# Endogenous Firm Competition and the Cyclicality of Markups*

Hassan Afrouzi†
Columbia University

December 4, 2019

**Abstract**

I show that the cyclicality of *output growth* is a sufficient predictor for the cyclicality of markups in a large class of models that micro-found variable markups through dynamic trade-offs. First, I show that this class of models imply a unified law of motion for markups in which current markup depends on firms' expectations over future *changes* in demand. Second, I use survey data on firms' expectations to test this law of motion and find evidence for this channel. Calibrating a micro-founded model to the U.S. data, I find that markups are procyclical when the model incorporates the observed hump-shaped response of output to shocks, as commonly found in the data. Furthermore, the model with hump-shaped response of output matches the cross-correlation of output and markups as untargeted moments.

*Keywords*: markup cyclicality, implicit collusion, customer-base, output inertia.
*JEL Classification*: E3

*I am grateful to Olivier Coibion, Saroj Bhattarai, Andrew Glover and Matthias Kehrig for very helpful comments as well as seminar participants at UT Austin, Midwest Macro Meeting, and Federal Reserve Banks of Dallas and Kansas City.

†Columbia University, Department of Economics, 420 West 118 St. New York, NY 10027. Email: hassan.afrouzi@columbia.edu.

# 1 Introduction

What determines the cyclicality of markups? The cyclicality of markups have been emphasized as a key transmission mechanism in the business cycle literature. Countercyclical markups, for instance, are broadly viewed as a propagation mechanism within New Keynesian models (Christiano, Eichenbaum and Rebelo, 2011), as well as a potential reason for positive comovement between hours and wages (Rotemberg and Woodford, 1992). Nonetheless, there is no empirical consensus on the direction of this cyclicality due to challenges in identifying and measuring the marginal costs of firms. Furthermore, theories that micro-found variable markups also have conflicting predictions regarding markups.[1]

The first main contribution of this paper is to show that in a broad class of models that micro-found variable markups through dynamic trade-offs, namely implicit collusion and customer-base models, cyclicality of *output growth* is a sufficient predictor for the cyclicality of markups.

For instance, within the implicit collusion model, oligopolistic firms always choose the highest markup from which no one has an incentive to deviate. This gives rise to an equilibrium in which firms do not explicitly coordinate or collude, but it is implicitly agreed upon to charge the highest incentive compatible price. Variable markups arise because these incentives are changing over time, in particular due to stochastic discount rates and expected changes in demand of firms. In times when an industry's demand is expected to decline, each firm wants to seize the day and charge a low price while demand is at its highest. Consequently, the oligopoly is forced to act more competitively and charge a lower markup than usual in times when demand is *expected to decline*. Therefore, what determines the response of markups is the expected *changes* in relative demand of firms, which in the general equilibrium setting is equal to the output growth.

In fact, the reason that implicit collusion models are believed to imply countercyclical markups is that in classic business cycle models output is always expected to decline towards its steady state level during a boom. The negative expected output growth increases firms' incentives to charge lower prices to gain higher market share before demand falls and forces the oligopoly to settle on lower than usual markups. However, if the model is rich enough to allow for a hump-shaped response of output, which is a well-documented empirical property for most identified shocks in the literature[2], then

[1] For instance, two common micro-foundations – implicit collusion and customer-base models – yield different cyclicalities: implicit collusion models are interpreted as implying countercyclical markups, while customer-base models have been used to generate both procyclical and countercyclical markups.

[2] See e.g. Ramey (2011); Ramey and Shapiro (1998); Monacelli and Perotti (2008) for government spend-

markups are *procyclical* during the periods that output is rising, even if it happens during a boom.

Furthermore, the inertial response of output not only changes the contemporaneous correlation of output and markups, but also is crucial in capturing the cross-correlation of the two as documented by Nekarda and Ramey (2013) who show that lags of markups are procyclical while leads of them are countercyclical. Calibrating an implicit collusion model to the U.S. data, I find that while the traditional implicit collusion model fails to capture this feature of the data, the same model almost completely captures these cross-correlations as *untargeted moments* once we allow for inertial response of output to shocks through investment adjustment costs (Figure 2).

Another class of models that micro-found variable markups through dynamic incentives are customer-base models in which firms' market shares are assumed to be inertial.[3] The dynamic incentives of firms in these models is to build market share through low markups when the opportunity cost of doing so is low and charge high markups when this cost is high. Again, the important factor that determines this opportunity cost is the expected demand growth of the firm.

In this paper, I do not take a stance on the micro-foundations of this friction.[4] Instead, using a simple customer-base model with an exogenous habit formation process on the side of customers, that is rich enough to produce both procyclical and countercyclical markups. Then, I show that within each version inertia in response of output reverses the cyclicality of markups. My reduced form approach in modeling customer-base model comes at the cost of less plausible quantitative implications, but it allows me to make a weaker claim for a broader class of models: as long as customer-base frictions introduce a dynamic trade-off for building market share, markup cyclicality is determined by the cyclicality of the output growth.

I formally show that, up to a first order approximation, both of these models yield the same reduced form expression for the dynamics of markups. Specifically, current markups depend on the net present value of all expected growth rates of output in the future, discounted by a stochastic discount factor. The two models differ, however, in terms of the sign restrictions that they imply for this reduced form representation of markup dynamics. Hence, one can test, and differentiate between, these two models by estimating the common law of motion for markups and comparing the empirical estimates with

ing shocks, Sims (2011); Smets and Wouters (2007) for productivity shocks, and Christiano et al. (2005) for monetary policy shocks.

[3]See, e.g., Phelps and Sidney (1970); Paciello et al. (2014); Ravn et al. (2006).

[4]There has been a tremendous amount of progress in recent years in micro-founding this friction using search and matching frameworks. See, e.g., Gourio and Rudanko (2014); Kaplan and Menzio (2016).

the sign restrictions implied by each model.

The second main contribution of this paper is to use survey data on firms' expectations to implement this empirical test. Because current markups depend on firms' expectations of future economic conditions, measuring these expectations is crucial to the endeavor. To do so, I rely on a recent survey of firms' expectations from New Zealand, introduced by Coibion et al. (2018). Consistent with the insights from the calibrated model, the results favor the implicit collusion model, and therefore given the high levels of inertia observed in output, point toward acyclical or procyclical markups.

**Literature Review.** Both implicit collusion and customer-base models are used within macroeconomic models to study the markup setting behavior of firms. In my analysis, I start by building the firm side of the implicit collusion model, and show that markups are determined by the joint distribution of expected growth of output and stochastic discount rates. This allows me to reconcile seemingly contradictory predictions of these models in a unified framework. For instance, Rotemberg and Saloner (1986) assume that demand shocks are i.i.d., implicitly implying that the expected demand growth is countercyclical, and conclude that markups are countercyclical. On the other hand, Kandori (1991); Haltiwanger and Harrington Jr (1991); Bagwell and Staiger (1997), each by assuming alternative processes for demand shocks, find that these models can produce procyclical markups.

Rotemberg and Woodford (1991, 1992) are the first to study the implicit collusion model within a DSGE model. Contrary to the partial equilibrium models, their general equilibrium setting endogenously pins down the joint distribution of output growth and stochastic discount rates which give rise to countercyclical markups; however, their result is not robust to the structure of the shocks, and is reversed by introducing a humped-shaped response for output.

Phelps and Sidney (1970) is the first paper that formalizes the idea for customer-base models, and show that firms will charge a lower markup than the static monopoly one. Various papers have used this idea to study the cyclical behavior of markups. Different versions of customer-base models have been shown to create either procyclical or countercyclical markups. For instance, among the most recent ones, by micro-founding the game between firms and customers, Paciello et al. (2014) find that markups are procyclical, but Ravn et al. (2006) argue that they are countercyclical.

Moreover, another class of models that generate variable markups use demand systems where the elasticity of demand is assumed to be a function of the level of demand. For instance, Kimball demand (Kimball, 1995) implies lower elasticity, and therefore higher markup, at higher levels of demand and generates procyclical markups. A recent application is Edmond et al. (2018) which uses this demand system to assess the cost of markups. In this paper, I do not consider these class of models and focus only on models where variable markups are generated through dynamic trade-offs.

Due to the difficulty of measuring marginal costs, there is a wide range of results on the cyclicality of markups in the empirical literature. For instance, Bils et al. (2012) find markups to be countercyclical, while Nekarda and Ramey (2013) argue the opposite. The theoretical observations in this paper support findings of Nekarda and Ramey (2013). A calibrated version of the implicit collusion model closely matches the cross-correlation of markups and output that they document.

Section 2 introduces the implicit collusion model in general equilibrium and within a calibrated version shows that markups are procyclical when we account for the inertia of output response to shocks. Section 3 derives the law of motion for the customer-base model and shows the equivalence of the law of motion for the two class of models. Section 4 presents the evidence for this law of motion using survey data on firms' expectations. Section 5 concludes.

# 2 Implicit Collusion Models

In this section, I embed the implicit collusion setup within a standard macro framework. While the general setup is similar to Rotemberg and Woodford (1991, 1992), I depart from their representation by deriving the law of motion for markups, and showing that markups depend on the expected *growth* of output. Then, I show that the cyclicality of markups is reversed once the model is calibrated to replicate the observed humped-shape response of output to shocks in the data, which I do by introducing investment adjustment costs

There is a final good of consumption in the economy which is produced using a large number of intermediate differentiated goods. There is a measure one of intermediate good sectors indexed by $i \in [0,1]$. In each sector, there are $N$ identical firms producing differentiated goods, indexed by $i, j$ where $j \in \{1, \ldots, N\}$.

## 2.1 The Final Good Producer

The final good producer takes the price of consumption good, $P_t$, as given and produces with

$$Y_t = \left[\int_0^1 Y_{i,t}^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}}$$

where

$$Y_{i,t} = \Phi(Y_{i1}, \ldots, Y_{iN}) \equiv \left[ N^{-\frac{1}{\eta}} \sum_{j=1}^{N} Y_{i,j,t}^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}$$

Therefore, $\sigma$ is the elasticity of substitution across sectors, and $\eta$ is the elasticity of substitution within sectors. The profit maximization problem of this firm gives

$$\frac{P_{i,j,t}}{P_t} = N^{-\frac{1}{\eta}} Y_{i,j,t}^{-\frac{1}{\eta}} Y_{i,t}^{\frac{1}{\eta}-\frac{1}{\sigma}} Y_t^{\frac{1}{\sigma}}$$

which defines the inverse demand function of the firm $ij$. One can invert the set of these functions to get the demand function of each firm, given by

$$Y_{i,j,t} = Y_t D(\frac{P_{i,j,t}}{P_t}; \frac{P_{i,-j,t}}{P_t}) \tag{1}$$

where

$$D(\frac{P_{i,j,t}}{P_t}; \frac{P_{i,-j,t}}{P_t}) \equiv \frac{1}{N} \left( \frac{P_{i,j,t}}{P_t} \right)^{-\eta} \left[ \frac{1}{N} \sum_{k=1}^{N} \left( \frac{P_{i,k,t}}{P_t} \right)^{1-\eta} \right]^{\frac{\eta-\sigma}{1-\eta}}.$$

A general assumption in two layer CES models is that $\eta > \sigma > 1$. The economic intuition for this assumption is that goods within sectors are closer substitutes than goods across sectors, and it also implies that the demand of a single firm is increasing in the price of its competitors.

## 2.2 Intermediate Goods

Within sector $i$, there are $N$ identical firms that use capital and labor to produce with a Cobb-Douglas production function, $Y_{i,j,t} = Z_t^a K_{i,j,t}^{\alpha} L_{i,j,t}^{1-\alpha}$, where $Z_t^a$ is an economy wide technology shock with an AR1 process:

$$\log(Z_t^a) = \rho_a \log(Z_{t-1}^a) + \sigma_a \varepsilon_{a,t}$$
$$\varepsilon_{a,t} \sim \mathcal{N}(0,1)$$

Also, intermediate firms know that their impact on total production is negligible and take $Y_t$ and $P_t$ as given. Moreover, I assume that there are competitive markets in place for renting labor and capital so that firms also take factor prices, $W_t$ and $R_t$, as given.

### 2.2.1 The Repeated Game of Sector $i$

Let $Y_{i,t} \equiv \Phi(Y_{i1,t}, \ldots, Y_{iN,t})$ denote the output of sector $i$ at time $t$. The $N$ firms in sector $i$ take the demand function of the final good producer, (1), as given, and play an infinitely repeated game. I assume that price (or alternatively, quantity) along with capital and labor demands are the only control variables of firms. The current price of future profits is a stochastic process, $\{Q_{0,t} : t \geq 0\}$, which firms take as given. In general equilibrium, these will be determined by the relative marginal utilities of households in different states.

As in every super-game, this repeated game has many potential equilibria. Although there is no rigorous way to rank the multiple equilibria of this game, the standard assumption in the literature is that given the structure of the game, firms choose the equilibrium that yields the highest possible profit stream. To this end, after fixing a punishment strategy for the firms, I will construct the best possible equilibrium in which firms always collude.[5]

Following the literature on implicit collusion models, I assume that in every sector there exists a perfect monitoring system that detects any cheating with probability one. Therefore, the best cheating strategy for a firm would be to best respond to collusion outputs of their rivals, as they know that even a slight deviation from collusion output will be noticed and the punishment strategy will be triggered.[6]

### 2.2.2 Characterization of the Repeated Game Equilibrium

The equilibrium will be constructed as follows; before time 0 firms get together and layout a contingent plan for all possible states in the future. For every single state at every point of time, they assign a markup level for every firm such that it yields the highest profit for the sector and is incentive compatible with collusion relative to the following punishment strategy: in case a firm cheats from the agreement, the game will enter a punishment stage at which firms will charge the static best response markup forever after; however, at every period there is a possibility that the industry will renegotiate this with probability $1-\gamma$ and will move back to the collusion stage. This probability $\gamma$ is in fact pinning down the

[5] Such an equilibrium is not necessarily the equilibrium with the highest net present value of profits; it may be the case that occasional deviations yield higher profits compared to staying in collusion forever. Therefore there might be an equilibrium with occasional collusion that dominates the best equilibrium in which firms always collude. I abstract from this case, following Rotemberg and Woodford (1991, 1992, 1999).

[6] In the absence of such a system, however, static best responding may not be the best cheating strategy for a firm. If small deviations were unnoticeable with some probability, characterizing the best strategy is nontrivial. For instance, in an environment with imperfect monitoring, Green and Porter (1984) characterize equilibria in which firms switch to punishment when their price falls below a trigger price, even if it is caused by a negative demand shock rather than a cheating competitor.

expected punishment length such that after a firm cheats, the industry expects to remain in punishment stage for an average of $1/(1-\gamma)$ periods.

Therefore, firms within every sector maximize the discounted value of the industry's life time profits such that no firm in no state has an incentive to cheat. Note that incentive compatibility is the only restricting concern in this setting. Without it, firms would choose the monopoly markup for the industry at every state. However, a firm's incentive to cheat is at its highest level when the rest of the firms are committed to producing the monopoly output of the industry. This incentive declines as the markups of the other firms decrease towards the one in the static equilibrium. Also, notice that charging the best response markup at every state is trivially a feasible markup sequence for the industry in terms of incentive compatibility, and accordingly an equilibrium.

After choosing the markup sequence, firms then start the game at the collusion stage, denoted by $C$. Notice that at any time and any state firms would prefer to commit to collusion, so the game will stay in stage $C$ forever. Furthermore, in the punishment stage, firms will play the static Nash equilibrium with probability $\gamma$ at every period, and they will prefer to go back to collusion when industry renegotiates since collusion is at least as good as static best responding. Therefore, the proposed strategy is a sub-game perfect Nash equilibrium.

An alternative interpretation for this strategy, which justifies the name "implicit collusion", is that there is an implicit agreement among firms according to which each firm chooses the highest markup from which no one in the industry has an incentive to deviate. However, if a firm observes that a competitor has deviated from this mutual understanding, the agreement breaks, and the oligopoly reverts back to static best responding.

With a CRS production function, firms' capital to labor ratios are independent of their output level. This can be interpreted as firms having a constant marginal cost of production in a given period that is pinned down by factor prices:

$$MC_t = \frac{1}{Z_t^a}\left(\frac{R_t}{\alpha}\right)^{\alpha}\left(\frac{W_t}{1-\alpha}\right)^{1-\alpha}$$

Given firms' price taking behavior for factor prices, we can treat them as only choosing their markup for every possible state, with input demands being determined automatically given the ratio induced by prices.

Since I only focus on symmetric equilibria, in characterizing the strategy of a firm, I assume that others are charging the same markup which is going to be the collusion markup in the equilibrium. Therefore, $ij$'s profit from the action profile $\mu_i^t \equiv (\mu_{i,j,t}; \mu_{i,t})$, where $\mu_{i,t}$ is the collusion markup chosen by the industry and $\mu_t$ is the average markup in the economy, is given by

$$\Pi_{i,j,t}(\mu_{i,j,t};\mu_{i,t}) = P_t Y_t(\frac{\mu_{i,j,t}}{\mu_{i,t}} - \frac{1}{\mu_{i,t}})\mu_{i,t}^{1-\sigma}\mu_t^{\sigma-1}D(\frac{\mu_{i,j,t}}{\mu_{i,t}};1)$$

The following Proposition formalizes the equilibrium.

**Proposition 1.** *Each firm in sector $i$, maximizes its net present value of future profits subject to no other firm having an incentive to undercut them:*

$$\max_{\{\mu_{i,t}\}_{t=0}^{\infty}} \frac{1}{N}\mathbb{E}_0 \sum_{t=0}^{\infty}(\beta\gamma)^t Q_{0,t} Y_t(1-\frac{1}{\mu_{i,t}})\mu_{i,t}^{1-\sigma}\mu_t^{\sigma-1}$$

$$s.t. \quad \max_{\rho_{i,t}}\left\{(\rho_{i,t}-\frac{1}{\mu_{i,t}})D(\rho_{i,t};1)\right\} - \frac{1}{N}(1-\frac{1}{\mu_{i,t}}) \leq \beta\gamma\mathbb{E}_t Q_{t,t+1}\frac{Y_{t+1}}{Y_t}\left(\frac{\mu_{t+1}/\mu_{i,t+1}}{\mu_t/\mu_{i,t}}\right)^{\sigma-1}\Gamma_{i,t+1} \quad (2)$$

$$\Gamma_{i,t} \equiv \frac{1}{N}\left[(1-\frac{1}{\mu_{i,t}}) - \mu_{COU}^{-\sigma}(\mu_{COU}-1)\mu_t^{\sigma-1}\right] + \beta\gamma\mathbb{E}_t Q_{t,t+1}\frac{Y_{t+1}}{Y_t}\left(\frac{\mu_{t+1}/\mu_{i,t+1}}{\mu_t/\mu_{i,t}}\right)^{\sigma-1}\Gamma_{i,t+1}$$

*where $\beta^\tau Q_{t,t+\tau}$ is the time $t$ price of a claim that pays a unit of consumption at $t+\tau$, and $\mu_{COU} \equiv \frac{(N-1)\eta+\sigma}{(N-1)\eta+\sigma-N}$ is the equilibrium markup of static best responding for firms at any state. $\eta > \sigma$ guarantees that $\mu_{COU} < \mu_{MON} \equiv \frac{\sigma}{\sigma-1}$. The solution to this problem $\{\mu_{i,t}\}_{t=0}^{\infty}$ exists, and it is a Sub-game Perfect Nash Equilibrium for the repeated game in sector $i$, in which firms always collude.*

Equation (2) is the incentive compatibility constraint which requires that all firms in a sector prefer collusion to cheating in every possible state. Accordingly, such a sequence of assigned collusion markups are incentive compatible by construction and therefore form an equilibrium.

Now, suppose that the model is calibrated such that the constraint binds in the steady state, then for small perturbations around that steady state, a first order approximation yields

$$\hat{\mu}_t = \psi_1\mathbb{E}_t\left[\Delta\hat{y}_{t+1} + \hat{q}_{t,t+1}\right] + \psi_2\mathbb{E}_t\left[\hat{\mu}_{t+1}\right] \quad (3)$$

where hats denote percentage deviations from the steady state level, and $\Delta\hat{y}_{t+1} \equiv \hat{y}_{t+1} - \hat{y}_t$.

$$\psi_1 \equiv \gamma\beta\frac{\bar{\mu}\bar{\Gamma}}{D(\bar{\rho};1)-1/N} \geq 0$$

$$\psi_2 \equiv \gamma\beta\frac{D(\bar{\rho};1)-(\sigma-1)(\mu_C-1)(\frac{\bar{\mu}}{\mu_C})^\sigma/N}{D(\bar{\rho};1)-1/N} \lesseqgtr 0$$

Equation 3 gives the law of motion for average markups in the partial equilibrium of the firm side. This is the key equation in this paper that will underlie all the results in later sections. Therefore, the following subsection is devoted to discussing this result.

### 2.2.3 Interpretation

Implicit collusion implies that markups are forward looking variables that depend on the expected change in demand in the next period, the changes in the price of future profits, and the expected change of markup in the next period. $\psi_1$, which is the coefficient on the first two, is a positive number that is increasing in steady state gains from collusion ($\gamma\beta\bar{\mu}\bar{\Gamma}$) and decreasing in the marginal revenue that a firm makes by cheating in the steady state ($D(\bar{\rho};1) - 1/N$). The intuition behind this equation is the key to understanding the main results of this paper. Two things between current period and the period ahead affect the current period's markup: first, the current price of next period's profit, which is the discount factor of the firms. The more patient the firms are in an industry, the higher their collusion markup will be today as they value future profits more. Second, the expected growth in demand from current period to next period. If firms expect that demand tomorrow will be higher than today, then they do not want to lose the chance of cheating tomorrow by cheating today. Basically, firms want to wait until demand is at its highest to take advantage of cheating, as in that case they will collect the highest cheating gains. This incentive to wait diminishes firms' cheating incentives in the current period, allowing the industry to sustain a higher collusion markup. Therefore, when firms expect output to grow, they will charge markups that are closer to the monopoly one.

$\psi_2$, however, can theoretically be positive or negative based on the calibration of the model. The reason is that there are two opposite forces that affect the firms' cheating incentives based on their expectation of the future markup. Before explaining these two forces, it is useful to recall that what ultimately determines the sign and the magnitude of the change in the markup is how hard it is to sustain the collusion markup, or in other words, how motivated firms are to cheat given a level of markup for the industry.

Suppose that firms expect that the markup next period will be higher than its steady state level. On one hand, they know that their industry is going to collude on a higher markup tomorrow, so they do not want to miss that chance by cheating today and pushing the industry to the punishment stage. On the other hand, since firms know that all other industries will also charge high markups, they expect to have a very high demand shift towards their industry from the final good producer, if industry as a whole charge the static best response markup. This second force gives an incentive to every single firm to push the industry to punishment stage by cheating today. Obviously the magnitude of this effect depends on how elastic the final good producer's demand for the industry is; as seen in the expression of $\psi_2$, when $\sigma$ is close to 1, this force is negligible because the firms do not expect to get a large demand shift if the industry moves to the static Nash equilibrium.

The previous results in the IO literature can be seen as special cases of equation 3. For example, Rotemberg and Saloner (1986) follows a case where $\hat{q}_{t,t+1} = 0$ due to a constant discount rate, and $\mathbb{E}_t\left[\Delta\hat{y}_{t+1}\right] = -\hat{y}_t$ as shocks are assumed to be i.i.d. over time. Therefore, in their model

$$\hat{\mu}_t = -\psi_1\hat{y}_t$$

which is a demonstration of their result that markups should be counter-cyclical. But as (3) implies, assuming other processes for these variables can give rise to different results. With two different random processes, $\Delta\hat{y}_{t+1}$ and $\hat{q}_{t,t+1}$, that are potentially correlated, the spectrum of possibilities for their underlying distribution is large enough to allow for *any* type of result in terms of the cyclicality of markups. Therefore, we need to pin down this joint distribution, which in the case of this paper will be done by introducing a household side for the model.

Finally, $\psi_1$ and $\psi_2$ are completely pinned down by the firm side parameters $\sigma, \eta, N, \gamma$ plus $\beta$ which is going to be the subjective discount factor of the households in the general equilibrium.

## 2.3 Households, the Government and Market Clearing

There is a representative household that solves the following standard problem with investment adjustment costs.

$$\begin{aligned} &\max_{\{(C_t,N_t,I_t,K_{t+1})\}_{t=0}^{\infty}} \mathbb{E}_0 \textstyle\sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{1-\theta}}{1-\theta} - \phi\frac{L_t^{1+1/\epsilon}}{1+1/\epsilon}\right] \\ s.t.\quad & P_tC_t + P_tI_t \leq W_tL_t + R_tK_t + \textstyle\int_0^1 \sum_{j=1}^{N} \Pi_{i,j,t}di - T_t \\ & K_{t+1} = (1-\delta)K_t + (1 - S(\tfrac{I_t}{I_{t-1}}))I_t \\ & S(\tfrac{I_t}{I_{t-1}}) \equiv \tfrac{a}{2}(1 - \tfrac{I_t}{I_{t-1}})^2 \end{aligned}$$

The investment adjustment cost is included to allow for inertia in the response of output to shocks. As I will discuss later the extent of this inertia will be crucial in determining the cyclicality of markups.

There is also a government that uses lump-sum taxes from households to conduct fiscal policy, $G_t$. I assume that $G_t$ follows an AR(2) stochastic process

$$
\begin{gathered}
G_t = \bar{G} Z_t^g \\
\log(Z_t^g) = \rho_1^g \log(Z_{t-1}^g) + \rho_2^g \log(Z_{t-2}^g) + \sigma_g \varepsilon_{g,t} \\
\varepsilon_{g,t} \sim \mathcal{N}(0,1)
\end{gathered}
$$

Again, the AR(2) assumption on government spending process is to allow for a humped-shape response of output to a fiscal policy shock.

Finally, the market clearing conditions are

$$
\begin{gathered}
C_t + I_t + G_t = Y_t \\
K_t = \int_0^1 \sum_{j=1}^N K_{i,j,t} di \\
L_t = \int_0^1 \sum_{j=1}^N L_{i,j,t} di
\end{gathered}
$$

## 2.4 Calibration and Simulation

In this section, by simulating a log-linearized version of the model around a steady state in which the incentive compatibility constraint binds, I will show (1) why these models are typically interpreted as implying counter-cyclical markups and (2) that markups are actually pro-cyclical once the model is calibrated to generate realistic amounts of inertia in economic activity.

### 2.4.1 Parameters

I have set $\beta = 0.993$ to match the a steady state annual real interest rate of 3 percent, $\alpha = 0.35$ to match a steady state share of capital income of 35 percent, $\delta = 0.025$ to match a 10 percent annual rate of depreciation on capital, $\phi = 8$ to match a steady state labor supply of 0.3, $\bar{G} = 0.2$ to match a steady state $G/Y$ of 20 percent, and $a = 2.48$ following Christiano et al. (2005). I also set the Frisch labor supply elasticity, $\epsilon$, to 2.5.

Moreover, I have set the elasticity of substitution across sectoral goods, $\sigma$, equal to 4, and the elasticity of substitution within sectoral goods,$\eta$, equal to 20. I set $\gamma = 0.8$ and $N = 15$ to match a steady state markup level of 20 percent.[7] Although these are calibrated in an arbitrary fashion, as I will show later in a series of robustness checks, for the given levels of $\sigma$ and $\eta$ the model is not very sensitive to these parameters, and reasonable

[7]Given $\sigma$ and $\eta$, this is the highest level for $\gamma$ for which the incentive compatibility constraint binds, and the Blanchard Kahn condition for the law of motion for markups holds.

variations in them do not affect the main results of my analysis. The qualitative results in terms of direction of cyclicality of markups are robust to any calibration as long as $\eta > \sigma$.

Finally, I have set the persistence of the technology shock to 0.95. For the persistence parameters of the government spending shock, I run the following regression on the quarterly data for real government consumption expenditures and gross investment from 1947Q1 to 2014Q1:

$$\log(G_t) = Constant + \rho_1^g \log(G_{t-1}) + \rho_2^g \log(G_{t-2}) + \varepsilon_t$$

which gives the estimates $\rho_1^g = 1.51$ and $\rho_2^g = -0.52$. I will also consider alternative persistence parameters for robustness checks in Section 2.4.4.

### 2.4.2 Impulse Response Functions

First, consider the case of no investment adjustment cost ($a = 0$). The dashed curves in Figure 1a show the impulse responses of this model to a 1 percent technology shock. The key observation is that in this setting, output jumps up on impact and converges back to zero as the effect of the transitory shock fades away. Moreover, the response of stochastic discount rate, which is given by $Q_{t,t+1} = \beta \frac{u'(C_{t+1})}{u'(C_t)}$, is countercyclical given that households are able to smooth their consumption without being restricted by costly investment. The fact that consumption has an inertial response to the technology shock is a crucial element to the countercyclicality of stochastic discount rates. On impact, households expect that their consumption will peak later in the expansion; therefore, they are not really concerned about future states as they know they will have a higher consumption.

By equation 3, the combination of counter-cyclical output growth and discount rates gives rise to counter-cyclical markups. The interpretation from the firm side is that on impact, firms know that demand is at its highest. This expectation along with the low price of future profits increases their cheating incentives to its highest, as they know that now is the best time to cheat and steal the market share of their rivals. The oligopoly, knowing this, is forced to settle on a low markup to eliminate this high cheating incentive.

A similar exercise can be done with the government spending shock. Suppose that $Z_t^g$ is an AR(1) process with persistence 0.95. The impulse response functions of the model to such a shock is illustrated by the dashed curves in Figure 1b. On impact, government spending is at its highest, which means that private consumption is at its lowest. First, since private consumption will increase to its steady state level, such a shock would give rise to counter-cyclical stochastic discount rates. Moreover, the income effect of $G$ is at its highest on impact, so that $Y$ will peak immediately due to a jump in labor supply and converge back to its steady state as the shock fades away. Again, the combination of counter-cyclical discount rates and output growth will translate into counter-cyclical markups.

However, empirical evidence on TFP shocks and government spending shocks suggests that the response of output to these shocks is inertial such that the peak effect happens not on impact but in later periods.[8] To allow for such inertia, I introduce investment adjustment costs and an AR(2) process for government spending. Solid curves in Figure 1a show the IRFs of the model to a 1% technology shock when $a = 2.48$. With positive adjustment costs, two things happen: first, investment does not jump on impact and has an inertial response, which translates to an inertial response in output, and second, households now face a stronger trade-off in smoothing their consumption because they face costly investment, which gives rise to pro-cyclical stochastic discount rates. Therefore, on impact firms expect their demand to increase in future periods, which gives them the incentive to avoid cheating until demand peaks as they do not want to force the oligopoly to the punishment stage of the game when demand is at its highest. This lower incentive to cheat allows the oligopoly to settle on higher incentive compatible markups. Hence, on impact one would expect a higher markup than the one in the steady state, making markups pro-cyclical.

A similar exercise can be done with the government spending shock by assuming an AR(2) process for $Z_t^g$. Figure 1b depicts the IRFs of the model to such a shock. The inertial implementation of the fiscal policy translates to an inertial output and consumption responses, as shown by solid curves in Figure 1b, which in turn produce procyclical markups for similar reasons to the case of the technology shock with $a > 0$.

### 2.4.3 Cross Correlation of Markups and Output

Another set of results that are emphasized in the literature is the cross correlation of markups and output over time. For instance, Nekarda and Ramey (2013) empirically document that in addition to the positive contemporaneous correlation of the two, lags of markups are pro-cyclical with output while the leads of it are counter-cyclical, as illustrated by the dashed curve with no markers in Figure 2. The goal of this section is to show that the model without inertia completely fails to match this evidence; however, the model with inertia is highly consistent with it, even though that they were not targeted in

[8] For empirical evidence on humped-shape response of output to productivity and government spending shocks, see, for example, Sims (2011); Smets and Wouters (2007); Ramey (2011); Ramey and Shapiro (1998); Christiano et al. (2005); Monacelli and Perotti (2008).

the calibration.

The dashed curve with circle markers in Figure 2 shows the simulated correlation of lags and leads of markups with output conditional on a TFP shock when there is no inertia in the response of output; and the solid curve in the same figure depicts the same graph for the model with inertia. While the model without inertia misses the direction of cyclicality for the most part, the model with inertia in the output response matches the empirical evidence closely.[9]

### 2.4.4 Robustness

In this section, I check the robustness of the predictions of the model with respect to different variables.

Figures 3a and 3c show the simulated correlations of leads and lags of markups with GDP conditional on a technology shock and a government spending shock respectively, for values of $\gamma$ between 0.4 and 0.8, such that darker curves correspond to higher levels of $\gamma$. Aside from the fact that lower $\gamma$'s create lower steady state markups because of the higher impatience of firms, they also produce lower correlations between GDP and markups. The reason for the latter is that variations in current markup are a weighted sum of all expected output changes and stochastic discount rate changes in the future, and as $\gamma$ gets smaller, they put lower weights on future values. Nevertheless, all values of $\gamma$ yield the same structure of correlations of lags and leads of the markup with the output.

Moreover, Figures 3b and 3d, respectively, show the correlation of leads and lags of markups with GDP conditional on a technology shock and government spending shock for values of $N$ between 5 and 25. Again darker curves represent higher values of $N$. Variation in number of competitors does not change the structure of correlations and has very small level effects. The reason is that what ultimately determines the cheating incentives of firms, and hence markups, is the elasticity of demand for a single firm which is equal to $\eta - \frac{\eta-\sigma}{(N-1)\rho^{\eta-1}+1} \in [\sigma, \eta]$. Note that for small amounts of $\eta$ ($\eta \leq 20$), which corresponds to a relatively high differentiation among within industry goods, the effect of $N$ on the structure and level of correlations is negligible.

In the model, investment adjustment costs are the mechanism generating the humped-shape response of output to technology shocks. While I have calibrated this parameter

[9] The same exercise can be done with government spending shocks, and while the results qualitatively remain the same – meaning that correlations in the inertial model are larger than in the model with no inertia in the output response – the inertia created by the AR(2) process is not enough to make the conditional correlation positive.

to the estimated value of Christiano et al. (2005), this section examines the question of how large this parameter needs to be for markups to be procyclical. Figure 4a depicts the number of periods that markups are procyclical after a 1% technology shock given different values of $a \in [0, 5]$. As soon as $a$ is larger than 0, markups are procyclical on impact. Also, the duration of procyclicality increases as $a$ gets larger. For my calibration of this parameter, markups are pro-cyclical for 5 quarters after the shock hits the economy.

Moreover, Figure 4b shows the contemporaneous correlation of the markup with GDP conditional on a 1% technology shock, which shows that the conditional correlation is increasing in $a$, and for $a > 1.2$ , it is positive. Hence, any empirically reasonable value of investment adjustment costs will generate procyclical markup in this model.

In the baseline calibration with inertia, I use estimated parameters for the AR(2) process of government spending to create the humped-shape response of the output to a $G$ shock. Now, I consider a wider range of persistence parameters to check for robustness of results in previous section. Consider the set $\{(\rho_1^g, \rho_2^g) | \rho_2^g \in [-0.7, 0], \rho_1^g + \rho_2^g = \rho_G\}$, where $\rho_G$ is the persistence of government spending shocks, fixed to an estimated value of 0.98. Therefore, this set defines a locus for persistence parameters of $G$ such that when $\rho_2^g = 0$ the process is AR(1) and when $\rho_2^g < 0$ the process is AR(2) with highest inertia achieved when $\rho_2^g = -0.7$. In fact, the magnitude of this parameter, $|\rho_2^g|$, determines the degree of inertia in the response of output. Figure 4c shows the number of periods that markups are procyclical after a 1% government spending shock given different values of $|\rho_2^g|$. Again, for the most part ($|\rho_2^g| > 0.1$), the inertia causes the markups to be procyclical on impact. For my estimate of persistence parameters, markups are procyclical for 2 periods after the impact. However, as Figure 4d shows, the inertia is not enough to make the conditional correlation of markup and GDP positive. Nevertheless, the correlation is still increasing with inertia.

# 3 Customer-base Models

Another class of models that micro-found variable markups is based on the notion that it is costly for customers to switch among firms. Accordingly, firms' pricing decisions in the current period affect their market share in future periods. In contrast to implicit collusion models, customer-base models have been shown to imply both procyclical and countercyclical markups.

The method through which customer-base frictions are modeled in the literature is by a habit formation process on the side of customers. This habit component makes it costly for customers to switch to other products, and hence creates dynamic considerations for firms as their future market share changes with their current decisions. The dynamics of markups, however, depend on how customers are reacting to pricing of the firms over time. Hence, the habit formation process is crucial for their dynamics.

In this section I build a simple reduced form customer-base model, with a rich enough external habit-formation process that could support both procyclical and countercyclical markups, and then show that introducing a humped-shape response for output changes the cyclicality of markups within each version. The reason is identical to the implicit collusion model; regardless of how the habit formation process works, the dynamic considerations that it creates for the firms are not in terms of the level of demand but its relative changes over time. Hence, similar to the implicit collusion model, since a humped-shape response of output changes the expectations of firms about changes in future demand over the business cycle, it also changes the cyclicality of markups.

## 3.1 Model Specification

Consider the final good producer of Section 2. To incorporate the customer-base model, I assume that this final good producer has an external habit formation over the goods within industries, meaning that

$$\begin{aligned} Y_t &= \left[\int_0^1 Y_{i,t}^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}} \\ Y_{i,t} &\equiv \Phi\left(\frac{Y_{i1,t}}{S_{i1,t}}, \ldots, \frac{Y_{iN,t}}{S_{iN,t}}\right) \\ \Phi(x_1, \ldots, x_N) &= \left[N^{-\frac{1}{\eta}} \sum_{j=1}^N x_j^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}} \end{aligned} \tag{4}$$

Where $S_{i,j,t}$ is the external habit of producer in using the input $Y_{i,j,t}$, which is taken as given by the final good producer at time $t$. I assume that $S_{i,j,t}$ has the following general law of evolution

$$S_{i,j,t} = h\left(\frac{\mu_{i,j,t}}{\mu_{i,t}}\right)(\gamma S_{i,j,t-1} + 1 - \gamma)$$

where $h(.)$ is differentiable, $h(1) = 1$, and $\gamma \in [0, 1)$. Therefore, the problem of the final good producer is

$$\max_{Y_{i,j,t}} P_t Y_t - \int_0^1 \sum_{j=1}^N P_{i,j,t} Y_{i,j,t} di$$

which implies that the demand for $Y_{i,j,t}$ is then given by

$$Y_{i,j,t} = Y_t D(P_{i,j,t} S_{i,j,t}; P_{i,-j,t} S_{i,-j,t})$$

where $D(.;.)$ is defined exactly as in section 3. Firm $ij$ takes demand as given and maximizes the net present value of all its future profits by choosing a relative markup $\frac{\mu_{i,j,t}}{\mu_{i,t}}$, and $S_{i,j,t}$, where $S_{i,t}$ is the the final good producer's habit for the others in the sector, in the symmetric equilibrium. Therefore, firm $ij$'s dynamic problem is

$$\begin{aligned} & \max_{\{\mu_{i,j,t}, S_{i,j,t}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} Q_{0,t} Y_t P_t^{-\sigma} \left(\frac{\mu_{i,t}}{\mu_t}\right)^{1-\sigma} \left(\frac{\mu_{i,j,t}-1}{\mu_{i,t}}\right) D\left(\frac{\mu_{i,j,t}}{\mu_{i,t}} \frac{S_{i,j,t}}{S_{i,t}}; 1\right) \\ s.t. \quad & S_{i,j,t} = h\left(\frac{\mu_{i,j,t}}{\mu_{i,t}}\right)(\gamma S_{i,j,t-1} + 1 - \gamma) \end{aligned}$$

**Proposition 2.** *In a symmetric equilibrium where all firms identically solve the problem above,*

1. *The law of motion for markups, up to a first order approximation, takes the same form as the implicit collusion model, i.e.*

$$\hat{\mu}_t = \psi_1 E_t\{\hat{q}_{t,t+1} + \Delta\hat{y}_{t+1}\} + \psi_2 E_t\{\hat{\mu}_{t+1}\}$$

   *where* $\psi_1 \equiv \beta\gamma\frac{\mu\mu_C^{-1}-1}{1+h'(1)} \lesseqgtr 0$, $\psi_2 \equiv \frac{\beta\gamma}{1+h'(1)} > 0$, $\mu_C \equiv \frac{(N-1)\eta+\sigma}{(N-1)\eta+\sigma-N} > 1$ *is the static best response markup in absence of the customer-base friction, and* $\mu$ *is the steady state markup in presence of the friction.*

2. *Markups move in the opposite direction of firms' expectations of output growth, meaning that* $\psi_1 < 0$, *if and only if the existence of customer-base frictions reduce the average markup, i.e.* $\mu < \mu_C$.

3. *Existence of customer-base frictions reduce the average markup if and only if* $h'(1) > 0$.

*Proof.* See Appendix. □

The first part of the proposition formalizes the idea that the law of motion implied by this model takes the same form as that of the implicit collusion model; however, the implications of the model on signs of $\psi$'s are different. While $\psi_2$ is unambiguously positive, the second part of the proposition relates the sign of $\psi_1$ to the nature of the customer-base friction. If the customer-base friction is such that it reduces the average markups compared to the frictionless economy, then markups should move in the opposite direction of expected output growth, and vice versa. In the context of models without inertia, the former would translate into procyclical markups. The latter, which predicts that markups should move in the same direction as the expected output growth if the average markup is higher than the frictionless one, is also consistent with the other side of the literature which predicts that markups are countercyclical. The third part gives a necessary and sufficient condition between the shape of the function $h(.)$ and the relationship between the average and frictionless markups.

Although they have opposing predictions, both versions of the model follow the same intuition. The reason that in this model markups are variable over time is that by altering their relative markup, and through the law of motion for habit, firms can change their future market shares. The fact that the law of motion depends on the net present value of all output growths in the future, implies that firms want to build high market share for periods in which they expect the output to be *higher*. However, how this incentive affects the markups, which is determined by the sign of $\psi_1$, depends on how potentially opposing forces aggregate together: there are two forces through which the relative markup affects firms' market shares. The first one is the direct relationship between price and demand; that a higher relative markup reduces current demand for firms. The second one affects market share through law of the motion for the habit: higher relative markup can either increase or decrease the habit of the final good producer for a particular firm.[10] While the first one always goes in one direction, the second force can work in either direction, causing a different law of motion for markups. For instance, if decreasing the relative markup increases the habit of the final good producer,[11] then markups should move in the opposite direction of the firms' expectations on output growth.

More importantly, even if we fix the sign of $\psi_1$, similar to the implicit collusion model, a humped-shaped response for the output would change the cyclicality of the markups. Figures 5a and 5b, shows the impulse responses of a version of the model with $\psi_1 < 0$,[12] to a 1% technology shock and government spending shock with and without inertial response for the output. Humped-shaped response of output changes the cyclicality of markups, even with a fixed $\psi_1$.

[10] By assuming a reduced form law of motion for final producer's habit, I abstract away from microfounding the slope of the $h(.)$ function which determines the sign of $\psi_1$. Nevertheless, the equivalence between slope of $h(.)$ and the relationship between average and frictionless markup, derived in Proposition 2, offers some intuition. The nature of the friction in terms of whether it decreases or increases the average markups is sufficient for identifying the sign of $\psi_1$. For instance, Paciello et al. (2014) mention that in their version of the customer-base model, the friction reduces the average markups and they find that markups are procyclical, which implies $\psi_1 < 0$.

[11] It is important to distinguish between how relative markup affects the habit of the final good producer, and how it affects the market share. The market share is determined by the two forces combined, and the assumption that a higher markup increases the habit, does not imply that the market share is also increasing in relative markup.

[12] The calibration is such that the frictionless markup, $\mu_C$, is 11.5%, and the average markup, $\mu$, is 10%. In terms of parameters, the only change compared to the previous section is $\eta = 10$.

# 4 Testing the Law of Motion for Markups

The two models considered in the previous sections both imply a law of motion for markups of the form

$$\hat{\mu}_t = \psi_1 \mathbb{E}_t\{\Delta \hat{y}_{t+1} + \hat{q}_{t,t+1}\} + \psi_2 \mathbb{E}_t\{\hat{\mu}_{t+1}\}$$

where the implication of each model for the signs of the coefficients are different. The following table summarizes these implications:

| | Implicit Collusion | Customer Base |
|---|---|---|
| $\psi_1$ | $> 0$ | $\lesseqgtr 0$ |
| $\psi_2$ | $\lesseqgtr 0$ | $> 0$ |

Table 1: Sign of Coefficients in the Law of Motion for Markups in Different Models

The goal of this section is to, first, empirically test this law of motion, and second, determine which model is more suited to explain the data based on the sign implications above. Because the question inherently relates to firms' expectations of future outcomes of the economy, I use a quantitative survey of firms' expectations from New Zealand introduced by **?**. Hence, if one could estimate the above parameters, rejecting that $\psi_1 > 0$ would reject the implicit collusion model; while rejecting the hypothesis that $\psi_2 > 0$ would imply the rejection of the customer-base model.

## 4.1 Identification

The survey includes data that allows me to directly test the law of motion for markups. Firms were asked to provide information about their number of competitors, average markup, current markup, their expected growth in sales, and their next expected price change. The following Proposition states that this cross-sectional data are enough to partially identify the coefficients in the law of motion up to the elasticity of substitution across industry goods.

**Proposition 3.** *Let industries be indexed by i and firms within them be indexed by j. Consider the following regression*

$$\begin{aligned}\hat{\mu}_{ij} - \sum_i \sum_j \hat{\mu}_{ij} &= Industry\_FE_i + \beta_1\{Ex\Delta Sales_{ij} - \sum_i \sum_j Ex\Delta Sales_{ij}\} \\ &+ \beta_2\{Ex\Delta Price_{ij} - \sum_i \sum_j Ex\Delta Price_{ij}\} + \varepsilon_{ij}\end{aligned}$$

*where $\hat{\mu}_{ij}$ is the deviation of current markup of firm ij from its average level, $Ex\Delta Sales_{ij}$ is the expected growth in sales for firm ij, and $Ex\Delta Price_{ij}$ is its next expected price change. Now consider the following decomposition of firms' errors in expecting stochastic discount rates and changes in marginal costs:*

$$\mathbb{E}_t^{ij}\{\hat{q}_{t,t+1}\} - \sum_i \sum_j \mathbb{E}_t^{ij}\{\hat{q}_{t,t+1}\} = u_{1,t}^i + u_{2,t}^{ij}$$
$$\mathbb{E}_t^{ij}\{\Delta\hat{mc}_{t+1}\} - \sum_i \sum_j \mathbb{E}_t^{ij}\{\Delta\hat{mc}_{t+1}\} = v_{1,t}^i + v_{2,t}^{ij}$$

*where $u_{1,t}^i$ and $v_{1,t}^i$ are industry specific errors that are orthogonal to the firm level errors $v_{2,t}^{ij}$ and $u_{2,t}^{ij}$. Assuming that $v_{2,t}^{ij}$ and $u_{2,t}^{ij}$ are independent across firms and are orthogonal to the other terms in the above regression, $\psi_1$ and $\psi_2$ can be identified from $\beta_1$ and $\beta_2$ up to the elasticity of substitution across sectors, $\sigma$.*

*Proof.* See Appendix. □

## 4.2 Results

A distinct characteristic of implicit collusion and customer-base models is that they are derived for oligopolistic firms. Hence, one would expect that in the data, while oligopolistic firms should conform to the above law of motion, competitive ones should not. With that in mind, to test the validity of the the law of motion, I divide the sample into two sub-samples, firms with more than 20 competitors, i.e. competitive firms, and firms with fewer than 20 competitors.[13] The prediction is that coefficients of the law of motion, identified through the regression in Proposition 3, should be significant for the second group, and not significant for the first. Table 2 shows the results of these two regressions, and this hypothesis cannot be rejected at a 95% confidence level.

| | $2 < N \leq 20$ | $N > 20$ |
|---|---|---|
| Expected growth in sales (demeaned) | **0.16**** (0.08) | -0.03 (0.11) |
| Expected size of next price change (demeaned) | **-0.18**** (0.06) | 0.02 (0.09) |
| Observations | 495 | 200 |

Table 2: The table reports the coefficients for the regression specified in Proposition 3, allowing for industry fixed effects. The first column reports the coefficients for firms that report less than 20 but more than 2 competitors. Second column reports the coefficients for firms that report more than 20 competitors.

[13] I also drop firms with less than 2 competitors considering the possibility of a non-binding incentive compatibility constraint for these firms.

Moreover, as the coefficients are significant for the oligopolistic firms, we can use the results of Proposition 3 to infer the implied coefficients on the law of motion for markups, and hence identify the model that is more consistent with the data. Specifically, rejecting the null hypothesis that $\psi_2 > 0$ would lead to rejection of the customer-base model, and similarly rejecting the null that $\psi_1 > 0$ would imply the rejection of the implicit collusion model. The following Proposition shows that while the first hypothesis is rejected by the data, the latter is not.

**Proposition 4.** *Given the estimates in Table 2, and imposing the theoretical bound $|\psi_2| < 1$, so that the law of motion is not divergent, the 0.95% confidence intervals for $\psi_1$ and $\psi_2$ are*

$$\psi_2 \in (-1, -0.06]$$
$$\psi_1 \in [0.003, 0.64]$$

*Therefore, we can reject the null that $\psi_2 > 0$, which implies the rejection of the customer-base models. On the other hand, the null hypothesis that $\psi_1 > 0$ (or $\psi_1 = 0$) cannot be rejected.*

*Proof.* See Appendix. □

Therefore, these results imply a rejection of the customer-base models but are consistent with the implicit collusion models. Given that this class of models points towards procyclical markups for any empirically realistic level of inertia in output, these results suggest that countercyclical markups are unlikely to be a key propagation mechanism of business cycles.

# 5 Conclusion

In this paper, I revisit the implicit collusion and customer-base models and show they both imply a forward looking law of motion for markups in which they depend on the firms' expectations of output growth and stochastic discount rates. Because markups are related to the expected output growth, and not to its level, the conditional expectations of firm for the dynamics of output are the key component of the cyclicality of markups. In particular, if firms expect a humped-shape response for output during the business cycle, the predictions of these models are reversed.

Previous work using these models has not allowed for sufficiently rich dynamics in output which has lead to the conclusion that implicit collusion models lead to countercyclical markups. I show that this prediction is overturned once empirically realistic dynamics of output are incorporated into the model. Doing so, also helps the implicit collusion model to match the empirical evidence on the dynamic cross correlation of output and markups documented in Nekarda and Ramey (2013).

Furthermore, while different versions of customer-base models have been used in the literature to model both procyclical and countercyclical markups, I show that the same result extends to each version of these models; meaning that, given a customer-base model, introducing inertia in response of the output will reverse the cyclicality of markups.

Therefore, matching the humped-shape response of output to shocks is crucial for choosing a proper model for understanding the cyclicality of markups. The common law of motion that I derive in this paper boils down this problem to estimating the signs of its two reduced-form parameters. Using survey data on firms' expectations from New Zealand, I find that firms increase their markup when they expect that their demand will be higher in the future. Combined with the evidence that output responds to shocks with inertia, as firms expect output to keep growing on the impact of an expansionary shock, this then implies that markups are procyclical.

# References

**Bagwell, Kyle and Robert W. Staiger**, "Collusion over the Business Cycle," *The RAND Journal of Economics*, 04 1997, *28* (1), 82–106.

**Bils, Mark, Peter J Klenow, and Benjamin A Malin**, "Testing for Keynesian labor demand," in "NBER Macroeconomics Annual 2012, Volume 27," University of Chicago Press, 2012, pp. 311–349.

**Christiano, Lawrence J, Martin Eichenbaum, and Charles L Evans**, "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of political Economy*, 2005, *113* (1), 1–45.

**Christiano, Lawrence, Martin Eichenbaum, and Sergio Rebelo**, "When Is the Government Spending Multiplier Large?," *Journal of Political Economy*, 2011, *119* (1), pp. 78–121.

**Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar**, "How do firms form their expectations? new survey evidence," *American Economic Review*, 2018, *108* (9), 2671–2713.

**Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu**, "How costly are markups?," Technical Report, National Bureau of Economic Research 2018.

**Gourio, Francois and Leena Rudanko**, "Customer capital," *Review of Economic Studies*, 2014, *81* (3), 1102–1136.

**Green, Edward J. and Robert H. Porter**, "Noncooperative Collusion under Imperfect Price Information," *Econometrica*, 01 1984, *52* (1), 87–100.

**Haltiwanger, John and Joseph E Harrington Jr**, "The impact of cyclical demand movements on collusive behavior," *The RAND Journal of Economics*, 1991, pp. 89–106.

**Kandori, Michihiro**, "Correlated demand shocks and price wars during booms," *The Review of Economic Studies*, 1991, *58* (1), 171–180.

**Kaplan, Greg and Guido Menzio**, "Shopping externalities and self-fulfilling unemployment fluctuations," *Journal of Political Economy*, 2016, *124* (3), 771–825.

**Kimball, Miles**, "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 1995, *27* (4), 1241–77.

**Monacelli, Tommaso and Roberto Perotti**, "Fiscal policy, wealth effects, and markups," Technical Report, National Bureau of Economic Research 2008.

**Nekarda, Christopher J. and Valerie A. Ramey**, "The Cyclical Behavior of the Price-Cost Markup," Working Paper 19099, National Bureau of Economic Research June 2013.

**Paciello, Luigi, Andrea Pozzi, and Nicholas Trachter**, "Markups Dynamics with Customer Markets," 2014.

**Phelps, Edmund S and G Sidney**, "Winter, 1970, Optimal price policy under atomistic competition," *Microeconomic Foundations of Employment and Inflation Theory, Norton New York*, 1970.

**Ramey, Valerie A.**, "Identifying Government Spending Shocks: It's all in the Timing*," *The Quarterly Journal of Economics*, 2011, *126* (1), 1–50.

**Ramey, Valerie A and Matthew D Shapiro**, "Costly capital reallocation and the effects of government spending," in "Carnegie-Rochester Conference Series on Public Policy," Vol. 48 1998, pp. 145–194.

**Ravn, Morten, Stephanie Schmitt-Grohe, and MartÃn Uribe**, "Deep Habits," *The Review of Economic Studies*, 2006, *73* (1), 195–218.

**Rotemberg, Julio J. and Garth Saloner**, "A Supergame-Theoretic Model of Price Wars during Booms," *The American Economic Review*, Jun. 1986, *76* (3), 390–407.

_ **and Michael Woodford**, "Markups and the Business Cycle," in "NBER Macroeconomics Annual 1991, Volume 6" NBER Chapters, National Bureau of Economic Research, Inc, September 1991, pp. 63–140.

**Rotemberg, Julio J and Michael Woodford**, "Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity," *Journal of Political Economy*, December 1992, *100* (6), 1153–1207.

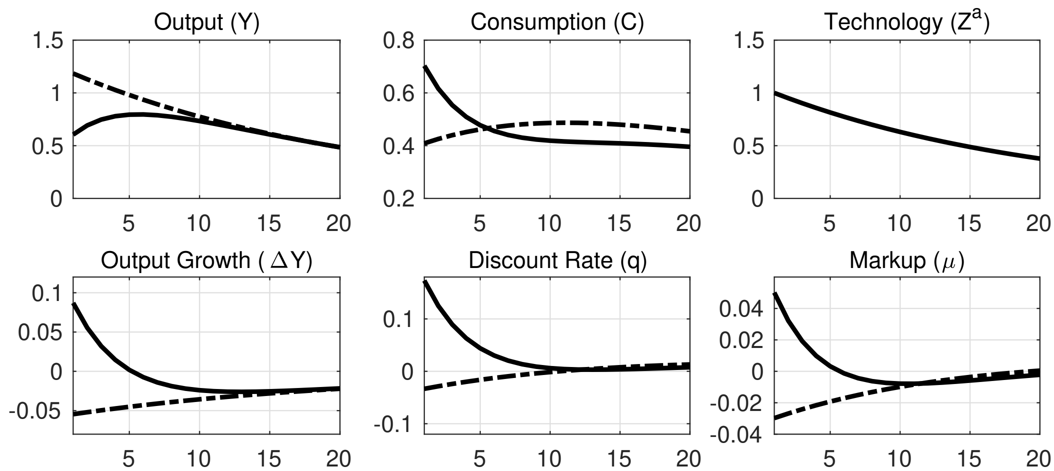
_ **and** _ , "The cyclical behavior of prices and costs," *Handbook of macroeconomics*, 1999, *1*, 1051–1135.

**Sims, Eric R.**, "Permanent and Transitory Technology Shocks and the Behavior of Hours: A Challenge for DSGE Models," 2011.

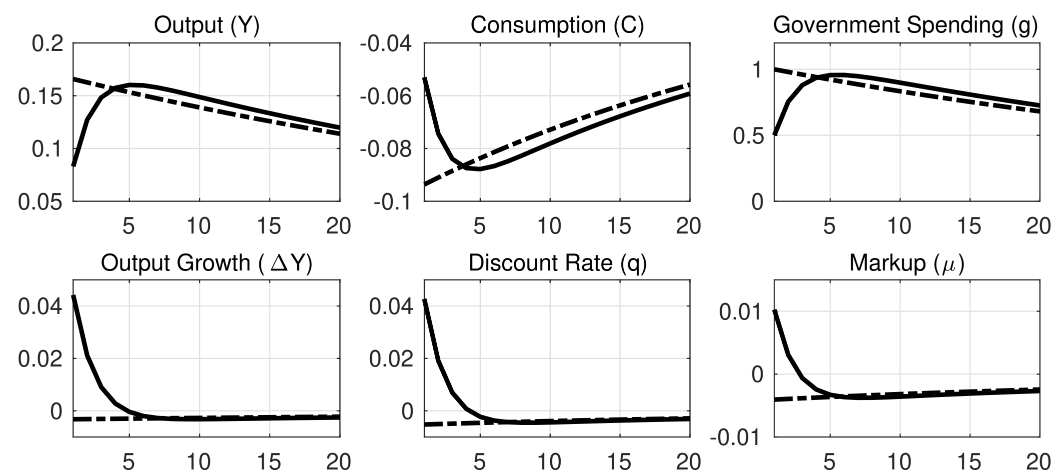
**Smets, Frank and Rafael Wouters**, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *The American Economic Review*, 2007, *97* (3), 586–606.

# A Figures

Figure 1: Impulse Response Functions: Implicit Collusion Model




(a) The dashed curves plot the impulse response functions of the implicit collusion model to a 1% technology shock with no adjustment cost in which markups are counter-cyclical as output growth and stochastic discount rates are counter-cyclical. Solid curves illustrate the impulse response functions of the same model to a 1% technology shock with investment adjustment cost. Markups are pro-cyclical as long as firms expect output to grow. See Section 2.4.2 for details.




(b) The dashed curves plot the impulse response functions of the implicit collusion model to a 1% government spending shock without inertia in which markups are counter-cyclical as output growth is negative during the expansion. Solid curves illustrate the impulse response functions of the same model to an inertial government spending shock that peaks at 1%. Markups are pro-cyclical on impact as output growth and stochastic discount rates are pro-cyclical. See Section 2.4.2 for details.

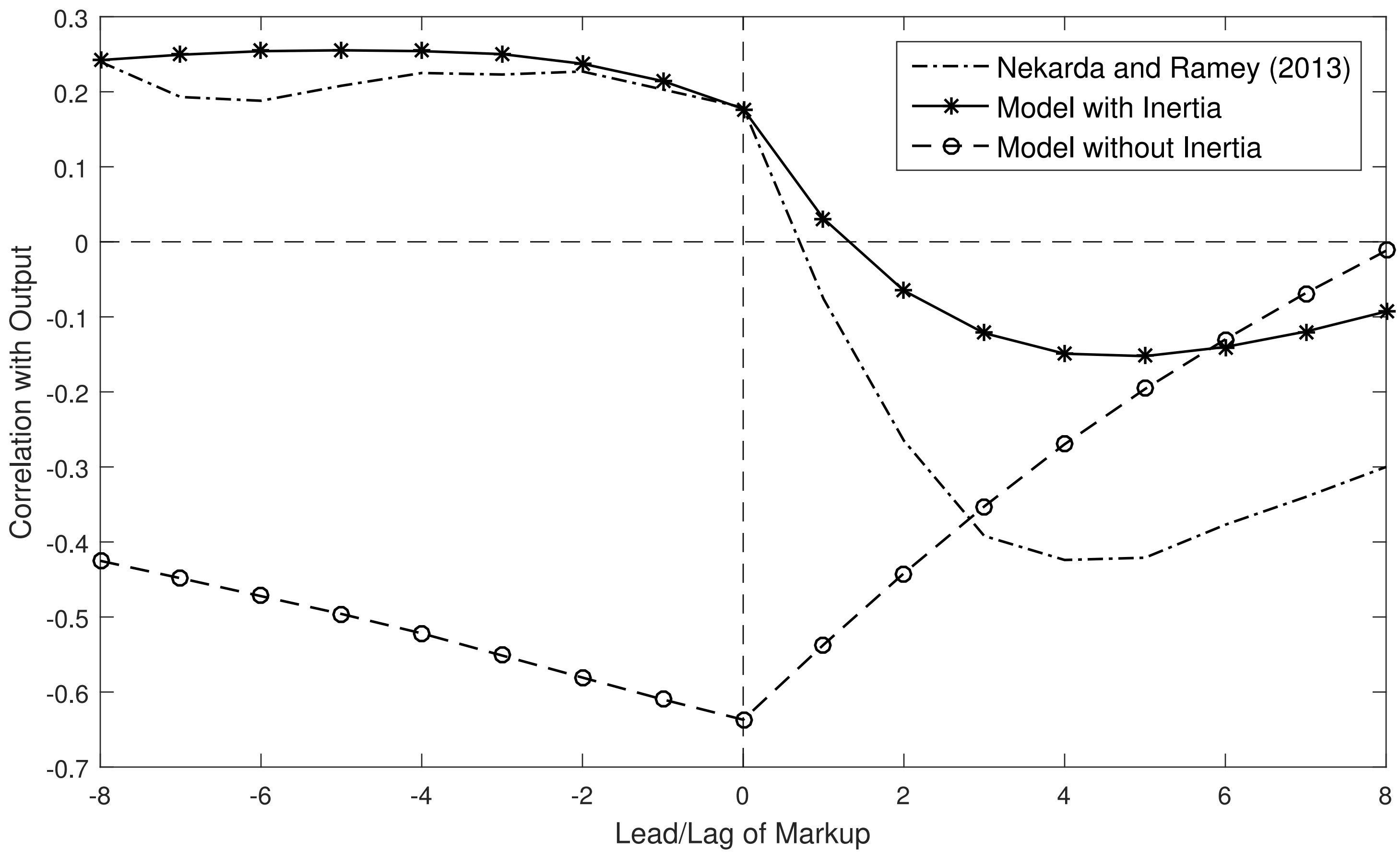
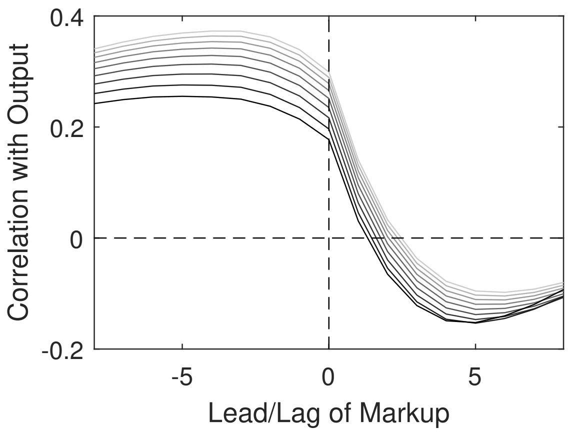
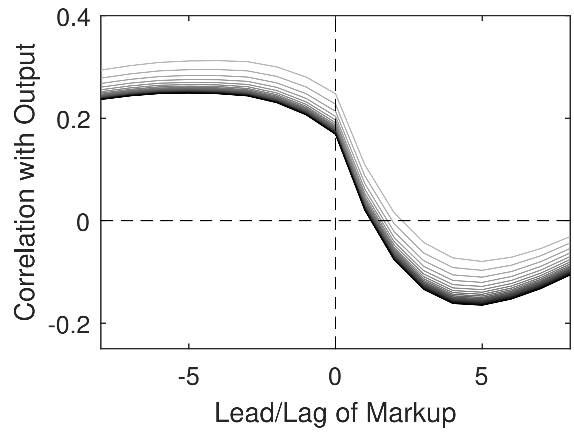



Figure 2: The black curve with square markers depicts correlation of $\mu_{t+j}$ with $Y_t$ from the simulated implicit collusion model without inertial response of output conditional on a TFP shock. The dotted curve shows the unconditional cross-correlation of the cyclical components of markups with real GDP from Nekarda and Ramey (2013). The black curve with circle markers illustrate this cross correlation from the simulated implicit collusion model with inertial response of output conditional on a TFP shock. Inertia is crucial in matching the data. See Section 2.4.3 for details.

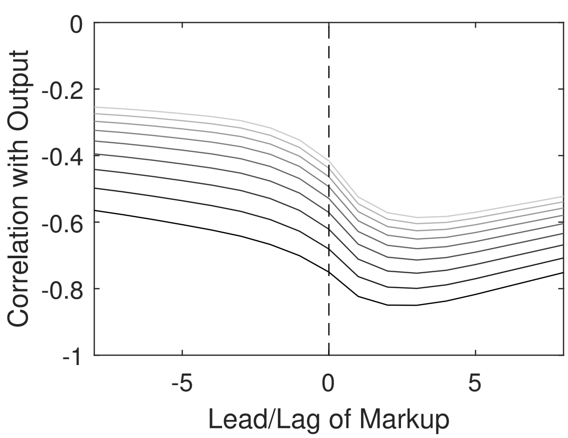
Figure 3: Robustness to number of firms in sectors $N$, and the renegotiation probability $\gamma$



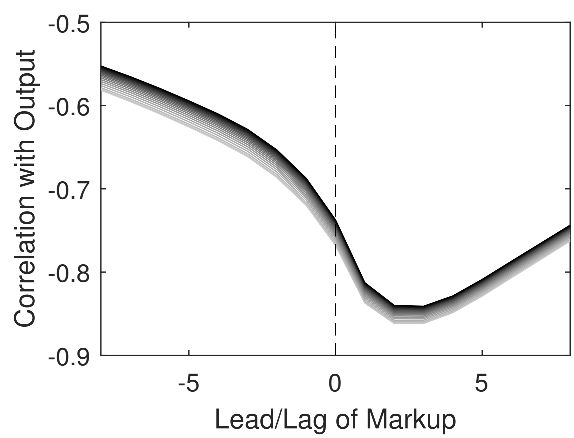

(a) Simulated correlation of $\mu_{t+j}$ with $Y_t$ conditional on a TFP shock for $\gamma \in [0.4, 0.8]$ . See section 2.4.4 for details.




(b) Simulated correlation of $\mu_{t+j}$ with $Y_t$ conditional on a TFP shock for $N \in \{5, \ldots, 25\}$ . See section 2.4.4 for details.

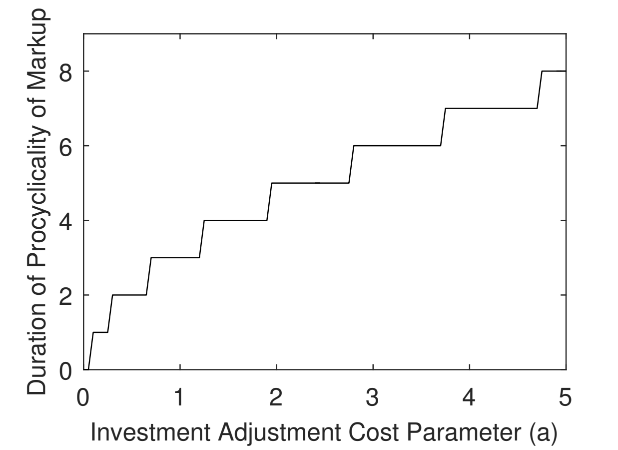



(c) Simulated correlation of $\mu_{t+j}$ with $Y_t$ conditional on a government spending shock for $\gamma \in [0.4, 0.8]$ . See section 2.4.4 for details.

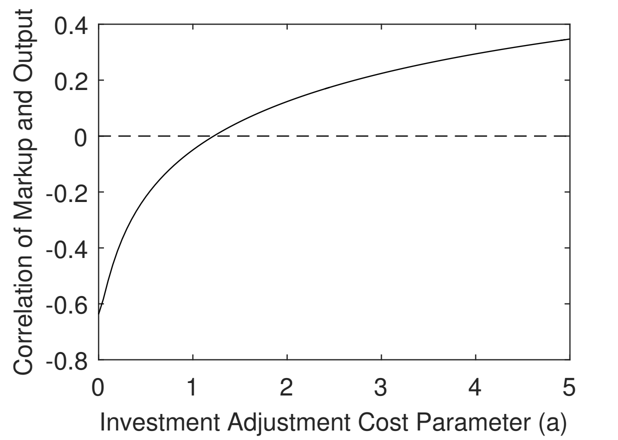



(d) Simulated correlation of $\mu_{t+j}$ with $Y_t$ conditional on a government spending shock for $N \in \{5, \ldots, 25\}$ . See section 2.4.4 for details.

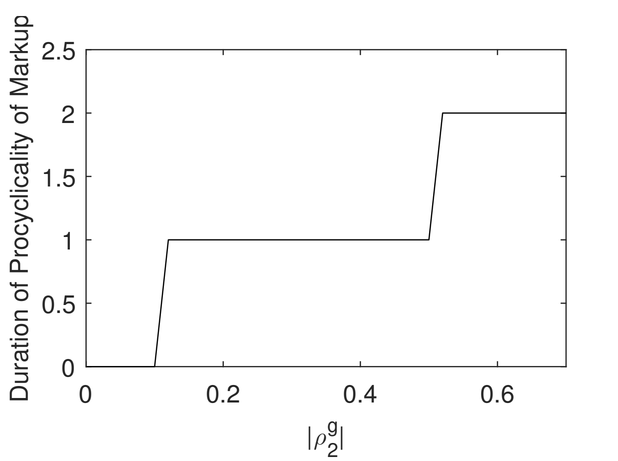
Figure 4: Robustness to inertia



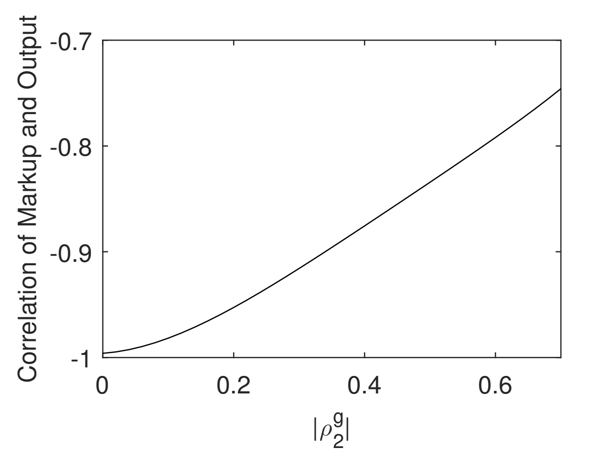

(a) Duration of procyclicality of markup after a 1% TFP shock for different value of investment adjustment cost parameter, $a \in [0,5]$ . See Section 2.4.4 for details.




(b) Simulated correlation of $\mu_t$ with $Y_t$ conditional on a TFP shock for different value of investment adjustment cost parameter, $a \in [0,5]$ . See Section 2.4.4 for details.

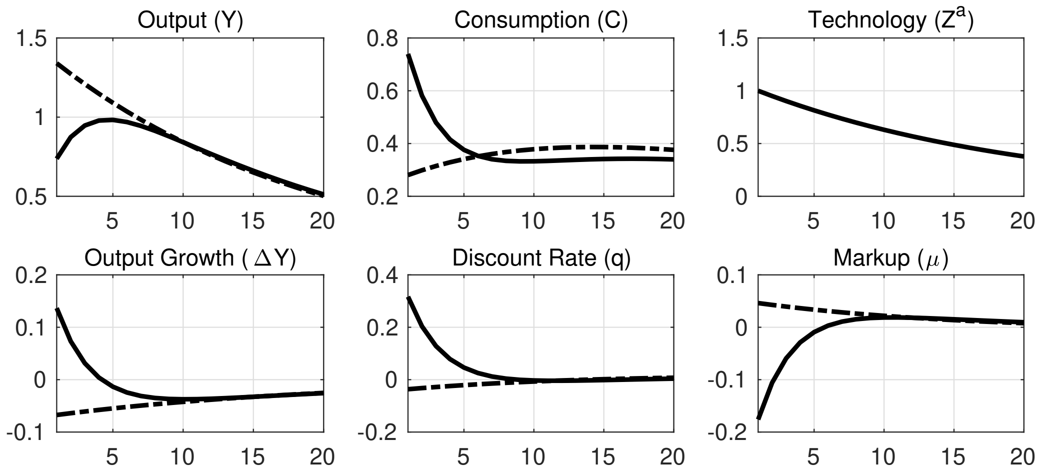



(c) Duration of procyclicality of markup after a 1% government spending shock for different values of the inertia parameter in the AR(2) process, $|\rho_2^g| \in [0, 0.7]$. See Section 2.4.4 for details.

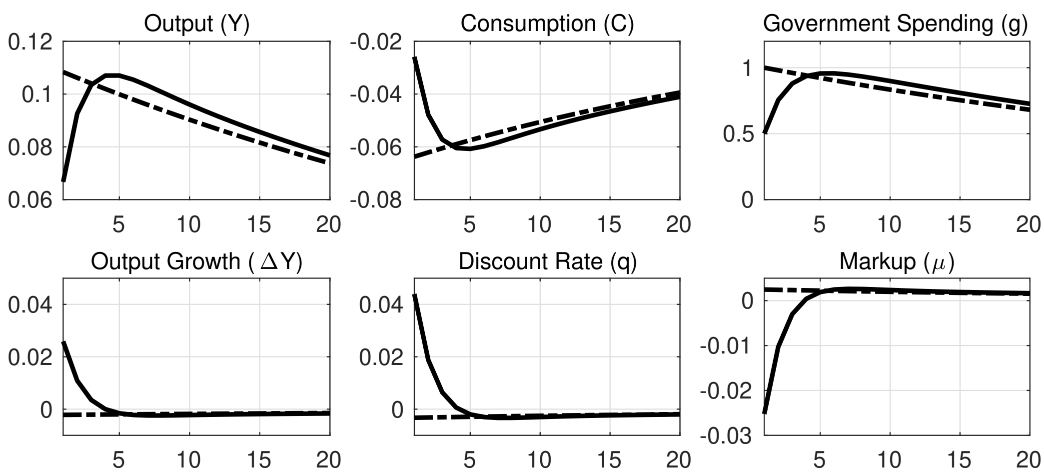



(d) Simulated correlation of $\mu_t$ with $Y_t$ conditional on a government spending shock for different values of the inertia parameter in the AR(2) process, $|\rho_2^g| \in [0, 0.7]$ . See Section 2.4.4 for details.

Figure 5: Impulse Response Functions: Customer-Base Model




(a) The dashed curves plot the impulse response functions of the customer-base model to a 1% technology shock with no adjustment cost in which markups are pro-cyclical as output growth and stochastic discount rates are counter-cyclical. Solid curves illustrate the impulse response functions of the same model to a 1% technology shock with investment adjustment cost. Markups are counter-cyclical as long as firms expect output to grow. See Section 3.1 for details.




(b) The dashed curves plot the impulse response functions of the customer-base model to a 1% government spending shock without inertia in which markups are pro-cyclical as output growth is negative during the expansion. Solid curves illustrate the impulse response functions of the same model to an inertial government spending shock that peaks at 1%. Markups are counter-cyclical on impact as output growth and stochastic discount rates are pro-cyclical. See Section 3.1 for details.

# B Proofs.

**Proof of Proposition 1.**

First, observe that the set of solutions is not empty as $\mu_{i,t} = \mu_{COU}, \forall t$ satisfies the constraint for all periods. Moreover, if the constraint is not binding, the firms will simply act like a monopoly and choose $\mu_{i,t} = \mu_{MON}$, as it maximizes their joint profits. Hence, the choice set of firms can be compactified so that $\mu_{i,t} \in [\mu_{COU}, \mu_{MON}]$, and as the usual assumption of continuity holds, the problem has a solution. Finally, for the solution to be a sub-game Nash equilibrium, two conditions have to hold: first, that firms do not have an incentive to deviate from the chosen markups in the equilibrium path, which is true by construction, and second, that if ever the game were to go to punishment stage, firms would have an incentive to revert back to this strategy, which is also true as collusion is always at least as good as best responding.

**Proof of Proposition 2.**

The first order conditions, after imposing the symmetric equilibrium conditions $\frac{\mu_{ijt}}{\mu_{i,t}} = 1, S_{i,j,t} = 1$ and normalizing $P_t = 1$ (as all firms solve the same problem), are

$$(1 + h'(1))(\mu_t^{-1} - \mu_C^{-1}) - h'(1)(1 - \mu_C^{-1}) = \gamma E_t\{Q_{t,t+1}\frac{Y_{t+1}}{Y_t}(\mu_{t+1}^{-1} - \mu_C^{-1})\}$$

Hence, in the steady state

$$\mu = \frac{1 - \beta\gamma + h'(1)}{1 - \beta\gamma + \mu_C h'(1)}\mu_C$$

Notice that $\mu \geq 1$ if and only if

$$1 - \beta\gamma + \mu_C h'(1) \geq 0 \Leftrightarrow h'(1) \geq -\frac{1 - \beta\gamma}{\mu_C} \tag{5}$$

also observe that

$$\mu - \mu_C = -h'(1)\frac{(\mu_C - 1)\mu_C}{1 - \beta\gamma + \mu_C h'(1)} < 0 \Leftrightarrow h'(1) > 0 \tag{6}$$

meaning that customer-base frictions reduce markups on average if and only if $h'(1)$ is positive.

Finally, a first order approximation to the above equation around the steady state yields:

$$\hat{\mu}_t = \psi_1 \mathbb{E}_t\{\hat{q}_{t,t+1} + \Delta\hat{y}_{t+1}\} + \psi_2 \mathbb{E}_t\{\hat{\mu}_{t+1}\}$$

such that $\psi_1 \equiv \beta\gamma\frac{\mu\mu_C^{-1}-1}{1+h'(1)}$ and $\psi_2 \equiv \frac{\beta\gamma}{1+h'(1)}$. This verifies that the customer-base model, up to a first order approximation, implies the same law of the motion for markups as the implicit collusion model. To infer the signs of $\psi_1$ and $\psi_2$, observe that 5 implies that $h'(1)+1$ as $-\frac{1-\beta\gamma}{\mu_C} > -1$. Thus, $\psi_2 > 1$ for any parametrization. Moreover, $\psi_1 > 0$ if and only if $\mu_C^{-1}\mu - 1 > 0$ which based on 6 happens if and only if $h'(1) < 0$, and vice versa.

**Proof of Proposition 3.**

Notice that

$$\begin{aligned}
Ex\Delta Sales_{i,j,t} &= E_t^{ij}\frac{P_{i,j,t+1}Y_{i,j,t+1} - P_{i,j,t}Y_{i,j,t}}{P_{i,j,t}Y_{i,j,t}} \\
&\approx E_t^{ij}\left[(1-\sigma)\Delta\hat{p}_{i,j,t+1} + \Delta\hat{y}_{t+1}\right] \\
&= (1-\sigma)Ex\Delta Price_{i,j,t} + E_t^{ij}\left[\Delta\hat{y}_{t+1}\right]
\end{aligned}$$

where the second line is derived using the demand structure $Y_{i,j,t} = Y_{i,t} = Y_t D(P_{i,t}; P_{i,t})$. Now, rewriting the law of motion

$$\begin{aligned}
\hat{\mu}_{i,j,t} &= \frac{\psi_1}{1-\psi_2}\mathbb{E}_t^{ij}\left\{\Delta\hat{y}_{t+1} + \hat{q}_{t,t+1}\right\} + \frac{\psi_2}{1-\psi_2}\mathbb{E}_t^{ij}\{\Delta\hat{\mu}_{i,t+1}\} \\
&= \frac{\psi_1}{1-\psi_2}\mathbb{E}_t^{ij}\left\{Ex\Delta Sales_{i,j,t} + (\sigma-1)\Delta\hat{p}_{i,t+1} + \hat{q}_{t,t+1}\right\} + \frac{\psi_2}{1-\psi_2}\mathbb{E}_t^{ij}\{\Delta\hat{p}_{i,j,t+1} - \Delta\hat{mc}_{t+1}\} \\
&= \frac{\psi_1}{1-\psi_2}\mathbb{E}_t^{ij}\left\{\hat{q}_{t,t+1}\right\} + \frac{\psi_1}{1-\psi_2}Ex\Delta Sales_{i,j,t} + \frac{(\sigma-1)\psi_1+\psi_2}{1-\psi_2}Ex\Delta Price_{i,j,t} - \frac{\psi_2}{1-\psi_2}\mathbb{E}_t^{ij}\{\Delta\hat{mc}_{t+1}\}
\end{aligned}$$

Now sum over $i$ and $j$ and subtract the two to get

$$\begin{aligned}
\hat{\mu}_{ij} - \sum_i\sum_j\hat{\mu}_{ij} &= \frac{\psi_1}{1-\psi_2}\{Ex\Delta Sales_{ij} - \sum_i\sum_j Ex\Delta Sales_{ij}\} \\
&+ \frac{(\sigma-1)\psi_1+\psi_2}{1-\psi_2}\{Ex\Delta Price_{ij} - \sum_i\sum_j Ex\Delta Price_{ij}\} \\
&+ \frac{\psi_1}{1-\psi_2}(\mathbb{E}_t^{ij}\{\hat{q}_{t,t+1}\} \\
&- \sum_i\sum_j\mathbb{E}_t^{ij}\{\hat{q}_{t,t+1}\}) - \frac{\psi_2}{1-\psi_2}(\mathbb{E}_t^{ij}\{\Delta\hat{mc}_{t+1}\} - \sum_i\sum_j\mathbb{E}_t^{ij}\{\Delta\hat{mc}_{t+1}\}) \\
&= \frac{\psi_1}{1-\psi_2}\{Ex\Delta Sales_{ij} - \sum_i\sum_j Ex\Delta Sales_{ij}\} \\
&+ \frac{(\sigma-1)\psi_1+\psi_2}{1-\psi_2}\{Ex\Delta Price_{ij} - \sum_i\sum_j Ex\Delta Price_{ij}\} + Industry_FE_i + \varepsilon_{i,j,t}
\end{aligned}$$

where

$$
\begin{aligned}
& Industry\_FE_i \equiv \tfrac{\psi_1}{1-\psi_2}u^i_{1,t} + \tfrac{\psi_2}{1-\psi_2}v^i_{1,t} \\
, \qquad & \varepsilon_{i,j,t} \equiv \tfrac{\psi_1}{1-\psi_2}u^{ij}_{2,t} + \tfrac{\psi_2}{1-\psi_2}v^{ij}_{2,t}
\end{aligned}
$$

Since $u^{ij}_{2,t}$ and $v^{ij}_{2,t}$ are independent of $Indsutry\_FE_i$ by construction and the other two terms by assumption, we have,

$$
\psi_1 = \frac{\hat{\beta}_1}{1+\hat{\beta}_2-(\sigma-1)\hat{\beta}_1},\ \psi_2 = 1 - \frac{1}{1+\hat{\beta}_2-(\sigma-1)\hat{\beta}_1}
$$

**Proof of Proposition 4.**

From the proof of last proposition, observe that

$$
\begin{aligned}
& \tfrac{1}{1-\psi_2} = 0.82 - 0.16(\sigma-1) \pm 1.96 \times (0.06 + 0.08(\sigma-1)) \\
\Rightarrow \quad & \tfrac{1}{1-\psi_2} \in [0.7 - 0.32(\sigma-1), 0.94]
\end{aligned}
$$

Also, since $|\psi_2| < 1$, we have

$$
\frac{1}{1-\psi_2} \in (\frac{1}{2}, \infty)
$$

combining the two we get

$$
\frac{1}{1-\psi_2} \in (0.5, 0.94], \forall \sigma > 1
$$

which implies that

$$
\psi_2 \in (-1, -0.06], \forall \sigma > 1
$$

Also

$$
\frac{\psi_1}{1-\psi_2} = \hat{\beta}_1 = 0.16 \pm 1.96 \times 0.08
$$

Hence,

$$
\begin{aligned}
\psi_1 & \in (1-\psi_2) \times [0.003, 0.32] \\
& \subset [0.003, 0.64], \forall \sigma > 1.
\end{aligned}
$$