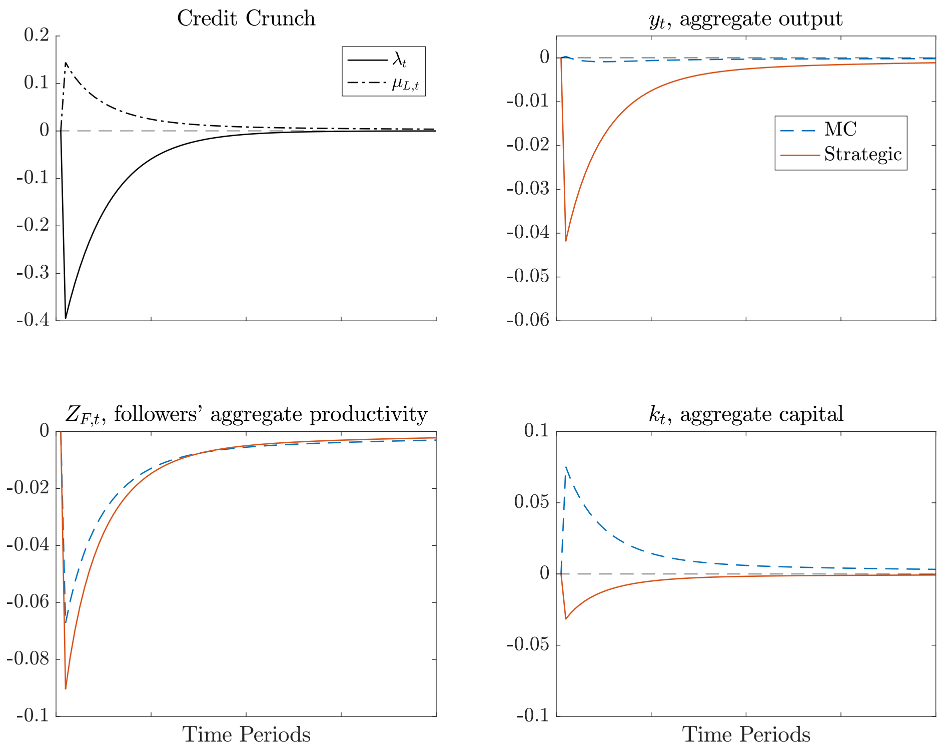
Figure 1: growth rate and markup along with cycles.

Figure 2: growth rate along with cycles, by Crouzet and Mehrotra (2020)



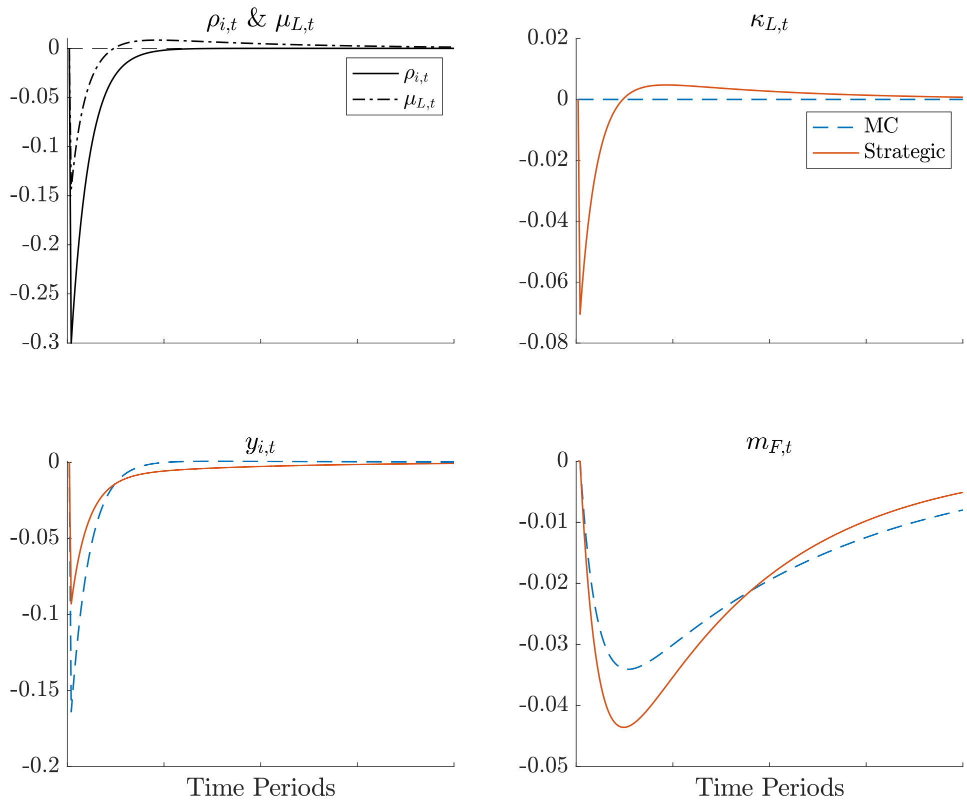
Note: the figure plots the evolve of the aggregate output, the market shares and the markup charged by the leaders. The shocks are designed to be agent-biased, persistent shocks on productivity of followers, i.e. $\rho_{F,1} < 1$, or on leaders, i.e. $\rho_{L,1} < 1$. The shocks are following AR(1) process with persistence $\delta = 0.9$ and the initial drops is given by $\rho_{j,1} = 0.9$, $j \in \{L, F\}$.

Figure 3: the evolve of real allocations and prices with a persistent agent biased technology shock.



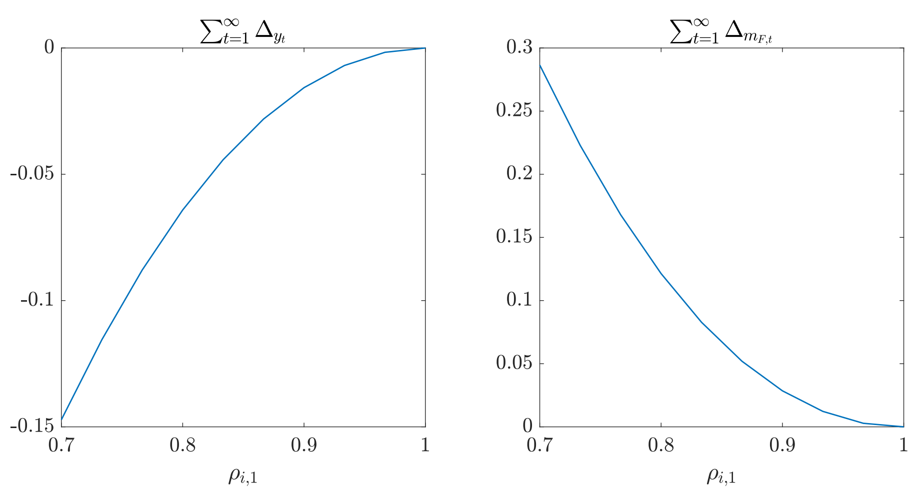
Note: the figure plots the evolve of the aggregate output, the market shares and the markup charged by the leaders. The shocks are designed to be agent-biased, persistent shocks on productivity of followers, i.e. $\rho_{F,1} < 1$, or on leaders, i.e. $\rho_{L,1} < 1$. The shocks are following AR(1) process with persistence $\delta = 0.9$ and the initial drops is given by $\rho_{j,1} = 0.9$, $j \in \{L, F\}$.

Figure 4: the evolve of real allocations and prices with a persistent agent biased technology shock.



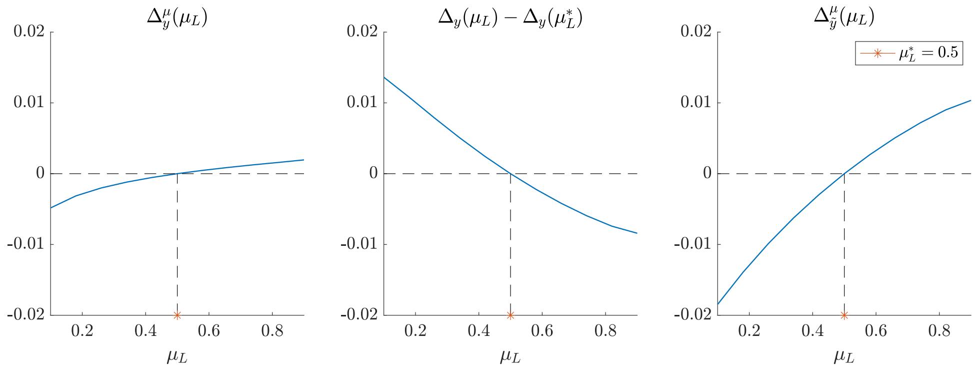
Note: the figure plots the dynamics of the economy after a homogeneous productivity shock. The shock is designed to be symmetric on productivity across all the sectors. The dashed-line represents dynamics of the monopolistic competitive benchmark.

Figure 5: dynamics of the economy with a homogeneous shock on productivity



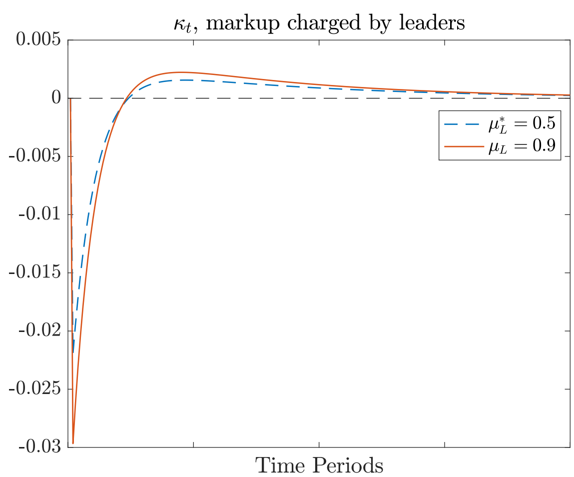
Note: the figure plots the aggregate effect on output and wealth after homogeneous productivity shocks. The shock is designed to be symmetric on productivity across all the sectors.

Figure 6: aggregate effect of the endogenous markup with alternative initial shocks on productivity



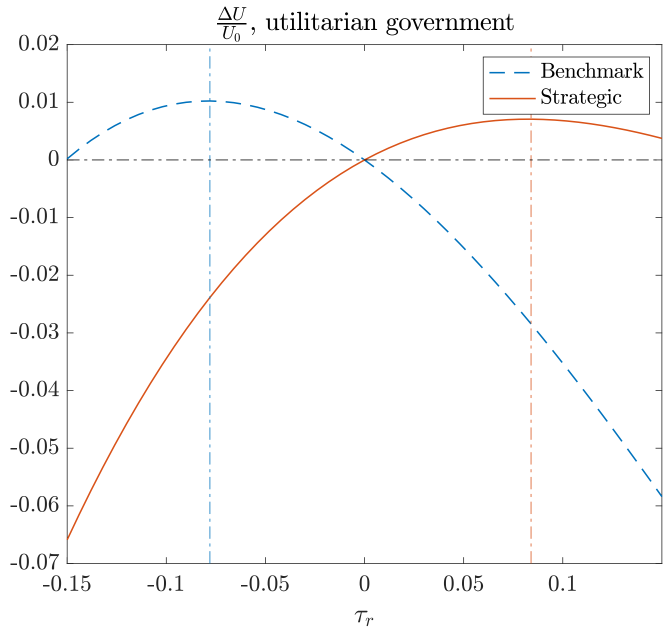
Note: the figure plots the aggregate effect of the change of concentration and decomposes the effect from the change in distribution and the change in endogenous markup. The x-axis plots the steady state market share of the leader.

Figure 7: the effect of the deepening of concentration after a homogeneous shock



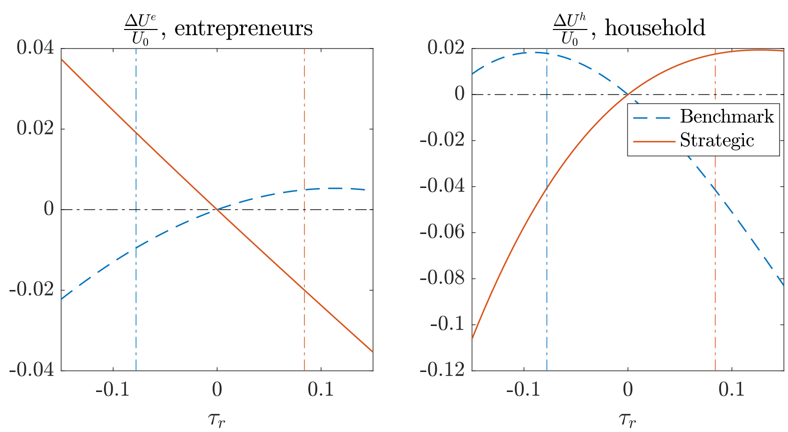
Note: the figure plots the change of markup charged by strategic leaders between the two economies with the different steady state concentrations after a homogeneous shock across all sectors.

Figure 8: the comparison of the endogenous markup between the economies with different concentration



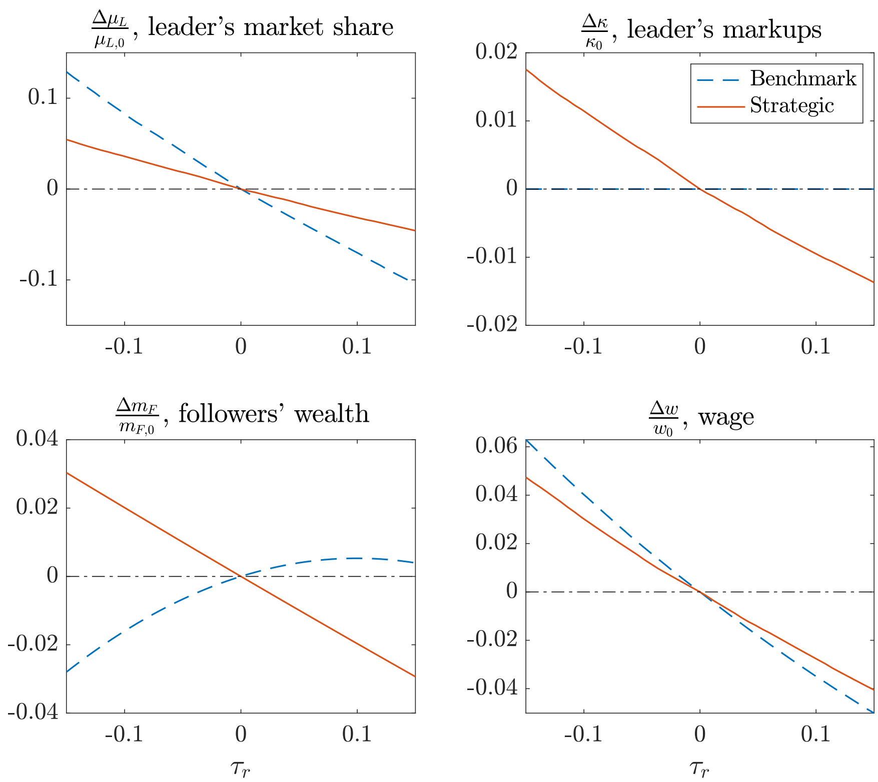
Note: the figure plots the discounted utility of the utilitarian government transitioned from the initial steady state to the new steady state with respective to the different subsidy rate on interest, i.e. τ_r. The dash-dotted line plots the optimal steady state subsidy rate of the model of concentrations and the competitive benchmark respectively.

Figure 9: the welfare of the utilitarian government transitioned from the steady state.



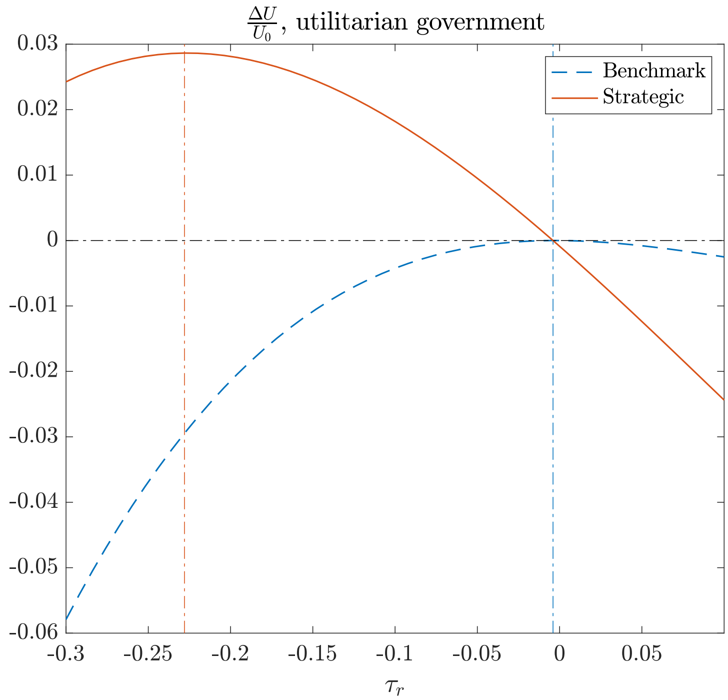
Note: the figure plots the discounted utility of entrepreneurs and households transitioned from the initial steady state to the new steady state with respective to the different subsidy rate on interest, i.e. τ_r. The dash-dotted line plots the optimal steady state subsidy rate of the utilitarian government of the model with concentrations and the competitive benchmark respectively.

Figure 10: the welfare of entrepreneurs and households with transitioned the steady state.



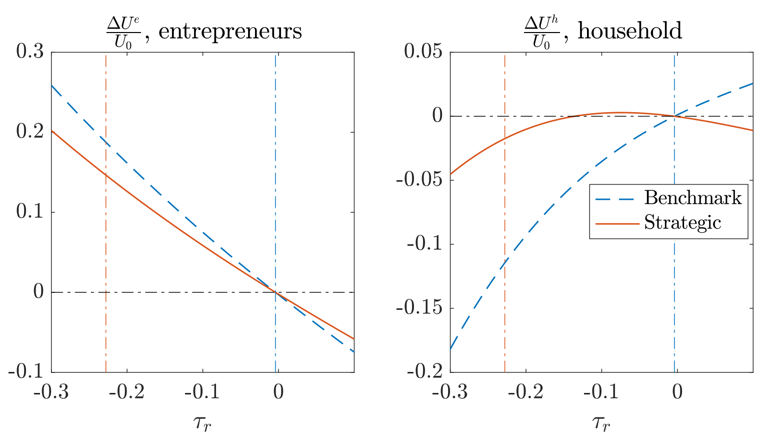
Note: the figure plots the prices and allocations at the new steady state with respective to the different subsidy rate on interest rate, i.e. τ_r.

Figure 11: the steady state prices and allocations with alternative τ_r.



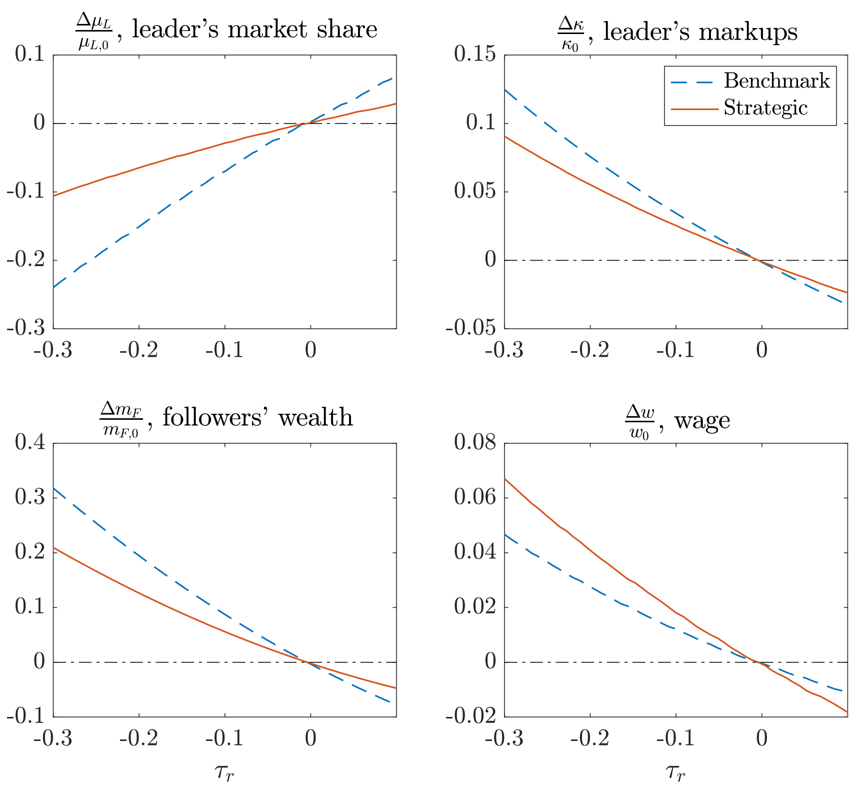
Note: the figure plots the discounted utility of the utilitarian government transitioned from the initial steady state, with respective to the different subsidies on interest rate that are biased to entrepreneurs, i.e. τ_r. The dash-dotted line plots the optimal steady state subsidy rate of the model of concentrations and the competitive benchmark respectively.

Figure 12: the welfare of the utilitarian government with subsidies on entrepreneurs transitioned from the steady state.



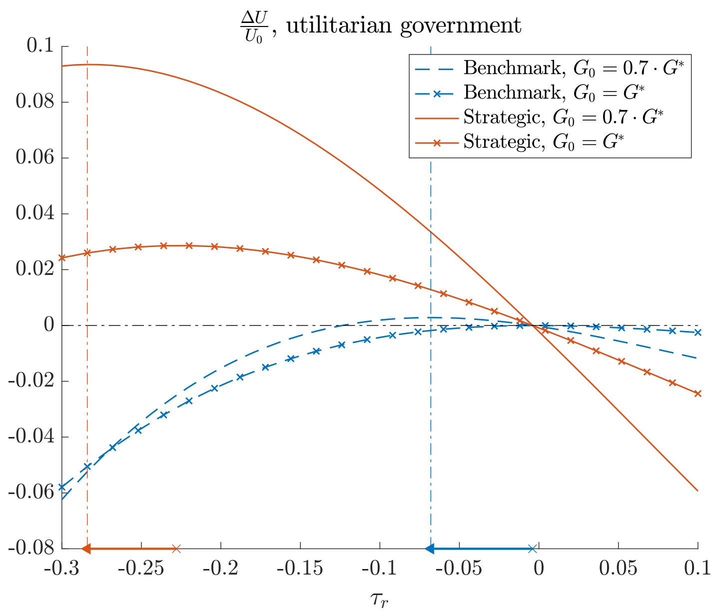
Note: the figure plots the discounted welfare of entrepreneurs and households transitioned from the steady state to the new steady state with respective to the different subsidies of interest, i.e. τ_r, that are biased to entrepreneurs. The dash-dotted line plots the optimal steady state subsidy rate of the utilitarian government of the model with concentrations and the competitive benchmark respectively.

Figure 13: the welfare of entrepreneurs and households with subsidies on entrepreneurs transitioned from the steady state.



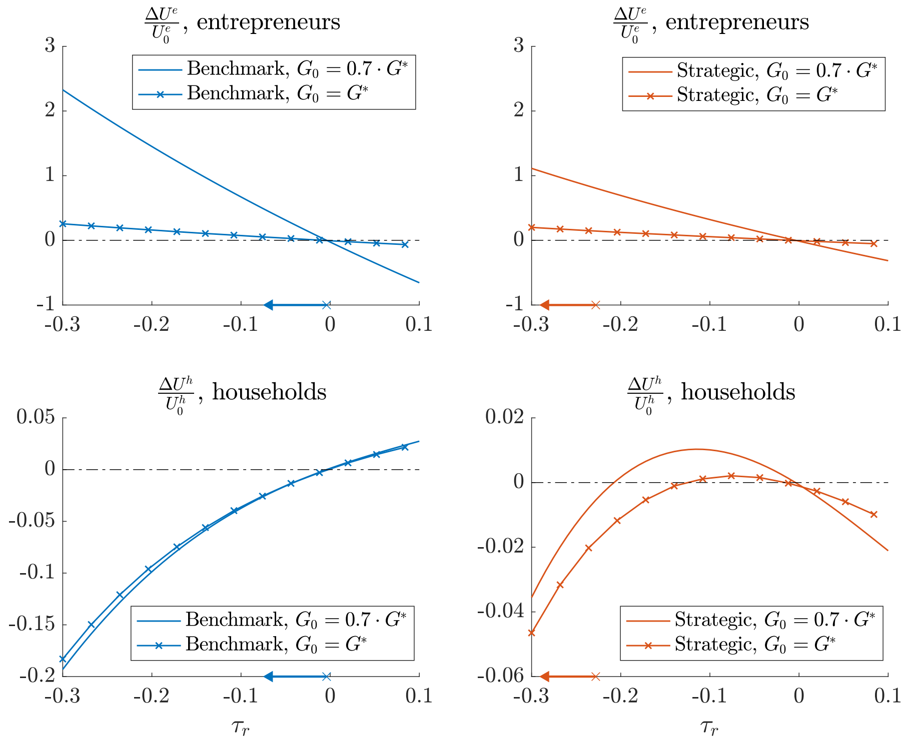
Note: the figure plots the prices and allocations at the new steady state with respective to the different subsidization on interest rate, i.e. τ_r, that is biased to entrepreneurs.

Figure 14: the steady state prices and allocations with alternative subsidies on entrepreneurs.



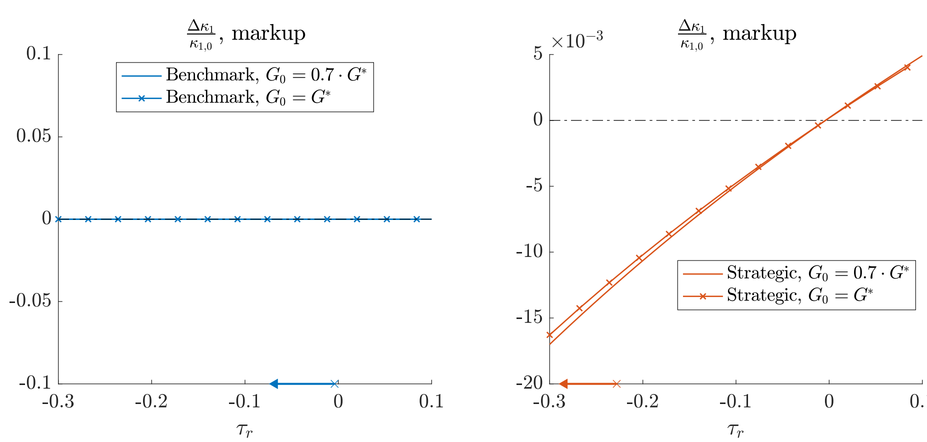
Note: the figure compares the discounted utility of the government between the case that the economy is transitioned from the steady state and the case from the state with an initial shock on the wealth of followers. The arrows on the x-axis show the change of the optimal rate of subsidization.

Figure 15: the welfare of the utilitarian government with alternative subsidies on entrepreneurs after an initial shock.



Note: the figure compares the discounted utility of entrepreneurs and households between the case that the economy is transitioned from the steady state and the case from the state with an initial shock on the wealth of followers. The arrows on the x-axis show the change of the optimal rate of subsidization because of the initial shock.

Figure 16: the welfare of entrepreneurs and households with alternative subsidies on entrepreneurs after an initial shock.



Note: the figure compares the markups of the unconstrained firms during the period of the shocks as a function of the subsidy biased to entrepreneurs. The arrows on the x-axis show the change of the optimal rate of subsidization because of the initial shock.

Figure 17: the change of the markup of the leaders at the first period of the shock

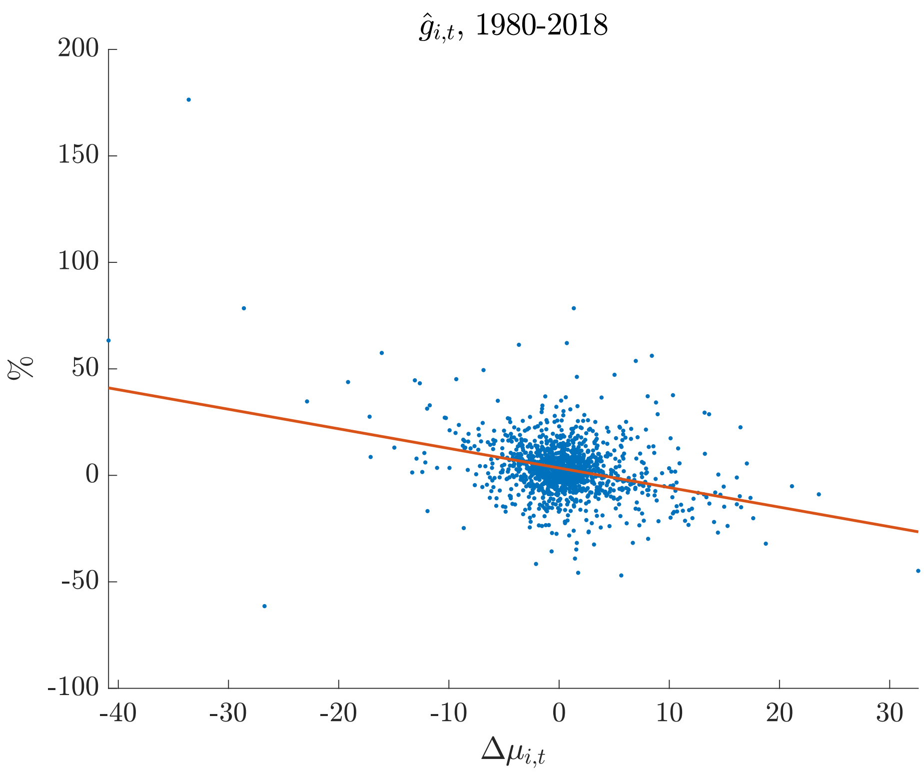


Figure 18: change of market share and growth rate

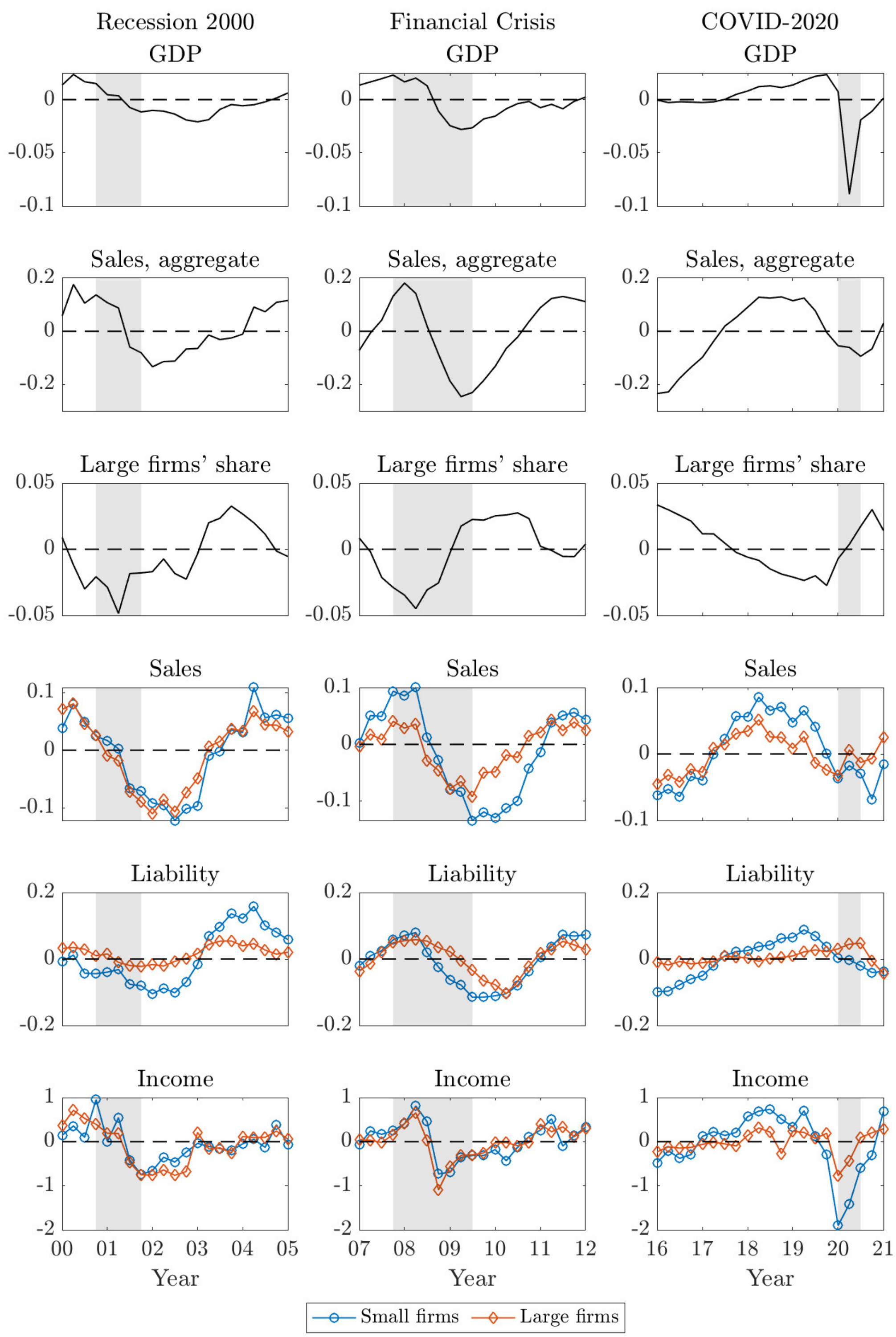


Figure 19: comparison between crisis, HP-filter detrended

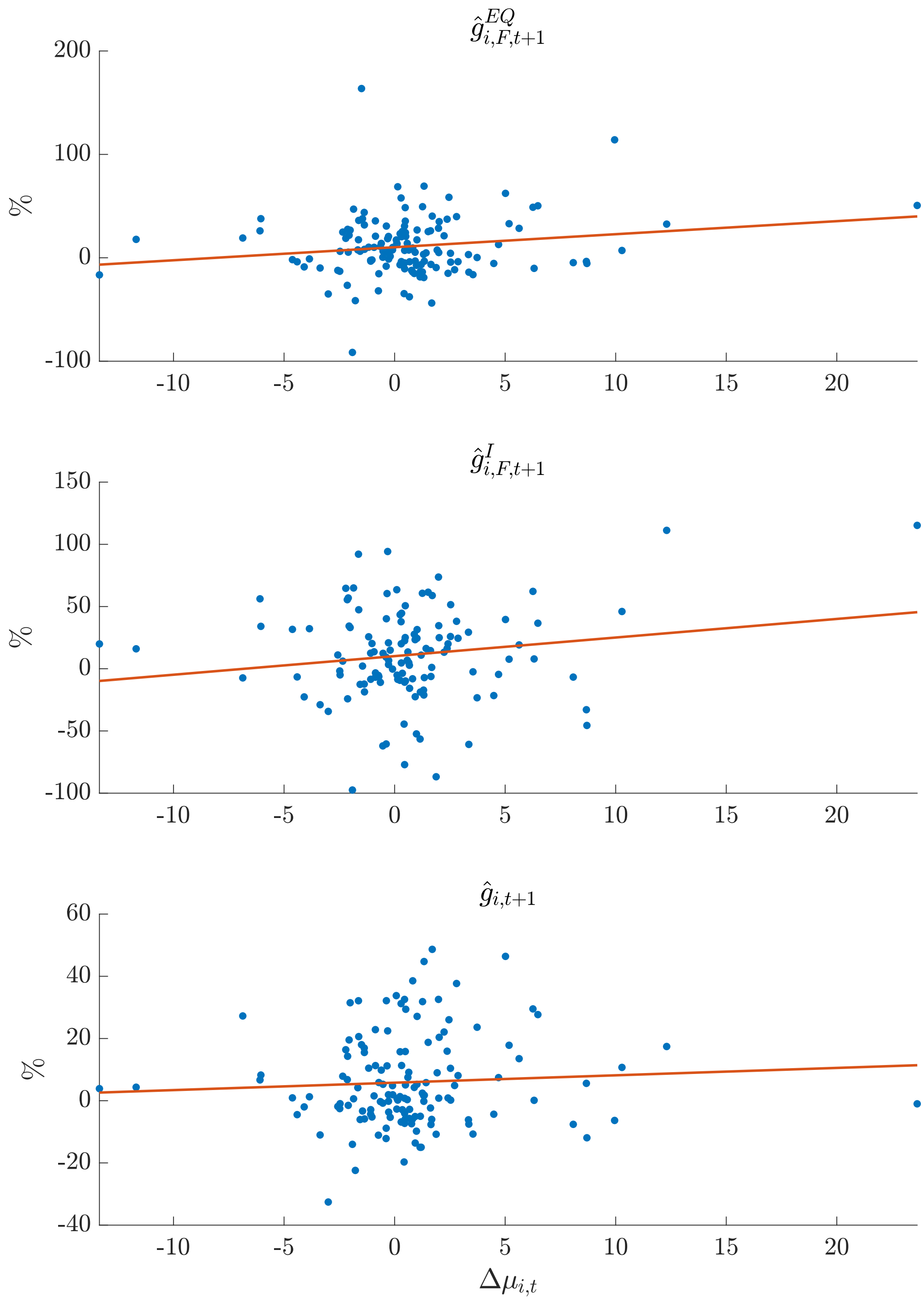


Figure 20: change of market share and followers' growth, financial crisis 2008

Table 1: sales and growth, across-industry

size group	0-90th	90-99th	99-99.5th	>99.5th
Assets ($ mil.)	$2.0	$48.8	$626.0	$6766.3
Sales ($ mil., quarterly)	$1.2	$18.8	$181.1	$1420.8
Sales growth (year-on-year)	0.19%	4.58%	4.34%	4.08%
Investment rate (year-on-year)	26.50%	24.91%	21.89%	20.36%

Note: documented by Crouzet and Mehrotra (2020) with QFR data, assets and sales are averages from 1977q1 to 2014q1 within category expressed in real 2009 dollar.

Table 2: share of sales, within-industry

	Number of firms	*Top 4*	*Top 8*	*Top 20*	*Top 50*
Mean	732	42.1%	54.9%	70.1%	82.0%
Median	328	34.5%	54.5%	72.5%	87.3%

Note: Manufacturing sector of Census 2012, share of the sales by size group. The industry is defined in NAICS-6 level.

Table 3: financial characteristics

size group	0-90th	90-99th	99-99.5th	>99.5th
Debt to asset ratio	0.35	0.29	0.30	0.28
Cash to asset ratio	0.15	0.10	0.07	0.06
Bank debt (fraction of total debt)	0.48	0.57	0.43	0.28
Zero leverage (% of tot. firm-quarter obs.)	20%	13%	8%	3%
Bank dependent (% of tot. firm-quarter obs.)	26%	29%	20%	10%

Note: documented by Crouzet and Mehrotra (2020) with QFR data, assets and sales are averages from 1977q1 to 2014q1 within category expressed in real 2009 dollar.

Table 4: parameterization

	Parameter	Model	Moment
α	capital share	1/3	-
β	discount factor	0.95	Buera and Moll (2015)
r	real interest rate	0.02	Buera and Moll (2015)
λ	collateral constraints	2	Buera and Moll (2015)
ϵ	within-sector elasticity	10	Atkeson and Burstein (2008)
σ	between-sector elasticity	1.01	Atkeson and Burstein (2008)
z_l	productivity of large firms	21.547	market share of large firms
γ	Pareto parameter	1.149	income distribution of firms
ω	population of entrepreneurs	0.2	population share of entrepreneurs

Table 5: growth rate and change in market share

	$\hat{g}_{i,t}$		
	(1) 1980-2000	(2) 2001-2018	(3) 1980-2018
$\Delta\mu_{i,L,t}$	-0.773^{***} (0.071)	-0.866^{***} (0.071)	-0.843^{***} (0.051)
Mean of Share ($\mu_{i,L}$)	-0.006 (0.024)	-0.007 (0.026)	-0.004 (0.019)
Total asset(K)	-0.011^{**} (0.004)	-0.004^{*} (0.002)	-0.009^{***} (0.002)
Mean of number of firms	-0.012 (0.011)	0.021^{***} (0.007)	0.007 (0.006)
Difference in number of firms	0.516^{***} (0.080)	0.387^{***} (0.067)	0.481^{***} (0.051)
Other controls	-	-	-

Table 6: growth rate and change in market share, financial crisis

	$\hat{g}^{EQ}_{i,F,t+1}$	$\hat{g}^{I}_{i,F,t+1}$	$\hat{g}_{i,t+1}$
$\Delta\mu_{i,L,t}$	1.348^{**} (0.630)	1.595^{*} (0.844)	0.007 (0.290)
Mean of Share ($\mu_{i,L}$)	-0.017 (0.190)	-0.117 (0.266)	-0.068 (0.086)
Difference in number of firms	0.371 (0.294)	0.091 (0.433)	0.408^{**} (0.137)
Lags ($\hat{x}_{i,t}$)	-0.014 (0.013)	-0.006 (0.006)	-0.349^{***} (0.054)
Other controls	-	-	-

Table 7: growth rate and concentration

	$\hat{g}_{i,t}$		
	(1) 1980-2000	(2) 2001-2018	(3) 1980-2018
$\mathbb{1}_{\Delta\mu_{i,L,t}}$	-12.566^{***} (2.366)	-13.054^{***} (2.071)	-13.039^{***} (1.635)
$\mu_{i,L,t}$	-0.137^{***} (0.027)	-0.157^{***} (0.026)	-0.151^{***} (0.019)
$\mathbb{1}_{\Delta\mu_{i,L,t}} \cdot \mu_{i,L,t}$	0.161^{***} (0.040)	0.194^{***} (0.035)	0.173^{***} (0.028)
Total Asset (K)	-0.015^{***} (0.005)	-0.003 (0.002)	-0.010^{***} (0.002)
Difference in number of firms	0.551^{***} (0.074)	0.413^{***} (0.067)	0.561^{***} (0.051)
Other controls	-	-	-

Appendices

A Proofs

Proof of Lemma 1 Given the prices, the constant return to scale implies that the the marginal cost of the entrepreneur j is a weakly increasing function of $y_{i,j}$. Therefore, the output, the labor demand and the profit are corner solutions and linear with respective to $a'_{i,j}$.

Proof of Lemma 2 Lemma 1 implies that the entrepreneur's rate of return on saving is a constant given by

$$r_{i,j} = (1+r) + \max\left\{0, (\frac{z_{i,j}}{\underline{z}_i} - 1)r\lambda\right\}. \tag{77}$$

Note that $p'_{i,l}$, p'_i, p' and productivity $z'_{i,j}$ are revealed before the saving decision. Therefore, the entrepreneur j has no precautionary saving motive. Together with the logarithmic utility function, the saving of entrepreneur is a constant fraction β of $m_{i,j}$.

Proof of Lemma 3 Since productivity shocks on $z_{i,j}$ is i.i.d., the distributions of wealth and productivity are independent. Therefore, $y_{i,f}$ and $m'_{i,f}$ are linear functions with respective to $m_{i,f}$, which implies that $m_{i,f}$ serves as the sufficient state variable of the model.

Proof of Lemma 4 The FOC of the Bellman equation of leaders (24) takes the form that

$$y_{i,l,t} + \frac{(\kappa_i - 1)\,\phi_{i,t} y_{i,l,t}}{p_{i,l,t}} \frac{\frac{\partial y_{i,l,t}}{y_{i,l,t}}}{\frac{\partial p_{i,l,t}}{p_{i,l,t}}} + \frac{1}{1+r}\frac{\partial v_l}{\partial m_{i,f,t+1}}\frac{\partial m_{i,f,t+1}}{\partial p_{i,f,t}}\frac{\partial p_{i,f,t}}{\partial p_{i,l,t}} \quad = \quad 0. \tag{78}$$

By the envelope theorem, the bellman equation of leaders implies

$$\frac{\mathrm{d}v_l}{\mathrm{d}m_{i,f,t}} \quad = \quad (\kappa_{i,t} - 1)\,\phi_t \frac{\partial y_{i,l,t}}{\partial m_{i,f,t}} + \frac{1}{1+r}\frac{\mathrm{d}v_l}{\mathrm{d}m_{i,f,t+1}}\frac{\partial m_{i,f,t+1}}{\partial m_{i,f,t}}. \tag{79}$$

Combining the FOC and the envelope theorem recursively, it gives that

$$\begin{aligned}
\kappa_t \quad = \quad & -(\kappa_t - 1)\frac{\frac{\partial y_{i,l,t}}{y_{i,l,t}}}{\frac{\partial p_{i,l,t}}{p_{i,l,t}}} && (80)\\
& -(\kappa_t - 1)\frac{1}{1+r}\frac{\pi_{i,l,t+1}}{\pi_{i,l,t}}\frac{\frac{\partial y_{i,l,t+1}}{y_{i,l,t+1}}}{\frac{\partial p_{i,l,t}}{p_{i,l,t}}} && (81)\\
& -(\kappa_t - 1)\left(\frac{1}{1+r}\right)^2\frac{\pi_{i,l,t+2}}{\pi_{i,l,t}}\frac{\frac{\partial y_{i,l,t+2}}{y_{i,l,t+2}}}{\frac{\partial p_{i,l,t}}{p_{i,l,t}}} && (82)\\
& \ldots && (83)
\end{aligned}$$

Alternatively,

$$\kappa_t = \frac{1}{1-\Upsilon_t^{-1}}, \tag{84}$$

where

$$\Upsilon_t = \frac{\frac{\partial y_{i,l,t}}{y_{i,l,t}}}{\frac{\partial p_{i,l,t}}{p_{i,l,t}}} + \sum_{t=k}^{\infty}\left(\frac{1}{1+r}\right)^k \frac{\pi_{i,l,t+k}}{\pi_{i,l,t}} \frac{\frac{\partial y_{i,l,t+k}}{y_{i,l,t+k}}}{\frac{\partial p_{i,l,t}}{p_{i,l,t}}}. \tag{85}$$

For static elasticity $\upsilon_{i,t}$, taken p and y as given, differentiation of equation (26), we have

$$\frac{\mathrm{d}y_{i,l}}{y_{i,l}} = -\epsilon\frac{\mathrm{d}p_{i,l}}{p_{i,l}} + (\epsilon-\sigma)\frac{\mathrm{d}p_i}{p_i}, \tag{86}$$

where $\frac{\mathrm{d}p_i}{p_i}$ is given by the differentiation of equation (27) that

$$\frac{\mathrm{d}p_i}{p_i} = \mu_{i,l}\frac{\mathrm{d}p_{i,l}}{p_{i,l}} + (1-\mu_{i,f})\frac{\mathrm{d}p_{i,f}}{p_{i,f}}. \tag{87}$$

Taken p and y as given, the differentiation of the equation (28) and (29) give that

$$\frac{\mathrm{d}p_{i,f}}{p_{i,f}} = \frac{\epsilon-\sigma}{\epsilon+\frac{\gamma}{\alpha}-1}\frac{\mathrm{d}p_i}{p_i}. \tag{88}$$

Substitute into $\frac{\mathrm{d}y_{i,l}}{y_{i,l}}$, we can derive that

$$\upsilon_{i,t} = -\epsilon + (\epsilon-\sigma)\left[1+\left(\mu_{i,l,t}^{-1}-1\right)\frac{\sigma+\frac{\gamma}{\alpha}-1}{\epsilon+\frac{\gamma}{\alpha}-1}\right]^{-1}. \tag{89}$$

Proof of Lemma 5 By the definition of leaders' markup, $\frac{\mathrm{d}p_{l,t}}{p_l}$ is given by

$$\frac{\mathrm{d}p_{l,t}}{p_l} = \frac{\mathrm{d}\kappa_t}{\kappa} + (1-\alpha)\frac{\mathrm{d}w_t}{w}. \tag{90}$$

In the symmetric equilibrium, $\frac{\mathrm{d}p}{p}=0$. Therefore

$$\frac{\mathrm{d}p_{f,t}}{p_f} = -\frac{\mathrm{d}\kappa_t}{\kappa} - (1-\alpha)\frac{\mathrm{d}w_t}{w} + \frac{\mathrm{d}\rho_{l,t}}{\rho_l} + \frac{\mathrm{d}\rho_{f,t}}{\rho_f}. \tag{91}$$

The aggregate labor demands for followers and leaders are

$$l_{f,t} = \left(p_{f,t}\frac{1-\alpha}{w}\right)^{\frac{1}{\alpha}} \lambda\beta m_{f,t}\frac{\gamma}{\gamma-1}\underline{z}_t^{1-\gamma}, \tag{92}$$

$$l_{l,t} = y_{l,t}z_l^{-\alpha}\left(\frac{1-\alpha}{w_t}\frac{r}{\alpha}\right)^{\alpha} \tag{93}$$

respectively. Therefore, the log-linearization implies that

$$\frac{\mathrm{d}l_{f,t}}{l_f} = \frac{\gamma}{\alpha}\frac{\mathrm{d}p_{f,t}}{p_f} - \left[\frac{\gamma}{\alpha}(1-\alpha)+1\right]\frac{\mathrm{d}w_t}{w} + \frac{\mathrm{d}m_{f,t}}{m_f}, \tag{94}$$

$$\frac{\mathrm{d}l_{l,t}}{l_l} = \frac{\mathrm{d}y_{l,t}}{y_l} - \alpha\frac{\mathrm{d}w_t}{w}. \tag{95}$$

Note that

$$\frac{l_f}{l_l} = \frac{p_f y_f}{p_l y_l}\left[\frac{p_l}{z_l^{-\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\left(\frac{r}{\alpha}\right)^{\alpha}}\right] \tag{96}$$

$$= \kappa, \tag{97}$$

substitute into labor market clearing condition, we can find that

$$\left[\kappa\left(\frac{\gamma}{\alpha}-\gamma+1\right)+\alpha\right]\frac{\mathrm{d}w_t}{w} = \frac{\mathrm{d}y_{l,t}}{y_l} + \kappa\frac{\gamma}{\alpha}\frac{\mathrm{d}p_{f,t}}{p_f} + \kappa\frac{\mathrm{d}m_{f,t}}{m_f}. \tag{98}$$

Log-linearization of the supply and the demand for followers, the output can be written as prices and states by

$$\frac{\mathrm{d}y_t}{y} = -\left(\frac{\gamma}{\alpha}+\epsilon-1\right)\frac{\mathrm{d}\kappa_t}{\kappa} - \left[\left(\frac{\gamma}{\alpha}+\epsilon-1\right)(1-\alpha)+\left(\frac{\gamma}{\alpha}-\gamma\right)\right]\frac{\mathrm{d}w_t}{w} \tag{99}$$

$$+\frac{\mathrm{d}m_{f,t}}{m_f} + \left(\frac{\gamma}{\alpha}+\epsilon-1\right)\frac{\mathrm{d}\rho_{l,t}}{\rho_l} + \frac{\gamma}{\alpha}\frac{\mathrm{d}\rho_{f,t}}{\rho_f}. \tag{100}$$

By linearization of the demand for leaders, the output of leaders takes the form that

$$\frac{\mathrm{d}y_{l,t}}{y_l} = -\epsilon\frac{\mathrm{d}p_{l,t}}{p_l} + (\epsilon-1)\frac{\mathrm{d}\rho_{l,t}}{\rho_l} + \frac{\mathrm{d}y_t}{y}. \tag{101}$$

Substitute $\frac{\mathrm{d}y_{l,t}}{y_l}$ and $\frac{\mathrm{d}p_{f,t}}{p_f}$ into the labor market clearing condition, the wage can be written as functions of markups and aggregate states by

$$\frac{\mathrm{d}w_t}{w} = -\frac{(1+\kappa)\frac{\gamma}{\alpha}+2\epsilon-1}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+1+\kappa+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\kappa_t}{\kappa} \tag{102}$$

$$+\frac{1+\kappa}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+1+\kappa+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}m_{f,t}}{m_f} \tag{103}$$

$$+\frac{(1+\kappa)\frac{\gamma}{\alpha}+2\epsilon-2}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+1+\kappa+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\rho_{l,t}}{\rho_l} \tag{104}$$

$$+\frac{(1+\kappa)\frac{\gamma}{\alpha}}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+1+\kappa+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\rho_{f,t}}{\rho_f}. \tag{105}$$

Substitute $\frac{w_t}{w}$ into $\frac{y_t}{y}$, the deviation of the output as a function of markups and aggregate states is then given by

$$\begin{aligned}
\frac{\mathrm{d}y_t}{y} &= \frac{1+\kappa+(1-\kappa)(1-\alpha)(\epsilon-1)}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+1+\kappa+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}m_{f,t}}{m_f} && (106)\\
&\quad-\frac{\epsilon(\kappa-1)(1-\alpha)\frac{\gamma}{\alpha}+\gamma(1+\kappa)+(\epsilon-1)(\kappa+\alpha)}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+1+\kappa+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\kappa_t}{\kappa} && (107)\\
&\quad+\frac{\left(\frac{\gamma}{\alpha}+\epsilon-1\right)(\kappa+1)+(\kappa-1)(\epsilon-1)(1-\alpha)\frac{\gamma}{\alpha}}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+1+\kappa+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\rho_{l,t}}{\rho_l} && (108)\\
&\quad+\frac{\frac{\gamma}{\alpha}\left[1+\kappa+(1-\kappa)(1-\alpha)(\epsilon-1)\right]}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+1+\kappa+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\rho_{f,t}}{\rho_f}. && (109)
\end{aligned}$$

For the myopic leader, the compound elasticity Υ_t is

$$\Upsilon_t = \epsilon-(\epsilon-\sigma)\left[1+\left(\mu_{l,t}^{-1}-1\right)\frac{\sigma+\frac{\gamma}{\alpha}-1}{\epsilon+\frac{\gamma}{\alpha}-1}\right]^{-1}, \quad (110)$$

where

$$\begin{aligned}
\mu_{l,t} &= \frac{p_{l,t}y_{l,t}}{p_{l,t}y_{l,t}+p_{f,t}y_{f,t}} && (111)\\
&= \left(\frac{p_{l,t}}{\rho_{l,t}}\right)^{1-\epsilon}p_t^{\epsilon-1} && (112)
\end{aligned}$$

is a function of ρ_t, κ_t and $m_{f,t}$. By applying implicit function theorem to $\kappa_t=\left(1-\Upsilon_t^{-1}\right)^{-1}$ and log-linearization around steady state, we have

$$\frac{\mathrm{d}\kappa_t}{\kappa} = \frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}\frac{\mathrm{d}\log\Upsilon}{\mathrm{d}\log\mu_l}\frac{\mathrm{d}\mu_{l,t}}{\mu_l}, \quad (113)$$

where $\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}$ and $\frac{\mathrm{d}\log\Upsilon}{\mathrm{d}\log\mu_l}$ are steady state constants given by

$$\begin{aligned}
\frac{\mathrm{d}\log\kappa_t}{\mathrm{d}\log\Upsilon_t} &= -\frac{\Upsilon^{-1}}{1-\Upsilon^{-1}} && (114)\\
\frac{\mathrm{d}\log\Upsilon_t}{\mathrm{d}\log\mu_{l,t}} &= -\frac{(\epsilon-\sigma)\left[1+\frac{\sigma+\frac{\gamma}{\alpha}-1}{\epsilon+\frac{\gamma}{\alpha}-1}\right]^{-1}}{\epsilon-(\epsilon-\sigma)\left[1+\frac{\sigma+\frac{\gamma}{\alpha}-1}{\epsilon+\frac{\gamma}{\alpha}-1}\right]^{-1}}\frac{2\frac{\sigma+\frac{\gamma}{\alpha}-1}{\epsilon+\frac{\gamma}{\alpha}-1}}{1+\frac{\sigma+\frac{\gamma}{\alpha}-1}{\epsilon+\frac{\gamma}{\alpha}-1}}. && (115)
\end{aligned}$$

Since in the symmetric equilibrium $\mathrm{d}p_t=0$,

$$\begin{aligned}
\frac{\mathrm{d}\mu_{l,t}}{\mu_l} &= (1-\epsilon)\frac{\mathrm{d}p_{l,t}}{p_l}-(1-\epsilon)\frac{\mathrm{d}\rho_{l,t}}{\rho_l} && (116)\\
&= (1-\epsilon)\left[\frac{\mathrm{d}\kappa_t}{\kappa}+(1-\alpha)\frac{\mathrm{d}w_t}{w}\right]-(1-\epsilon)\frac{\mathrm{d}\rho_{l,t}}{\rho_l}. && (117)
\end{aligned}$$

Substitute $\frac{\mathrm{d}w_t}{w}$ into $\frac{\mathrm{d}\mu_{l,t}}{\mu_l}$, we can find that the leaders' market share as a function of markups and aggregate states by

$$\frac{\mathrm{d}\mu_{l,t}}{\mu_l} = \frac{(1-\epsilon)\left[(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+\alpha\right]}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\kappa_t}{\kappa} \tag{118}$$

$$+\frac{(\epsilon-1)\left[(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1\right]}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\rho_{l,t}}{\rho_l} \tag{119}$$

$$+\frac{(1-\epsilon)(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\rho_{f,t}}{\rho_f} \tag{120}$$

$$+\frac{(1-\epsilon)(1-\alpha)(1+\kappa)}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}m_{f,t}}{m_f}. \tag{121}$$

Then substitute $\frac{\mathrm{d}\mu_{l,t}}{\mu_l}$ into equation (113), we can pin down the κ as aggregate states by

$$\left[1+\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}\frac{\mathrm{d}\log\Upsilon}{\mathrm{d}\log\mu_l}\frac{(\epsilon-1)\left[(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+\alpha\right]}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}\right]\frac{\mathrm{d}\kappa_t}{\kappa} \tag{122}$$

$$= \frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}\frac{\mathrm{d}\log\Upsilon}{\mathrm{d}\log\mu_l}\frac{(\epsilon-1)\left[(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1\right]}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\rho_{l,t}}{\rho_l} \tag{123}$$

$$-\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}\frac{\mathrm{d}\log\Upsilon}{\mathrm{d}\log\mu_l}\frac{(\epsilon-1)(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\rho_{f,t}}{\rho_f} \tag{124}$$

$$-\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}\frac{\mathrm{d}\log\Upsilon}{\mathrm{d}\log\mu_l}\frac{(\epsilon-1)(1-\alpha)(1+\kappa)}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}m_{f,t}}{m_f}. \tag{125}$$

Therefore, the elasiticities in the definition of Δy_1^1 and Δy_1^2 related to the direct effect and the indirect effect are

$$\frac{\partial\log y}{\partial\log\rho_f} = \frac{\frac{\gamma}{\alpha}\left[1+\kappa+(1-\kappa)(1-\alpha)(\epsilon-1)\right]}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+1+\kappa+2(1-\alpha)(\epsilon-1)}, \tag{126}$$

$$\frac{\partial\log y}{\partial\log\rho_l} = \frac{\left(\frac{\gamma}{\alpha}+\epsilon-1\right)(\kappa+1)+(\kappa-1)(\epsilon-1)(1-\alpha)\frac{\gamma}{\alpha}}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+1+\kappa+2(1-\alpha)(\epsilon-1)}, \tag{127}$$

$$\frac{\partial\log y}{\partial\log\kappa} = -\frac{\epsilon(\kappa-1)(1-\alpha)\frac{\gamma}{\alpha}+\gamma(1+\kappa)+(\epsilon-1)(\kappa+\alpha)}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+1+\kappa+2(1-\alpha)(\epsilon-1)}, \tag{128}$$

$$\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\rho_f} = -\frac{\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}\frac{\mathrm{d}\log\Upsilon}{\mathrm{d}\log\mu_l}\frac{(\epsilon-1)(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}}{1+\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}\frac{\mathrm{d}\log\Upsilon}{\mathrm{d}\log\mu_l}\frac{(\epsilon-1)\left[(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+\alpha\right]}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}}, \tag{129}$$

$$\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\rho_l} = \frac{\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}\frac{\mathrm{d}\log\Upsilon}{\mathrm{d}\log\mu_l}\frac{(\epsilon-1)\left[(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1\right]}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}}{1+\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}\frac{\mathrm{d}\log\Upsilon}{\mathrm{d}\log\mu_l}\frac{(\epsilon-1)\left[(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+\alpha\right]}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}}. \tag{130}$$

Note that for the competitive benchmark, $\frac{\partial\log\tilde{y}}{\partial\log\rho_f}$ and $\frac{\partial\log\tilde{y}}{\partial\log\rho_l}$ are homomorphic to that of strategic leaders, except that the steady state κ is substitute by the $\tilde{\kappa}$ of the competitive benchmark with $\kappa>\tilde{\kappa}$. In particular, it is easy to check that $\frac{\partial\log y}{\partial\log\rho_f}$ is a decreasing function with respective to κ, while $\frac{\partial \log y}{\partial \log \rho_l}$ is increasing with κ. Therefore, $\frac{\partial \log y}{\partial \log \rho_f} < \frac{\partial \log \tilde{y}}{\partial \log \rho_f}$ and $\frac{\partial \log y}{\partial \log \rho_l} > \frac{\partial \log \tilde{y}}{\partial \log \rho_l}$. Therefore when $\mathrm{d}\log \rho_{l,1} < 0$ and $\mathrm{d}\log \rho_{f,1} < 0$,

$$\begin{aligned}
\Delta^1_{y_1}(\rho_{f,1}) &> 0, && (131)\\
\Delta^1_{y_1}(\rho_{l,1}) &< 0. && (132)
\end{aligned}$$

Note that $\frac{\partial \log y}{\partial \log \kappa} < 0$, $\frac{\partial \log \kappa}{\partial \log \rho_f} < 0$ and $\frac{\partial \log \kappa}{\partial \log \rho_l} > 0$, it is straight forward to show that

$$\begin{aligned}
\Delta^2_{y_1}(\rho_{f,1}) &< 0, && (133)\\
\Delta^2_{y_1}(\rho_{l,1}) &> 0. && (134)
\end{aligned}$$

Proof of Lemma 6 The aggregate saving function of followers is given by

$$m_{f,t+1} = \beta(1+r)\, m_{f,t} + \frac{\alpha}{\gamma} p_{f,t} y_{f,t}, \tag{135}$$

which implies that, around steady state,

$$\begin{aligned}
\mathrm{d}\log m_{f,t+1} &= \beta(1+r)\,\mathrm{d}\log m_{f,t} && (136)\\
&\quad + [1-\beta(1+r)]\,\mathrm{d}\log p_{f,t} && (137)\\
&\quad + [1-\beta(1+r)]\,\mathrm{d}\log y_{f,t}. && (138)
\end{aligned}$$

Substitute out $\frac{\mathrm{d}y_{f,t}}{y_f}$, it gives that

$$\begin{aligned}
\frac{\mathrm{d}m_{f,t+1}}{m_f} &= \frac{\mathrm{d}m_{f,t}}{m_f} + [1-\beta(1+r)]\frac{\gamma}{\alpha}\frac{\mathrm{d}p_{f,t}}{p_f} && (139)\\
&\quad - [1-\beta(1+r)](1-\alpha)\frac{\gamma}{\alpha}\frac{\mathrm{d}w_t}{w}. && (140)
\end{aligned}$$

Combining with $\frac{\mathrm{d}p_{f,t}}{p_f}$, $\frac{\mathrm{d}p_{l,t}}{p_l}$, $\frac{\mathrm{d}\kappa_t}{\kappa}$ and $\frac{\mathrm{d}\kappa_t}{\kappa}$, one can find that

$$\begin{aligned}
\frac{\mathrm{d}m_{f,t+1}}{m_f} &= \frac{2(1-\alpha)(\epsilon-1)+\kappa+1+2(1-\alpha)\beta(1+r)(1+\kappa)\frac{\gamma}{\alpha}}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}m_{f,t}}{m_f} && (141)\\
&\quad + [1-\beta(1+r)]\frac{\gamma}{\alpha}\frac{2(1-\alpha)\epsilon-(\kappa+1)}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\kappa_t}{\kappa} && (142)\\
&\quad + [1-\beta(1+r)]\frac{\gamma}{\alpha}\frac{\kappa+1-2(1-\alpha)(\epsilon-1)}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\rho_{l,t}}{\rho_l} && (143)\\
&\quad + [1-\beta(1+r)]\frac{\gamma}{\alpha}\frac{\kappa+1+2(1-\alpha)(\epsilon-1)}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}\frac{\mathrm{d}\rho_{f,t}}{\rho_f}. && (144)
\end{aligned}$$

Therefore, together with the equation (125), the elasticities in the definition of $\Delta m^1_{f,t}$ and $\Delta m^2_{f,t}$ are given by

$$\frac{\partial \log m'_f}{\partial \log \rho_l} = [1-\beta(1+r)]\frac{\gamma}{\alpha}\frac{\kappa+1-2(1-\alpha)(\epsilon-1)}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}, \tag{145}$$

$$\frac{\partial \log m'_f}{\partial \log \rho_f} = [1-\beta(1+r)]\frac{\gamma}{\alpha}\frac{\kappa+1+2(1-\alpha)(\epsilon-1)}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}, \tag{146}$$

$$\frac{\partial \log m'_f}{\partial \log m_f} = \frac{\beta(1+r)\,2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}, \tag{147}$$

$$\frac{\partial \log m'_f}{\partial \log \kappa} = [1-\beta(1+r)]\frac{\gamma}{\alpha}\frac{2(1-\alpha)\epsilon-(\kappa+1)}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}, \tag{148}$$

$$\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log m_f} = -\frac{\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}\frac{\mathrm{d}\log\Upsilon}{\mathrm{d}\log\mu_l}\frac{(\epsilon-1)(1-\alpha)(1+\kappa)}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}}{1+\frac{\mathrm{d}\log\kappa}{\mathrm{d}\log\Upsilon}\frac{\mathrm{d}\log\Upsilon}{\mathrm{d}\log\mu_l}\frac{(\epsilon-1)\left[(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+\alpha\right]}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}}. \tag{149}$$

Consider the case that the condition (50) satisfies. For the direct effect, i.e. $\sum_{t=2}^{\infty}\Delta^1_{m_t}(\rho_{j,1})$ for $j\in\{l,f\}$. Proof by mathematical induction.

1. When $t=2$, it is easy to check that $\frac{\partial \log m'_f}{\partial \log \rho_f}>0$ and a decreasing function of κ. Meanwhile, $\frac{\partial \log m'_f}{\partial \log \rho_l}<0$ and is an increasing function of κ. Note that $\kappa>\tilde{\kappa}$, it implies that

$$0<\frac{\partial \log m'_f}{\partial \log \rho_f}<\frac{\partial \log \tilde{m}'_f}{\partial \log \rho_f}, \tag{150}$$

$$0>\frac{\partial \log m'_f}{\partial \log \rho_l}>\frac{\partial \log \tilde{m}'_f}{\partial \log \rho_l}. \tag{151}$$

Therefore with $\mathrm{d}\log\rho_{j,1}<0$,

$$0>\frac{\partial \log m'_f}{\partial \log \rho_f}\mathrm{d}\log\rho_{f,1}>\frac{\partial \log \tilde{m}'_f}{\partial \log \rho_f}\mathrm{d}\log\rho_{f,1}, \tag{152}$$

$$0<\frac{\partial \log m'_f}{\partial \log \rho_l}\mathrm{d}\log\rho_{l,1}<\frac{\partial \log \tilde{m}'_f}{\partial \log \rho_l}\mathrm{d}\log\rho_{l,1}, \tag{153}$$

or,

$$\Delta^1_{m_f,1}(\rho_{f,1}) > 0, \tag{154}$$

$$\Delta^1_{m_f,1}(\rho_{l,1}) < 0. \tag{155}$$

2. When $t \geq 2$, suppose that

$$0 > \left(\frac{\partial \log m'_f}{\partial \log m_f}\right)^{t-1} \frac{\partial \log m'_f}{\partial \log \rho_f} \mathrm{d}\log \rho_{f,1} > \left(\frac{\partial \log \tilde{m}'_f}{\partial \log \tilde{m}_f}\right)^{t-1} \frac{\partial \log \tilde{m}'_f}{\partial \log \rho_f} \mathrm{d}\log \rho_{f,1}, \tag{156}$$

$$0 < \left(\frac{\partial \log m'_f}{\partial \log m_f}\right)^{t-1} \frac{\partial \log m'_f}{\partial \log \rho_l} \mathrm{d}\log \rho_{l,1} < \left(\frac{\partial \log \tilde{m}'_f}{\partial \log \tilde{m}_f}\right)^{t-1} \frac{\partial \log \tilde{m}'_f}{\partial \log \rho_l} \mathrm{d}\log \rho_{l,1}. \tag{157}$$

Since $\frac{\partial \log m'_f}{\partial \log m_f}$ is a decreasing function with respective to κ, or,

$$\frac{\partial \log \tilde{m}'_f}{\partial \log \tilde{m}_f} > \frac{\partial \log m'_f}{\partial \log m_f} > 0, \tag{158}$$

it must have

$$0 > \left(\frac{\partial \log m'_f}{\partial \log m_f}\right)^{t} \frac{\partial \log m'_f}{\partial \log \rho_f} \mathrm{d}\log \rho_{f,1} > \left(\frac{\partial \log \tilde{m}'_f}{\partial \log \tilde{m}_f}\right)^{t} \frac{\partial \log \tilde{m}'_f}{\partial \log \rho_f} \mathrm{d}\log \rho_{f,1}, \tag{159}$$

$$0 < \left(\frac{\partial \log m'_f}{\partial \log m_f}\right)^{t} \frac{\partial \log m'_f}{\partial \log \rho_l} \mathrm{d}\log \rho_{l,1} < \left(\frac{\partial \log \tilde{m}'_f}{\partial \log \tilde{m}_f}\right)^{t} \frac{\partial \log \tilde{m}'_f}{\partial \log \rho_l} \mathrm{d}\log \rho_{l,1}, \tag{160}$$

or

$$\Delta^1_{m_f,t}(\rho_{f,1}) \quad > \quad 0, \tag{161}$$

$$\Delta^1_{m_f,t}(\rho_{l,1}) \quad < \quad 0. \tag{162}$$

Therefore, for all $t \geq 1$, it turns out that $\Delta^1_{m_t}(\rho_{f,1}) > 0$ and $\Delta^1_{m_t}(\rho_{l,1}) < 0$, which implies that

$$\sum_{t=1}^{\infty} \Delta^1_{m_f,t}(\rho_{f,1}) \quad > \quad 0, \tag{163}$$

$$\sum_{t=1}^{\infty} \Delta^1_{m_f,t}(\rho_{f,1}) \quad < \quad 0. \tag{164}$$

For the indirect effect, similarly we although show the proof with mathematical induction.

1. When $t = 1$, since $\frac{\partial \log m'_f}{\partial \log \kappa} > 0$, $\frac{\mathrm{d}\log \kappa}{\mathrm{d}\log \rho_f} < 0$ and $\frac{\mathrm{d}\log \kappa}{\mathrm{d}\log \rho_l} > 0$,

$$0 > \left(\frac{\partial \log m'_f}{\partial \log \rho_f} + \frac{\partial \log m'_f}{\partial \log \kappa} \frac{\mathrm{d}\log \kappa}{\mathrm{d}\log \rho_f}\right) \mathrm{d}\log \rho_{f,1} > \frac{\partial \log \tilde{m}'_f}{\partial \log \rho_f} \mathrm{d}\log \rho_{f,1}, \tag{165}$$

$$0 < \left(\frac{\partial \log m'_f}{\partial \log \rho_l} + \frac{\partial \log m'_f}{\partial \log \kappa} \frac{\mathrm{d}\log \kappa}{\mathrm{d}\log \rho_l}\right) \mathrm{d}\log \rho_{l,1} < \frac{\partial \log \tilde{m}'_f}{\partial \log \rho_l} \mathrm{d}\log \rho_{l,1}, \tag{166}$$

or,

$$\begin{aligned}\Delta^2_{m_{f_1}}(\rho_{f,1}) &> 0, &(167)\\ \Delta^2_{m_{f_1}}(\rho_{l,1}) &< 0. &(168)\end{aligned}$$

2. When $t \geq 2$, suppose that $\Delta^2_{m_{f,t}}(\rho_{f,1}) > 0$ and $\Delta^2_{m_{f,t}}(\rho_{l,1}) < 0$. Note that

$$0 < \frac{\partial \log m'_f}{\partial \log m_f} + \frac{\partial \log m'_f}{\partial \log \kappa}\frac{\mathrm{d}\log \kappa}{\mathrm{d}\log m_f} < \frac{\partial \log \tilde{m}'_f}{\partial \log \tilde{m}_f}, \tag{169}$$

it must be that

$$\begin{aligned}\Delta^2_{m_{f,t+1}}(\rho_{f,1}) &> 0, &(170)\\ \Delta^2_{m_{l,t+1}}(\rho_{l,1}) &< 0. &(171)\end{aligned}$$

Therefore, for all $t \geq 1$, it turns out that $\Delta^2_{m_t}(\rho_{f,1}) > 0$ and $\Delta^2_{m_t}(\rho_{l,1}) < 0$, which implies that

$$\begin{aligned}\sum_{t=1}^{\infty}\Delta^2_{m_f,t}(\rho_{f,1}) &> 0, &(172)\\ \sum_{t=1}^{\infty}\Delta^2_{m_f,t}(\rho_{f,1}) &< 0. &(173)\end{aligned}$$

Proof of Lemma 7 By the definition of Υ_t,

$$\Upsilon_t = \upsilon_t + \frac{1}{1+r}\frac{\pi_{l,t+1}}{\pi_{l,t}}\left(\hat{\upsilon}_{t+1} - \upsilon_{t+1}\right) + \frac{1}{1+r}\frac{\pi_{l,t+1}}{\pi_{l,t}}\Upsilon_{t+1}. \tag{174}$$

Given that $\kappa_t = \frac{1}{1-\Upsilon_t^{-1}}$, it implies that the Eular equation of the leader takes the form that

$$\left(1-\kappa_t^{-1}\right)^{-1} = \upsilon_t + \frac{1}{1+r}\frac{\pi_{l,t+1}}{\pi_{l,t}}\left(\hat{\upsilon}_{t+1} - \upsilon_{t+1}\right) + \frac{1}{1+r}\frac{\pi_{l,t+1}}{\pi_{l,t}}\left(1-\kappa_{t+1}^{-1}\right)^{-1}. \tag{175}$$

Therefore, at steady state,

$$\Upsilon = \upsilon + \frac{\hat{\upsilon}_1}{r}, \tag{176}$$

where

$$\begin{aligned}\upsilon &= \epsilon - (\epsilon - \sigma)\left[1 + \left(\mu_l^{-1} - 1\right)\frac{\sigma + \frac{\gamma}{\alpha} - 1}{\epsilon + \frac{\gamma}{\alpha} - 1}\right]^{-1}, &(177)\\ \hat{\upsilon}_1 &= -\frac{\frac{\partial y_l}{y_l}}{\frac{\partial m_f}{m_f}}\frac{\frac{\partial m'_f}{m_f}}{\frac{\partial p_f}{p_f}}\frac{\frac{\partial p_f}{p_f}}{\frac{\partial p_l}{p_l}}. &(178)\end{aligned}$$

In further, taken p, y as given, the differentiation of equation (28), (26) and (27) gives

$$\frac{\frac{\partial p_f}{p_f}}{\frac{\partial p_l}{p_l}} = \frac{(\epsilon-\sigma)\,\mu_l}{\epsilon+\frac{\gamma}{\alpha}-1-(\epsilon-\sigma)\,\mu_f}. \tag{179}$$

The differentiation of equation (30) implies that

$$\frac{\frac{\partial m_f'}{m_f}}{\frac{\partial p_f}{p_f}} = \frac{\gamma}{\alpha}\left[1-\beta\,(1+r)\right]. \tag{180}$$

Taken p_l, y, p and w as given, applying the implicit function theorem to the differentiation of equation (26), (28), (29) and (27), we can find that

$$\frac{\frac{\partial y_l}{y_l}}{\frac{\partial m_f}{m_f}} = -\frac{(\epsilon-\sigma)\,\mu_f}{\epsilon+\frac{\gamma}{\alpha}-1-(\epsilon-\sigma)\,\mu_f}. \tag{181}$$

Proof of Lemma 8 Log-linearization of equation (175) and $\left(1-\kappa^{-1}\right)^{-1}=\Upsilon$, we can derive the system of difference equations as follows.

$$\begin{bmatrix} \frac{\mathrm{d}\kappa_{t+1}}{\kappa} \\ \frac{\mathrm{d}m_{f,t+1}}{m_f} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{bmatrix} \frac{\mathrm{d}\kappa_{t}}{\kappa} \\ \frac{\mathrm{d}m_{f,t}}{m_f} \end{bmatrix}, \tag{182}$$

where

$$x_{21} = \frac{\frac{\mathrm{d}m_f'}{m_f}}{\frac{\mathrm{d}\kappa}{\kappa}} > 0, \tag{183}$$

$$x_{22} = \frac{\frac{\mathrm{d}m_f'}{m_f}}{\frac{\mathrm{d}m_f}{m_f}} \in (0,1)\,, \tag{184}$$

$$x_{11} = \frac{\frac{1}{1+r}\left(\hat{\upsilon}_1\frac{\frac{\mathrm{d}\hat{\upsilon}_1}{\hat{\upsilon}_1}}{\frac{\mathrm{d}\mu}{\mu}}-\upsilon\frac{\frac{\mathrm{d}\upsilon}{\upsilon}}{\frac{\mathrm{d}\mu}{\mu}}\right)\frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_f}{m_f}}\frac{\frac{\partial m_f'}{m_f}}{\frac{\partial\kappa}{\kappa}}+\Upsilon\frac{\kappa^{-1}}{1-\kappa^{-1}}-\frac{\hat{\upsilon}_1}{r}\frac{\frac{\partial\pi}{\pi}}{\frac{\partial\kappa}{\kappa}}+\upsilon\frac{\frac{\mathrm{d}\upsilon}{\upsilon}}{\frac{\mathrm{d}\mu}{\mu}}\frac{\frac{\partial\mu}{\mu}}{\frac{\partial\kappa}{\kappa}}+\frac{\hat{\upsilon}_1}{r}\frac{\frac{\partial\pi}{\pi}}{\frac{\partial m_f}{m_f}}\frac{\frac{\partial m_f'}{m_f}}{\frac{\partial\kappa}{\kappa}}}{-\left(\hat{\upsilon}_1\frac{\frac{\mathrm{d}\hat{\upsilon}_1}{\hat{\upsilon}_1}}{\frac{\mathrm{d}\mu}{\mu}}-\upsilon\frac{\frac{\mathrm{d}\upsilon}{\upsilon}}{\frac{\mathrm{d}\mu}{\mu}}\right)\frac{1}{1+r}\frac{\frac{\partial\mu}{\mu}}{\frac{\partial\kappa}{\kappa}}+\frac{1}{1+r}\Upsilon\frac{\kappa^{-1}}{1-\kappa^{-1}}-\frac{\hat{\upsilon}_1}{r}\frac{\frac{\partial\pi}{\pi}}{\frac{\partial\kappa}{\kappa}}} > 0, \tag{185}$$

$$x_{12} = \frac{\frac{1}{1+r}\left(\hat{\upsilon}_1\frac{\frac{\mathrm{d}\hat{\upsilon}_1}{\hat{\upsilon}_1}}{\frac{\mathrm{d}\mu}{\mu}}-\upsilon\frac{\frac{\mathrm{d}\upsilon}{\upsilon}}{\frac{\mathrm{d}\mu}{\mu}}\right)\frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_f}{m_f}}\frac{\frac{\partial m_f'}{m_f}}{\frac{\partial m_f}{m_f}}+\frac{\hat{\upsilon}_1}{r}\frac{\frac{\partial\pi}{\pi}}{\frac{\partial m_f}{m_f}}\left(\frac{\frac{\partial m_f'}{m_f}}{\frac{\partial m_f}{m_f}}-1\right)+\upsilon\frac{\frac{\mathrm{d}\upsilon}{\upsilon}}{\frac{\mathrm{d}\mu}{\mu}}\frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_f}{m_f}}}{-\frac{1}{1+r}\left(\hat{\upsilon}_1\frac{\frac{\mathrm{d}\hat{\upsilon}_1}{\hat{\upsilon}_1}}{\frac{\mathrm{d}\mu}{\mu}}-\upsilon\frac{\frac{\mathrm{d}\upsilon}{\upsilon}}{\frac{\mathrm{d}\mu}{\mu}}\right)\frac{\frac{\partial\mu}{\mu}}{\frac{\partial\kappa}{\kappa}}+\frac{1}{1+r}\Upsilon\frac{\kappa^{-1}}{1-\kappa^{-1}}-\frac{\hat{\upsilon}_1}{r}\frac{\frac{\partial\pi}{\pi}}{\frac{\partial\kappa}{\kappa}}} \tag{186}$$

Define

$$f(\lambda) \equiv \lambda^2-(x_{11}+x_{22})\,\lambda+x_{11}x_{22}-x_{12}x_{21} \tag{187}$$

such that the eigenvalue of matrix $\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$ are the roots of the function $f(\lambda) = 0$. Note that $\frac{x_{11}+x_{22}}{2} > 0$ implies that the speed of convergence depends on the root

$$\lambda^* = \frac{x_{11} + x_{22} - \sqrt{(x_{11} - x_{22})^2 + 4x_{12}x_{21}}}{2}. \quad (188)$$

In particular, if $x_{12}x_{21} > 0$, it implies $\lambda^* < x_{22}$.

In the case of competitive benchmark, the system can be written as the similar difference equations as

$$\begin{bmatrix} \frac{\mathrm{d}\tilde{\kappa}_{t+1}}{\tilde{\kappa}} \\ \frac{\mathrm{d}\tilde{m}_{f,t+1}}{\tilde{m}_f} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & x_{22} \end{bmatrix} \begin{bmatrix} \frac{\mathrm{d}\tilde{\kappa}_t}{\tilde{\kappa}} \\ \frac{\mathrm{d}\tilde{m}_{f,t}}{\tilde{m}_f} \end{bmatrix}, \quad (189)$$

where the speed of convergence depends on x_{22}.

Therefore, if $x_{21}x_{21} > 0$, or $x_{21} > 0$, the speed of convergence would be higher in the forward-looking case. Note that it requires that

$$\frac{\hat{\upsilon}_1}{1+r} \frac{\frac{\mathrm{d}\hat{\upsilon}_1}{\hat{\upsilon}_1}}{\frac{\mathrm{d}\mu}{\mu}} \frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_f}{m_f}} \frac{\frac{\partial m_f'}{m_f}}{\frac{\partial m_f}{m_f}} + \underbrace{\frac{\hat{\upsilon}_1}{r} \frac{\frac{\partial\pi}{\pi}}{\frac{\partial m_f}{m_f}} \left(\frac{\frac{\partial m_f'}{m_f}}{\frac{\partial m_f}{m_f}} - 1 \right)}_{\text{positive}} + \underbrace{\upsilon \frac{\frac{\mathrm{d}\upsilon}{\upsilon}}{\frac{\mathrm{d}\mu}{\mu}} \frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_f}{m_f}} \left(1 - \frac{1}{1+r} \frac{\frac{\partial m_f'}{m_f}}{\frac{\partial m_f}{m_f}} \right)}_{\text{positive}} > 0. \quad (190)$$

A sufficient condition is that $\frac{\frac{\mathrm{d}\hat{\upsilon}_1}{\hat{\upsilon}_1}}{\frac{\mathrm{d}\mu}{\mu}} < 0$, or,

$$\epsilon > \underline{\epsilon} \equiv \left(\mu_l^{-1} - 1\right)\left(\sigma + \frac{\gamma}{\alpha} - 1\right) - \left(\frac{\gamma}{\alpha} - 1\right). \quad (191)$$

Furthermore, denote ϱ_1 and ϱ_2 such that

$$\frac{\mathrm{d}m_{f,t+1}}{m_f} = \varrho_1 \frac{\mathrm{d}m_{f,t}}{m_f}, \quad (192)$$

$$\frac{\mathrm{d}\kappa_t}{\kappa} = \varrho_2 \frac{\mathrm{d}m_{f,t}}{m_f}. \quad (193)$$

Substitute into the system of equations (182), it turns out that the speed of convergence ϱ_1 is the root of the function

$$\varrho_1^2 - (x_{11} + x_{22})\,\varrho_1 + x_{11}x_{22} - x_{12}x_{21} = 0. \quad (194)$$

Let $\varrho_1 = \lambda^*$. When $x_{12}x_{21} > 0$, it must be that

$$\varrho_2 = \frac{\lambda - x_{22}}{x_{12}} < 0, \quad (195)$$

which implies, not only the faster convergence, but also $\frac{\mathrm{d}\log\kappa_t}{\mathrm{d}\log m_{f,t}} < 0$: the sign of indirect effect remains consistent as in Lemma 6.

At $t=1$, for $k\in\{L,F\}$, log-linearization over the Eular Equation gives the system

$$\begin{bmatrix} \frac{\mathrm{d}\kappa_2}{\kappa} \\ \frac{\mathrm{d}m_{f,2}}{m_f} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12,k} \\ u_{21} & u_{22,k} \end{bmatrix} \begin{bmatrix} \frac{\mathrm{d}\kappa_1}{\kappa} \\ \frac{\mathrm{d}\rho_1}{\rho} \end{bmatrix}, \tag{196}$$

where

$$u_{11} = x_{11}, \tag{197}$$

$$u_{12,k} = \frac{\frac{1}{1+r}\left(\hat{\upsilon}\frac{\frac{\mathrm{d}\hat{\upsilon}}{\hat{\upsilon}}}{\frac{\mathrm{d}\mu}{\mu}} - \upsilon\frac{\frac{\mathrm{d}\upsilon}{\upsilon}}{\frac{\mathrm{d}\mu}{\mu}}\right)\frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_f}{m_f}}\frac{\frac{\partial m_f'}{m_f}}{\frac{\partial\rho}{\rho}} + \frac{\hat{\upsilon}}{r}\frac{\frac{\partial\pi}{\pi}}{\frac{\partial m_f}{m_f}}\frac{\frac{\partial m_f'}{m_f}}{\frac{\partial\rho}{\rho}} - \frac{\hat{\upsilon}}{r}\frac{\frac{\partial\pi}{\pi}}{\frac{\partial\rho}{\rho}} + \upsilon\frac{\frac{\mathrm{d}\upsilon}{\upsilon}}{\frac{\mathrm{d}\mu}{\mu}}\frac{\frac{\partial\mu}{\mu}}{\frac{\partial\rho}{\rho}}}{-\left(\hat{\upsilon}\frac{\frac{\mathrm{d}\hat{\upsilon}}{\hat{\upsilon}}}{\frac{\mathrm{d}\mu}{\mu}} - \upsilon\frac{\frac{\mathrm{d}\upsilon}{\upsilon}}{\frac{\mathrm{d}\mu}{\mu}}\right)\frac{1}{1+r}\frac{\frac{\partial\mu}{\mu}}{\frac{\partial\kappa}{\kappa}} + \frac{1}{1+r}\Upsilon\frac{\kappa^{-1}}{1-\kappa^{-1}} - \frac{\hat{\upsilon}}{r}\frac{\frac{\partial\pi}{\pi}}{\frac{\partial\kappa}{\kappa}}}, \tag{198}$$

$$u_{21} = \frac{\frac{\partial m_f'}{m_f}}{\frac{\partial\kappa}{\kappa}} > 0, \tag{199}$$

$$u_{22,k} = \frac{\frac{\partial m_f'}{m_f}}{\frac{\partial\rho}{\rho}}. \tag{200}$$

Consider the shock $\mathrm{d}\log\rho_{f,1}<0$ and $\mathrm{d}\log\rho_{l,1}<0$ separately.

1. When $\mathrm{d}\log\rho_{f,1}<0$, it must be that $\mathrm{d}\log m_{f,2}<0$, $\mathrm{d}\log\kappa_2>0$. Note that

$$\mathrm{d}\log\kappa_2 - u_{12,F}\mathrm{d}\log\rho_{1,F} = u_{11}\mathrm{d}\log\kappa_1, \tag{201}$$

a sufficient condition such that $\mathrm{d}\log\kappa_1>0$ is that $u_{12,F}>0$, or

$$\underbrace{\frac{1}{1+r}\hat{\upsilon}\frac{\frac{\mathrm{d}\hat{\upsilon}}{\hat{\upsilon}}}{\frac{\mathrm{d}\mu}{\mu}}\frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_F}{m_F}}\frac{\frac{\partial m_F'}{m_F}}{\frac{\partial\rho_F}{\rho_F}}}_{\text{positive}} + \upsilon\frac{\frac{\mathrm{d}\upsilon}{\upsilon}}{\frac{\mathrm{d}\mu}{\mu}}\left(\frac{\frac{\partial\mu}{\mu}}{\frac{\partial\rho_F}{\rho_F}} - \frac{1}{1+r}\frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_F}{m_F}}\frac{\frac{\partial m_F'}{m_F}}{\frac{\partial\rho_F}{\rho_F}}\right) + \frac{\hat{\upsilon}}{r}\left(\frac{\frac{\partial\pi}{\pi}}{\frac{\partial m_F}{m_F}}\frac{\frac{\partial m_F'}{m_F}}{\frac{\partial\rho_F}{\rho_F}} - \frac{\frac{\partial\pi}{\pi}}{\frac{\partial\rho_F}{\rho_F}}\right) > 0. \tag{202}$$

Given that

$$\frac{\frac{\partial\mu}{\mu}}{\frac{\partial\rho_F}{\rho_f}} = \frac{(1-\epsilon)(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}, \tag{203}$$

$$\frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_F}{m_F}} = \frac{(1-\epsilon)(1-\alpha)(1+\kappa)}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}, \tag{204}$$

$$\frac{\frac{\partial m_F'}{m_F}}{\frac{\partial\rho_F}{\rho_F}} = \frac{[1-\beta(1+r)]\frac{\gamma}{\alpha}[\kappa+1+2(1-\alpha)(\epsilon-1)]}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}, \tag{205}$$

$$\frac{\frac{\partial\pi}{\pi}}{\frac{\partial\rho_F}{\rho_F}} = \frac{\gamma}{\alpha}\frac{\frac{\partial\pi}{\pi}}{\frac{\partial m_F}{m_F}}, \tag{206}$$

it must be that

$$\frac{\frac{\partial\mu}{\mu}}{\frac{\partial\rho_F}{\rho_F}} < \frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_F}{m_F}}\frac{\frac{\partial m_F'}{m_F}}{\frac{\partial\rho_F}{\rho_F}}, \tag{207}$$

$$\frac{\frac{\partial\pi}{\pi}}{\frac{\partial\rho_F}{\rho_F}} < \frac{\frac{\partial\pi}{\pi}}{\frac{\partial m_F}{m_F}}\frac{\frac{\partial m_F'}{m_F}}{\frac{\partial\rho_F}{\rho_F}}. \tag{208}$$

Therefore, the inequality (202) holds true.

2. When $\mathrm{d}\log\rho_{L,1} < 0$, it must be that $\mathrm{d}\log m_{F,2} > 0$, $\mathrm{d}\log\kappa_2 < 0$. Note that

$$\mathrm{d}\log\kappa_2 - u_{12}\mathrm{d}\log\rho_{1,F} = u_{11}\mathrm{d}\log\kappa_1, \tag{209}$$

a sufficient condition such that $\mathrm{d}\log\kappa_2 < 0$ is that $u_{12,L} < 0$, or

$$\underbrace{\frac{1}{1+r}\hat{\upsilon}\frac{\frac{\mathrm{d}\hat{\upsilon}}{\hat{\upsilon}}}{\frac{\mathrm{d}\mu}{\mu}}\frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_f}{m_f}}\frac{\frac{\partial m_f'}{m_f}}{\frac{\partial\rho_l}{\rho_l}}}_{\text{negative}} + \frac{\hat{\upsilon}}{r}\left(\frac{\frac{\partial\pi}{\pi}}{\frac{\partial m_f}{m_f}}\frac{\frac{\partial m_f'}{m_f}}{\frac{\partial\rho_l}{\rho_l}} - \frac{\frac{\partial\pi}{\pi}}{\frac{\partial\rho_l}{\rho_l}}\right) + \upsilon\frac{\frac{\mathrm{d}\upsilon}{\upsilon}}{\frac{\mathrm{d}\mu}{\mu}}\left(\frac{\frac{\partial\mu}{\mu}}{\frac{\partial\rho_l}{\rho_l}} - \frac{1}{1+r}\frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_f}{m_f}}\frac{\frac{\partial m_f'}{m_f}}{\frac{\partial\rho_l}{\rho_l}}\right) < 0,. \tag{210}$$

Given that

$$
\begin{aligned}
\frac{\frac{\partial\mu}{\mu}}{\frac{\partial\rho_L}{\rho_L}} &= \frac{(\epsilon-1)\left[(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1\right]}{2(1-\alpha)(1+\kappa)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}, && (211)\\
\frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_F}{m_F}} &= \frac{(1-\epsilon)(1-\alpha)(1+\kappa)}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}, && (212)\\
\frac{\frac{\partial m'_F}{m_F}}{\frac{\partial\rho_L}{\rho_L}} &= \frac{[1-\beta(1+r)]\frac{\gamma}{\alpha}[\kappa+1-2(1-\alpha)(\epsilon-1)]}{2(1+\kappa)(1-\alpha)\frac{\gamma}{\alpha}+\kappa+1+2(1-\alpha)(\epsilon-1)}, && (213)\\
\frac{\frac{\partial\pi}{\pi}}{\frac{\partial\rho_L}{\rho_L}} &> -\frac{\frac{\partial\pi}{\pi}}{\frac{\partial m_F}{m_F}}, && (214)
\end{aligned}
$$

it must be that

$$
\begin{aligned}
\frac{\frac{\partial\mu}{\mu}}{\frac{\partial\rho_L}{\rho_L}} &> \frac{\frac{\partial\mu}{\mu}}{\frac{\partial m_F}{m_F}}\frac{\frac{\partial m'_F}{m_F}}{\frac{\partial\rho_L}{\rho_L}}, && (215)\\
\frac{\frac{\partial\pi}{\pi}}{\frac{\partial\rho_L}{\rho_L}} &> \frac{\frac{\partial\pi}{\pi}}{\frac{\partial m_F}{m_F}}\frac{\frac{\partial m'_F}{m_F}}{\frac{\partial\rho_F}{\rho_F}}. && (216)
\end{aligned}
$$

Therefore, the inequality (210) holds true.

Combining the two cases, it implies that for the forward-looking leader, $\frac{\mathrm{d}\log\kappa_1}{\mathrm{d}\log\rho_{1,F}}>0$, $\frac{\mathrm{d}\log\kappa_1}{\mathrm{d}\log\rho_{1,L}}<0$. The signs of the elasticity shown in Lemma 4 remain consistent with that of the myopic leader.