Dynamic Oligopoly and Price Stickiness*

Olivier Wang NYU Stern

Iván Werning MIT

March 2022

How does market concentration affect the potency of monetary policy? To tackle this question we build a model with oligopolistic sectors. We provide a formula for the response of aggregate output to monetary shocks in terms of sufficient statistics: demand elasticities, concentration, and markups. We calibrate our model to the evidence on pass-through, and find that higher concentration significantly amplifies non-neutrality. To isolate the strategic effects of oligopoly, we compare our model to one with monopolistic competition recalibrated to ensure firms face comparable demand functions. Finally, we compute an exact Phillips curve for our model. Qualitatively, our Phillips curve incorporates extra terms relative to the standard New Keynesian one. However, quantitatively, we show that a standard Phillips curve, appropriately recalibrated, provides an excellent approximation.

*In memory of Julio Rotemberg, who was ahead of his time, many times. We thank excellent research assistance by Marc de la Barrera, Rebekah Dix, Juliette Fournier, Alex Carrasco and Pedro Martinez Bruera, as well as comments by Fernando Alvarez, Anmol Bhandari, Ariel Burstein, Glenn Ellison, Fabio Ghironi, Francesco Lippi, Simon Mongey, and seminar and conference participants at MIT, UTDT lecture in honor of Julio Rotemberg, EIEF Rome, SED 2018 meetings, Minneapolis Federal Reserve Bank, NBER-SI 2020, Central Bank of Chile, Wisconsin, NYU, PSE, Wharton, Bank of Israel, Yale, and Harvard. All remaining errors are ours.

1 Introduction

The recent rise in product-market concentration has been viewed as a driving force behind several macroeconomic trends.[1] What are the implications of trends in concentration or market power for the transmission of monetary policy? Do strategic interactions in pricing between increasingly large firms amplify or dampen the real effects of monetary shocks? The baseline New Keynesian model, built on the tractable paradigm of monopolistic competition, is not designed to address these questions as there is no notion of market concentration.

In this paper, we provide a new framework to study the link between market structure and monetary policy. We generalize the New Keynesian model by allowing for dynamic oligopolistic competition between any finite number of firms in each sector of the economy. Firms compete by setting their prices, but they do so in a staggered and infrequent manner due to nominal rigidities. We study Markov equilibria of our dynamic game, where the pricing strategy, or reaction function, of every firm is a function of the prices of its competitors. We use this model to study the aggregate real effects of monetary shocks.

Departing from monopolistic competition to oligopoly poses new challenges, as it requires solving a dynamic game with strategic interactions and embedding it into a general equilibrium macroeconomic model. Despite these complexities, our first results derive a closed-form solution for the response of the aggregate price level and output to small monetary shocks. We show that the degree of aggregate price stickiness is conveniently captured by a single measure of strategic complementarities given by the slope of the price-setting reaction function to competitors' prices. Indeed, only the reaction function at a steady state is required. In this way, our result links the partial equilibrium industry dynamics, with the general equilibrium macro response to a monetary shock.

Given the importance of this slope, we next investigate its determinants. We provide a formula for it that inputs three sufficient statistics: market concentration as captured by the effective number of firms within a sector, demand elasticities, and markups.[2] According to our formula, all other things the same, higher observed

[1] For instance, Gutiérrez and Philippon (2017) document an increase in the mean Herfindahl index since the mid-nineties, and argue that it has weakened investment. Autor, Dorn, Katz, Patterson and Van Reenen (2017) and Barkai (2020) relate the rising concentration of sales over the past 30 years in most US sectors to the fall in the labor share.

[2] In the standard monopolistic competition model desired markups are constant and only a function

markups predict larger output responses to monetary shocks. At the heart of this result is the notion that, away from the monopolistic limit, markups reveal the strength of strategic complementarities. This is the case because greater strategic complementarities lead to higher markups in our dynamic oligopoly game.

Our sufficient statistic formula can be used to gauge the present non-neutrality, for a given estimate of these three statistics. All three statistics are endogenous, however, so this formula should not be used for comparative statics. For example, an increase in concentration is likely to affect both demand elasticities and equilibrium markups.

To perform counterfactual experiments, we take a more structural approach and solve the oligopolistic equilibrium in terms of fundamentals. We go beyond CES and use a general homothetic demand with flexible elasticities and superelasticities, as the latter can affect monetary policy transmission through variable markups even under monopolistic competition.

We first vary concentration in each sector while keeping preference parameters fixed. We find that higher concentration can significantly amplify or dampen aggregate price stickiness and therefore the real effects of monetary policy, depending on how properties of demand vary with n.[3] On the one hand, when preferences are restricted to CES, higher concentration unambiguously amplifies stickiness. Maximal effects are attained under duopoly, for which the half-life of the price level and output in reaction to monetary shocks is 40% higher than under monopolistic competition. Translated to a more common measure of non-neutrality, this is a significant effect: we show it is equivalent to dividing the slope of the Phillips curve by $1.4^2 \approx 2$.[4] On the other hand, with non-CES preferences (e.g., Kimball 1995), higher concentration amplifies aggregate stickiness if the superelasticity (the elasticity of the elasticity) of demand is low, as under CES, but has the opposite effect if the superelasticity is high. It is thus essential to first understand the link between concentration and finer properties of demand functions.

We use evidence on the heterogeneity in the pass-through of idiosyncratic cost shocks across small and large firms from Amiti, Itskhoki and Konings (2019) to calibrate how concentration affects the superelasticity of demand, and find that con-

of the demand elasticity. However, in a strategic and dynamic environment the endogenous markup is no longer a simple function of the demand elasticity.

[3]In Section 6 we extend the model to allow for firm heterogeneity within sectors and show that the model with n symmetric firms is an excellent approximation to a model with heterogeneous firms and inverse Herfindahl index $1/n$. Thus a rise in concentration has the same effect for monetary policy whether it comes from higher market shares for larger firms, or from a decrease in the number of firms.

[4]We explain later why the slope of the Phillips curve is the inverse of the square of the half life.

centration amplifies stickiness substantially, even more than under CES: the half-life doubles when going from $n = \infty$ to $n = 3$. Translating, this corresponds to dividing the slope of the Phillips curve by four! Under this calibration the rise in the average Herfindahl index observed in the U.S. since 1990 increases the half-life by 15%, or a one-third reduction in the slope of the Phillips curve.

What explains these results? The number of competitors in a market has an effect on firms' dynamic strategic incentives, but also on the residual demand faced by each firm. On the one hand, "feedback effects" make each firm care about its rivals' current and future prices when setting its price, due to the shape of demand. On the other hand, "strategic effects" arise because each firm realizes its current pricing decision can affect how its rivals will set their prices in the future. Feedback effects are present in monopolistic competitive models with non-CES demand, but strategic effects can only exist when firms are not atomistic.

To isolate these two effects for each n, we compare the oligopolistic model with n firms to a "naive" equilibrium where the n firms correctly predict the path of prices on the equilibrium path, but ignore the off-equilibrium effect of their own price choices on rivals' future prices. The naive equilibrium is also equivalent to a "recalibrated" rational model with $n = \infty$ and Kimball preferences set to match the elasticity and superelasticity of residual demand with finite n.[5]

We find that the feedback effects dictated by the shape of demand explain most of our results. While strategic effects matter for the level of steady state markups, they only have a modest impact on monetary policy transmission. Of course, this quantitative conclusion that strategic effects play a small role cannot be reached offhand, but only after solving the full, strategic, model as we have done.

We conclude by generalizing our model and analysis: we allow for more general preferences and go beyond permanent money supply shocks to derive an exact "Phillips curve", that is, an equilibrium relationship between inflation and other equilibrium variables such as real output. The Phillips curve can be used to study any type of shock, with any degree of persistence. For example, we study interest rate shocks under a Taylor rule, or the effects of hitting the zero lower bound, or news shocks about these changes. One can also use it to study real shocks, for a given monetary policy rule.

Relative to the standard New Keynesian model, which is a simply first-order dy-

[5] Another natural benchmark is a model with $n = \infty$ that matches the same own-cost pass-through as the oligopolistic model. It is not equivalent but very close to the naive model.

namical system, our Phillips curve includes higher-order terms. Thus, these terms can in principle generate endogenous inflation persistence and cost-push shocks. However, we find again that for a wide range of shocks the equilibrium with naive firms provides an accurate approximation to the strategic model. Since the naive model is equivalent to a monopolistic setup, this implies that a standard (first-order) New Keynesian Phillips curve provides an excellent approximation to the actual (higher-order) Phillips curve.

Overall, our results show that the monopolistic model with an appropriately chosen Kimball demand provides an approximation to an oligopolistic reality. Although this virtual equivalence is true and is useful in a reduced form way, in a deeper sense, it does not imply that oligopoly is irrelevant.

First, one must choose the Kimball demand correctly, in a manner that depends on market concentration. Our framework provides a rigorous mapping from micro-evidence on pass-through and concentration to the reduced-form Kimball parameter driving these models.

Second, the oligopoly model yields a unique link between markups and monetary policy transmission, in the aggregate and in the cross-section, that cannot arise under monopolistic competition, even with non-CES demand. Under monopolistic competition, predictions of the model depend on calibrating two independent parameters of demand functions: the markup only depends on the elasticity, and the price response to monetary policy only depends on the superelasticity. Oligopolistic competition, on the other hand, highlights a tight connection: the superelasticity of demand has a positive effect on both markups and the pass-through of monetary policy. Therefore, our model predicts that monetary policy is transmitted relatively more through sectors or regions with higher markups all else equal, because they are the ones featuring the slowest price adjustment following monetary shocks.

Related Literature

An important early exception to the complete domination of monopolistic competition in the macroeconomics literature on firm pricing is Rotemberg and Saloner (1986), who propose a model of oligopolistic competition to explain the cyclical behavior of markups. Rotemberg and Woodford (1992) later embed their model into a general equilibrium framework with aggregate demand shocks driven by government spending. These two papers assume flexible prices and abstract from monetary policy.[6] Another important difference is that we focus on Markov equilibria, in line with the more recent industrial organization literature, rather than trigger-strategy price-war equilibria.

The first paper to combine non-monopolistic competition and nominal rigidities in general equilibrium is Mongey (2018). This paper uses a rich quantitative model with two firms, menu costs, and idiosyncratic shocks to show that duopoly can generate significant non-neutrality relative to the Golosov and Lucas (2007) benchmark. It also finds that duopoly is closer to monopolistic competition under Calvo price-setting than with menu costs. Our paper takes a complementary approach, more analytical but assuming Calvo pricing and abstracting from idiosyncratic shocks.[7] This allows us to go beyond two firms and explore different questions, in particular by changing industry concentration, separating strategic complementarities from residual demand effects, and allowing for arbitrary shocks through the Phillips curve.[8] Modeling more than two firms also lets us incorporate recent evidence linking cost pass-through and market shares from Amiti, Itskhoki and Konings (2019) to infer the relation between concentration and monetary non-neutrality. This evidence implies that even under Calvo pricing, oligopoly leads to significant amplification.

The literature on variable markups in international trade highlights the importance of market structure for cost (e.g., exchange rate) pass-through in static settings (e.g., Atkeson and Burstein 2008). We study a dynamic general equilibrium version of these models, as is needed to analyze monetary policy, and show how to map pass-through estimates to aggregate effects of monetary policy. Our results also share some of the mechanisms studied in partial equilibrium in the industrial organization literature exploring the link between market structure, demand systems and pass-through of costs to prices in models featuring menu costs (Slade 1998, Neiman 2011), non-CES demand systems (Goldberg and Verboven 2001), or both (Nakamura and Zerom 2010).

[6] Rotemberg and Saloner (1987) study a static partial-equilibrium menu-cost model, comparing the incentive to change prices under monopoly and duopoly.

[7] Calvo pricing remains an important benchmark in the literature on price stickiness, due to its tractability, but additionally, recent work on menu costs, such as Gertler and Leahy (2008), Midrigan (2011), Alvarez, Le Bihan and Lippi (2016b) and Alvarez, Lippi and Passadore (2016a), show that certain menu-cost models may actually behave close to Calvo pricing.

[8] Several papers, including Benigno and Faia (2016) and Corhay, Kung and Schmid (2020) with Rotemberg pricing and Etro and Rossi (2015) and Andrés and Burriel (2018) with Calvo pricing, consider models of monopolistic competition that depart from the standard CES setting because the demand curve faced by a firm depends on the number of competitors; but firms still behave atomistically, taking rivals' current and future prices as given.

Kimball (1995) introduced non-CES aggregators that increase non-neutrality even under monopolistic competition.[9] As we show in Section 5, there is a close connection between this class of models (e.g., Klenow and Willis 2016, Gopinath and Itskhoki 2010) and our oligopolistic model. By making the market structure explicit, our paper provides foundations for the dynamic pricing complementarities embedded in the monopolistic Kimball aggregator, in a way consistent with the data on firm size and long-run pass-through. Relative to this strand of the literature, the oligopolistic model also generates unique predictions on the cross-sectional relation between markups, concentration, and monetary policy transmission.

In addition to the dynamic pricing with staggered price stickiness we focus on, market structure can affect the degree of monetary non-neutrality through other margins. Nakamura and Steinsson (2013) organize sources of complementarities in pricing into "micro" (e.g., variable markups or decreasing returns to scale) and "macro" complementarities (e.g., intermediate inputs). Afrouzi (2020) studies the incentives to acquire information in a flexible prices rational-inattention oligopolistic model.

We focus on short-run dynamics holding concentration fixed. It would be interesting to incorporate endogenous entry and exit to study the feedback between fluctuations in output and concentration, as in, e.g., Bilbiie, Ghironi and Melitz (2007). The challenge would be to solve for an additional fixed point: monetary shocks leading to higher output may stimulate entry; but the resulting higher number of firms would imply a faster aggregate price adjustment, which dampens the output response that stimulated entry in the first place.

2 A Macro Model with Oligopolies

The household side of our model is standard. Things are more interesting on the firm side: we depart from monopolistic competition to introduce oligopolies.

Basics. Time is continuous with an infinite horizon $t \in [0, \infty)$.[10] We abstract from aggregate uncertainty and focus on an unanticipated shock.

[9] Other non-CES preferences achieve the same purpose (e.g., translog preferences in Bergin and Feenstra, 2000). We focus on Kimball preferences for concreteness, but our results apply to any homothetic preferences, including those in the wide class studied in Matsuyama and Uschchev (2017).

[10] Our analysis translates easily to a discrete-time setup, but continuous time has a few advantages and permits comparisons with the menu-cost literature (e.g., Alvarez and Lippi, 2014).

There are three types of economic agents: households, firms and the government. Households are described by a continuum of infinitely lived agents that consume nondurable goods and supply labor. The government controls the money supply, provides transfers and issues bonds, to ensure it satisfies its budget constraint.

Firms produce across a continuum of sectors $s \in S$. Each sector is oligopolistic, with a finite number n_s of firms $i \in I_s$, each producing a differentiated variety. Firms can only reset prices at randomly spaced times, so the price vector within a sector is a state variable. By setting $n_s \to \infty$ or $n_s = 1$ we obtain a standard monopolistic setup, where each firm has a negligible effect on competitors. Away from these limit cases there are strategic interactions across firms within a sector, but not across sectors. As we spell out below, this induces a dynamic game in each sector. We focus on Markov equilibria.

Household Preferences. Utility is given by

$$\int_0^\infty e^{-\rho t} U(C(t), \ell(t), m(t)) dt,$$

where $\ell(t)$ denotes labor, $m(t) = M(t)/P(t)$ denotes real money balances and aggregate consumption at any point in time satisfies

$$C = \left(\int_S C_s^{1-\frac{1}{\omega}} ds \right)^{\frac{1}{1-\frac{1}{\omega}}}$$
$$C_s = H_s\left(\{c_{i,s}\}_{i \in I_s} \right),$$

with $C = \exp \int_S \log C_s ds$ when $\omega = 1$ where $\{C_s\}$ and $\{c_{i,s}\}$ are sectoral consumption across sectors $s \in S$ and good varieties across firms $i \in I_s$ within each sector. H_s is homogeneous of degree one and can be more general than CES (e.g. Kimball).

In most of the paper we adopt the Golosov and Lucas (2007) specification

$$U(C, \ell, m) = \frac{C^{1-\sigma}}{1-\sigma} + \psi \log m - \ell.$$

As is well known, these preferences help simplify the aggregate equilibrium dynamics; however, we consider more general preferences in Section 7.

Household Budget Constraints. The flow budget constraint at any $t \geq 0$ is

$$P(t)C(t) + \dot{B}(t) + \dot{M}(t) = W(t)\ell(t) + \Pi(t) + T(t) + R(t)B(t),$$

where (dropping the t dependence) B are nominal bond holdings, R is the nominal interest rate, M money holdings, W the nominal wage, T nominal lump-sum transfers, $\Pi = \int \sum_{i \in I_s} \Pi^{i,s}\, ds$ denotes aggregate firm nominal profits, and P the price index described below. Imposing the No Ponzi condition $\lim_{t \to \infty} e^{-\int_0^t R(s)ds}(B(t) + M(t)) \geq 0$ gives the present value condition

$$\int_0^{\infty} e^{-\int_0^t R(s)ds}(P(t)C(t) + R(t)M(t) - T(t) - W(t)\ell(t) - \Pi(t))dt = M(0) + B(0).$$

Prices and Price Indices. At every point in time, let the vector of prices within a sector s be

$$p_s = (p_{1,s}, p_{2,s}, \ldots, p_{n_s,s})$$

and let $p_{-i,s} = (p_{1,s}, \ldots, p_{i-1,s}, p_{i+1,s}, \ldots, p_{n,s})$ so that $p_s = (p_{i,s}, p_{-i,s})$.

The aggregate price index is given by $P = \left(\int P_s^{1-\omega} ds\right)^{\frac{1}{1-\omega}}$ for $\omega \neq 1$ or $P = \exp \int \log P_s ds$ for $\omega = 1$ where P_s is the sectoral price index, defined by the unit cost condition $P_s = \min_{c_{i,s}} \sum p_{i,s} c_{i,s}$ subject to $H_s(\{c_{i,s}\}_{i \in I_s}) = 1$.

Demand. The demand for firm $i \in I_s$ can be written as

$$y_{i,s}(t) = C(t)P(t)^{\omega}\ d^{i,s}(p_{i,s}(t))$$

for a demand function $d^{i,s}$ that depends only on prices in sector s. The term $C(t)P(t)^{\omega}$ captures aggregate time-varying effects on demand. The individual demand function $d^{i,s}$ captures both within-sector substitution and across-sector substitution.[11] We assume that goods within a sector are gross substitutes: $d^i_j > 0$ for $i \neq j$. This generalizes the standard assumption that goods are more substitutable within than across sectors (i.e., if H_s is a CES aggregator with elasticity of substitution η then $d^i_j > 0$ is equivalent to $\eta > \omega$).

Firms. Each firm $i \in I_s$ in sector $s \in S$ produces from labor according to the production function,

$$y_{i,s}(t) = z_{i,s} f\left(\ell_{i,s}(t)\right)$$

where f is increasing and differentiable. If f is concave then this captures decreasing returns. We first assume no differences in productivity within sectors, so that the n_s

[11]Indeed, $d^{i,s}(p_s)$ is homogeneous of degree $-\omega$.

firms are symmetric, and thus normalize to $z_{i,s} = 1$.[12] Section 6 extends the analysis to heterogeneous firms.

Profits for firm i are

$$\Pi^{i,s}(p_{i,s}, p_{-i,s}; t) = C(t)P(t)^{\omega} d^i(p_{i,s}, p_{-i,s})p_{i,s} - W(t)f^{-1}\left(C(t)P(t)^{\omega} d^i(p_{i,s}, p_{-i,s})\right).$$

Firms receive opportunities to change their price $p_{i,s}$ at random intervals of time determined by a Poisson arrival rate $\lambda_s > 0$, the realizations of which are independent across firms and sectors. Between price changes, firms meet demand at their posted prices. They maximize the present value of profits

$$\mathbb{E}_0 \int_0^{\infty} e^{-\int_0^t R(s)ds}\ \Pi^{i,s}(p_{i,s}(t), p_{-i,s}(t); t)dt.$$

Although there is no aggregate uncertainty, the expectations averages over idiosyncratic realization of times at which firms can change their prices.

Markov Equilibrium. A strategy for firm i specifies its desired reset price at any time t should it have an opportunity to change its price. A Markov equilibrium involves a strategy that is a function only of the price of its rivals and calendar time t,

$$g^{i,s}(p_{-i}; t).$$

The dependence of $g^{i,s}$ on t is required to accommodate monetary shocks and the ensuing transition with possibly time-varying aggregates $C(t)P(t)^{\omega}$, $W(t)$ and $R(t)$. The general non-stationary Hamilton-Jacobi-Bellman equation and optimality condition are detailed in Appendix B. In Section 2.1 below we describe the stationary case.

Given that firms are symmetric within sectors, we consider strategies that are symmetric $g^{i,s} = g^s$. We do not require the equilibrium to be unique: if there are multiple equilibria, our results apply to each one of them.

Equilibrium Definition. Given initial prices $\{p_{i,s}(0)\}$, an equilibrium is given by paths for the aggregate price $P(t)$, wage $W(t)$, interest rate $R(t)$, consumption $C(t)$, labor $\ell(t)$ and money supply $M(t)$, as well as demand functions for consumers $d^{i,s}$ and strategy functions for firms $g^{i,s}$ such that: (a) consumers optimize quantities taking as given the sequence of prices and interest rates; (b) each firm's reset price

[12] Any differences across sectors $z_{i,s} = z_s$ can be absorbed into the units used to measure consumption in sector s, so that setting $z_s = 1$ is without loss in generality.

strategy $g^{i,s}$ is optimal, given the path for $P(t), C(t)$, its rivals' strategies $g^{j,s}$ and the demand functions d; (c) consistency: the aggregate price level evolves in accordance with the reset strategy g employed by firms; (d) markets clear: firms meet demand for goods, $y_{i,s}(t) = c_{i,s}(t)$, the supply of labor equals demand

$$\ell(t) = \int \sum_{i \in I_s} \ell_{i,s}(t) ds$$

and the demand for money equals supply, both denoted by $M(t)$ so implicitly imposed already.

2.1 Stationary Markov Equilibrium

We first study the dynamics within a sector in partial equilibrium, that is, assuming all conditions external to the sector (i.e., the wage, the nominal discount rate, aggregate consumption and price) are constant. The resulting oligopoly game within such as sector is then stationary. This partial equilibrium analysis also characterizes a steady state in general equilibrium. We later show that these within sector dynamics also help characterize the aggregate macroeconomic adjustment to a permanent monetary shock.

We focus on a sector and omit the notation $s \in S$. In a stationary game we can suppress the dependence on t in the Bellman equation, and a stationary Markov equilibrium is characterized by

$$\rho V^i(p) = \Pi^i(p) + \lambda \sum_j \left(V^i(g(p_{-j}), p_{-j}) - V^i(p) \right) \tag{1}$$

where $g(p_{-i}) \in \arg\max_{p_i} V^i(p_i, p_{-i})$ with necessary condition

$$V^i_{p_i}(g(p_{-i}), p_{-i}) = 0. \tag{2}$$

With Poisson rate λ, one of the n firms indexed by j (including firm i) will adjust its price to $g(p_{-j})$, which will make firm i's value jump to $V^i \left(g^j(p_{-j}), p_{-j} \right)$.[13] A useful simple observation is that g only depends on ρ and λ through the ratio λ/ρ. Let $\bar{p}$ denote the steady state price, satisfying $\bar{p} = g(\bar{p})$.

[13] $V^i \left(g^j(p_{-j}), p_{-j} \right)$ serves as shorthand notation for $V^i \left(p_1, \ldots, p_{j-1}, g^j(p_{-j}), p_{j+1}, \ldots, p_n \right)$.

Reaction Slope. We focus on equilibria with differentiable value and reaction functions. By symmetry, at the steady state price $\bar{p}$, the slope $\frac{\partial g}{\partial p_j}(\bar{p})$ does not depend on j. We scale this slope by the number of rivals and define

$$B = (n-1)\frac{\partial g}{\partial p_j}(\bar{p}).$$

To a first order approximation, firm i resets its price to

$$\log p_i = \log \bar{p} + B \times \frac{\sum_{j\neq i}\left(\log p_j - \log \bar{p}\right)}{n-1}. \tag{3}$$

Thus, B represents the reaction to average rival prices.

The slope parameter B will play a starring role in our analysis. It will serve as a unifying concept capturing strategic complementarities in pricing, whether they arise from dynamic oligopoly, non-CES demand, or decreasing returns to scale in production. In the basic model with monopolistic competition, CES demand, and constant returns, $B = 0$.

Limit $\lambda/\rho \to 0$: Static Bertrand-Nash equilibrium. When prices are infinitely sticky or firms are infinitely impatient, so that $\lambda/\rho \to 0$ then (1) implies that $V(p) \to \Pi(p)$ and, thus, firms play a static best-response. Intuitively, they take the current prices of other firms as fixed forever. The equilibrium then converges to a static Bertrand-Nash:

$$\lim_{\lambda/\rho\to 0} g(p_{-i}) = g^{\text{Nash}}(p_{-i}) = \arg\max_{p_i} \Pi^i\left(p_i, p_{-i}\right).$$

The steady state price $\bar{p}$ converges to the Bertrand-Nash price denoted p^{Nash}.

Remark 1. Our focus on differentiable equilibria rules out the typer of "kinked demand curve" and "Edgeworth cycles" Markov equilibria studied by Maskin and Tirole (1988). They considered a Bertrand duopoly ($n = 2$) model with *perfectly substitutable* goods (a particular limit of the d function) as firms become infinitely patient ($\rho \to 0$); in our setting the latter is isomorphic to fixing any $\rho > 0$ but taking the flexible prices limit $\lambda \to \infty$. Under these conditions, they showed that firms can effectively "collude" around the joint monopoly price by using strategies that are non-monotonic in the rival's price.

Similar non-monotone equilibria of this type are possible in our model in some cases. However, in practice, numerical explorations away from their limiting assumptions of perfect substitutability and price flexibility show that for a wide range of parameters Markov equilibria strategies are indeed monotonic and consistent with the differentiable ones we study.[14] Indeed, with linear demand functions d the game has a Markov equilibrium in linear strategies. This is also consistent with the subsequent IO literature, which has focused on Markov equilibria not displaying the form of tacit "Edgeworth cycle" collusion explored in Maskin and Tirole (1988).

3 Monetary Shocks: Dynamics and Sufficient Statistics

We now study an unanticipated permanent shock to money. We suppose the economy is initially in a steady state: constant aggregates P_-, M_-, C_-, ℓ_-, W_-, $R_- = \rho$ and prices in each sector at their steady state $p_{i,s} = \bar{p}_s$. Consider a permanent monetary shock arriving at $t = 0$ so that $M(t) = M_+ = (1+\delta)M_-$ for all $t \geq 0$.

After the shock, sectors readjust towards their steady state, but do so in a random manner that depends on the realizations of the random Calvo price adjustment opportunities across the finite number of firms within a sector. At the aggregate level, however, sectoral idiosyncratic uncertainty averages out, producing deterministic paths for the aggregate price level and consumption.

3.1 Exact Dynamics: Partial Equilibrium To General Equilibrium

Firms must forecast the path that macroeconomic variables will take after the shock. Any given path for aggregates determines a Markov reset price strategy $g^{i,s}$. These strategies, in turn, determine the evolution of aggregates. It is possible to solve this fixed-point problem quite generally as we do in Section 7, but we first focus on a simple case. In the spirit of Golosov and Lucas (2007), our assumptions on preferences lead to the following simplification:

[14] Figure A.11 displays the locus of existence of these monotone equilibria in the (λ, η)-space (where η is the within-sector elasticity of substitution). While the curse of dimensionality prevents us from solving numerically for the full MPE with general n, we conjecture that the region of existence of these equilibria increases with the number of firms, since a higher n reduces the potential monopoly profit (the case of monopolistic competition $n \to \infty$ being an extreme example). Similarly, a higher outer elasticity ω lowers the joint monopoly profit, which should also enlarge the region of existence of the monotone equilibrium.

Proposition 1. *Equilibrium aggregates satisfy*

$$W(t) = (1+\delta)W_{-}, \quad R(t) = \rho, \quad P(t)C(t)^{\sigma} = \rho M_{+}. \tag{4}$$

If in addition

$$\omega\sigma = 1 \tag{5}$$

then the equilibrium prices after the shock satisfy $p_{i,s} = (1+\delta)\,\hat{p}_{i,s}$ *and the normalized prices* $\hat{p}$ *are initialized at* $\hat{p}_{i,s}(0) = \frac{p_{i,s}(0)}{1+\delta}$ *and evolve according to the reaction function* g *of the stationary game in Section 2.1, so that upon resetting firm* i *sets*

$$\hat{p}_{i,s} = g\left(\hat{p}_{-i,s}\right).$$

Proposition 1, proved in Appendix A.1, is extremely useful. It provides conditions under which firm reset prices may ignore the transitional dynamics of macroeconomic variables following the monetary shock. This result allows us to extend the partial equilibrium analysis to general equilibrium. This is an exact result, not an approximation for small monetary shocks (as in Alvarez and Lippi, 2014). Until Section 7 , we consider preferences that satisfy $\omega\sigma = 1$.

For concreteness imagine a positive shock $\delta > 0$. Following the shock the interest rate is unchanged and the nominal wage rises permanently in proportion with money. The left panel of Figure 1 displays price dynamics within a sector (for a duopoly), following a cobweb adjustment process, but with the times of adjustment randomly determined. The right panel displays the paths for aggregates $P(t)$ and $C(t)$. In the long run, normalized prices $\hat{p}$ converge back to their steady state, so that actual prices adjust proportionally by the factor $1+\delta$ (or δ in logs). On impact, prices are unchanged, but $C(t)$ rises above its steady state value by a factor $(1+\delta)^{\frac{1}{\sigma}}$. Over time, as prices rise, $P(t)C(t)^{\sigma}$ remains constant, so consumption falls, eventually returning to its steady state value. Although the nominal interest rate is unchanged, the real interest rate falls along the transition due to the rise in inflation, explaining the temporary rise in consumption.

The classic paper by Rotemberg and Saloner (1986) studied a partial equilibrium model of oligopoly, facing exogenous fluctuations in demand without price rigidities. They assumed a fixed real interest rate. Their analysis focused on non-Markov trigger strategies that sustain "collusive" prices in bad times, but lead to price wars during booms (creating an amplification mechanism for output). In their model, price wars occur because booms are periods with higher demand and, thus, high temporary

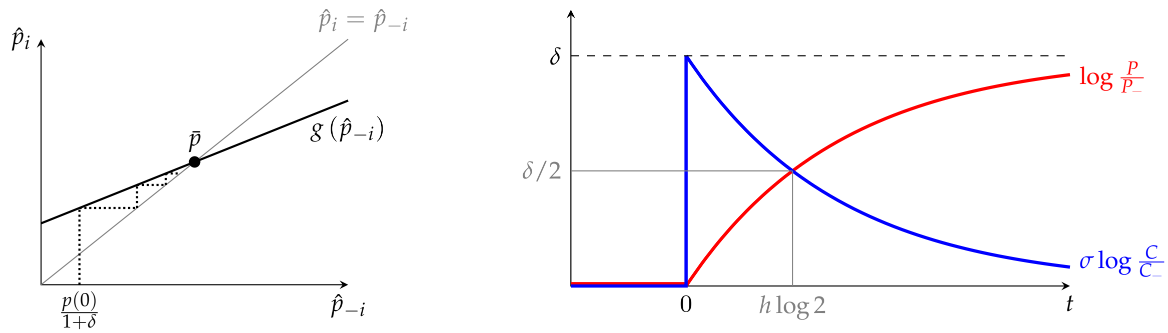

Figure 1: Dynamics following a monetary shock δ. Left: Price dynamics within a sector. Right: Aggregate dynamics. $h \log 2$ denotes the half-life of the price level.

profits to compete over, so the incentive to raise prices is greater. A similar effect is generally present in our model. However, Proposition 1 shows that when $\omega\sigma = 1$ these effects may not be present in general equilibrium since $P(t)$ moves in the opposite direction. An equivalent way to describe this is that Rotemberg and Saloner (1986) assumed a fixed real interest rate, but in our model a boom lowers real interest rates, which makes firms care about the future more, exactly counterbalancing the increase in stakes from higher current profits.

3.2 Approximate Dynamics: The Importance of B_s

We are interested in the speed of convergence of the price level to its new steady state $\bar{P} = (1+\delta) P_-$. From (4), $\log P(t) + \sigma \log C(t)$ is constant after the monetary shock so this also gives us the speed of convergence of output. The next proposition studies the approximate dynamics of these paths (proof in Appendix A.2).

Proposition 2. *Suppose $\omega\sigma = 1$, then to first order in the size of the monetary shock δ,*

$$\log P(t) - \log \bar{P} = -\delta \int_s \zeta_s e^{-\lambda_s (1-B_s) t} ds, \tag{6}$$

$$\log C(t) - \log \bar{C} = \frac{\delta}{\sigma} \int_s \zeta_s e^{-\lambda_s (1-B_s) t} ds, \tag{7}$$

where $\zeta_s = \bar{P}_s \bar{C}_s / (\bar{P}\bar{C})$ is the steady state expenditure share of sector s.[15] *The present-value*

[15]Away from $\omega\sigma = 1$, one can show that $P(t)$ follows (6) except that B_s is replaced with $\alpha_s + B_s$, where α_s captures firms' incentives to increase or decrease prices in response to aggregate fluctuations; $\alpha_s = 0$ in each sector when $\omega\sigma = 1$ or in the limit $\lambda_s/\rho \to 0$.

output effect of the shock discounted at any rate $r \geq 0$ is

$$\int_0^\infty e^{-rt} \log\left(\frac{C(t)}{\bar{C}}\right) dt = \frac{\delta}{\sigma} \int_s \frac{\zeta_s ds}{r + \lambda_s (1 - B_s)}. \tag{8}$$

This proposition summarizes the approximate dynamics for money shocks for our oligopolistic model. It takes the reduction in complexity obtained in Proposition 1 one step further, to closed-form solutions requiring only a handful of statistics, including the frequency of price adjustments λ_s and the slopes B_s. Intuitively, for small shocks the dynamics are dictated by the slope of g around a steady state.

Our proposition highlights a crucial role for the statistic B_s that provides a strong unifying principle: complementarities in pricing at the sectoral level affect the aggregate response to monetary shocks captured in a simple partial equilibrium statistic, B_s. If B_s is high then prices converge more slowly contributing towards larger real effects. Our result also allows for heterogeneity in price frequency λ_s and the slope parameter B_s across sectors, showing how how these are aggregated; we come back to this aspect later.[16]

There are a few ways to depart from the benchmark $B = 0$ case and obtain strategic complementarities $B > 0$. Two are standard in the literature, even under monopolistic competition ($n = \infty$): non-CES (e.g., Kimball) demand and decreasing returns to scale.[17] The third way is oligopolistic competition $n < \infty$, which can generate $B > 0$ even with CES demand and constant returns to scale. Proposition 2 nests all these cases and puts them on equal footing through the single statistic B. We shall also show that, naturally, as $n \to \infty$ the effect of oligopoly disappears and the slope B converges to the complementarity under monopolistic competition.

3.3 On Half Lives and Phillips Curves

Without sectoral heterogeneity, a useful statistic is the half-life of the price level (displayed in Figure 1) which by (7) is also equal to the half-life of the output gap. It is given by $h \cdot \log 2$ with

$$h = \frac{1}{\lambda(1 - B)}.$$

[16]When $B_s = 0$, equation (8) specializes to Proposition 1 in Carvalho (2006) about the cumulative output effect of a monetary shock with heterogeneous price stickiness λ_s.

[17]These are often discussed under the rubric of "real rigidities". A similar effect is obtained by allowing for input-output linkages, which we have abstracted from here.

To simplify, we henceforth refer to h simply as the half life, rather than $h \cdot \log 2$. In the basic New Keynesian model with monopolistic competition, CES demand, and constant returns to scale, $B = 0$ hence the half-life of the price level following a monetary shock is simply $1/\lambda$ and the cumulative output effect (i.e., (8) with $r = 0$) is $\frac{\delta}{\sigma\lambda}$.

The slope of the Phillips curve often serves as a measure of non-neutrality, which can be used to compare oligopoly to monopolistic competition. It turns out that a standard New Keynesian Phillips curve perfectly fits the reaction to a permanent money shock when $\omega\sigma = 1$ and abstracting from sectoral heterogeneity.

Proposition 3. *Suppose $\omega\sigma = 1$ and no heterogeneity across sectors, then to a first order*

$$\dot{\pi}(t) = \rho\pi(t) - \kappa\, mc(t) \qquad \pi(t) = \kappa \int e^{-\rho s} mc(t+s) ds$$

$$\kappa = \hat{\lambda}(\rho + \hat{\lambda}) \qquad \hat{\lambda} = \lambda(1 - B)$$

where $\pi(t) = \frac{\partial}{\partial t} \log P(t)$ and $mc(t) = \frac{1}{\sigma}\left(\log C(t) - \log \bar{C}\right)$ is the log-deviation of real marginal cost.

Up to constants of proportionality and for ρ small

$$\kappa \approx \frac{1}{h^2}.$$

This implies that for any percent increase in the half life h the slope of the Phillips curve is reduced by about twice this percentage. The metric κ is arguably more directly relevant for describing the tradeoff between inflation and output. The half-life h, however, is commonly reported in the menu-cost literature. In what follows we discuss both measures.

In Section 7 we investigate general preferences and develop a Phillips curve relationship that can be used for shocks of any kind, not just permanent money shocks. We show that the exact Phillips curve no longer generally takes the simple New Keynesian form above. However, despite these qualitative differences, we also find that the simple Phillips curve above continues to provide a quantitatively excellent approximation to the dynamics of inflation.

3.4 Sufficient Statistics for B: Markups and Elasticities

We now provide a key expression for the slope B in each sector, in terms of observable sufficient statistics. We focus on one sector and omit the s notation. For any n, we obtain B in terms of two steady state objects, the demand elasticity and the markup:

Proposition 4. *In a sector with $n \geq 2$ firms, the slope of the reaction function around the steady state satisfies*

$$B = \frac{1 + \frac{\rho}{\lambda}}{1 + \frac{1-(\mu-1)(\omega-1)}{(n-1)[(\epsilon-1)(\mu-1)-1]}} \tag{9}$$

where $\epsilon = -\frac{\partial \log d^i}{\partial \log p_i}(\bar{p})$ is the demand elasticity and $\mu = \frac{\bar{p}}{W/f'\left(f^{-1}\left(d^i(\bar{p})\right)\right)}$ is the steady state markup (i.e., price over marginal cost).

Proposition 4 shows how to locally infer unobserved steady state strategies from a small number of potentially observed statistics. Taking as given market concentration n and the demand elasticity ϵ, a higher steady state markup μ is associated with a higher slope B since $\epsilon \geq \omega$.[18] The advantage of Proposition 4 is that in order to infer the slope, we do not need to know the factors behind an observed markup, which is particularly useful as we show later that markups depend on many objects beyond ϵ. This is the sense in which μ is a sufficient statistic.[19]

The intuition behind this result is best seen in the other logical direction. If firm i deviates to a price above the equilibrium markup μ, a high B means that its rivals will react strongly and increase prices as well; this limits how much demand firm i loses from its deviation. Intuitively, this leads to a higher markup. Indeed, when $\omega = 1$ and $\rho \to 0$ to simplify,

$$\frac{\mu - 1}{\mu^{\text{Nash}} - 1} = 1 + \frac{1}{n-1} \cdot \frac{B}{1-B}$$

where $\mu^{\text{Nash}} = \frac{\epsilon}{\epsilon - 1}$. Thus, a high equilibrium markup must be a consequence of steep reaction functions. Note also that, according to this formula, what matters for B is the net markup relative to net Nash markup and n—the elasticity of demand plays no further role once we condition on this markup ratio.

Combining the results in this section, the response of the aggregate price level and output to a permanent monetary shock depends on three steady state statistics: markups, demand elasticities and market concentration. Armed with these sufficient

[18] The formula and its proof rely on strategic interactions between at least $n = 2$ firms. Recall that the case $n = 1$ recovers monopolistic competition: the demand elasticity is $\epsilon = \omega$ and the markup is $\mu = \frac{\omega}{\omega - 1}$.

[19] Equation (9) is a relation between endogenous objects. In particular, we cannot take limits in n or λ while holding μ fixed. In Section 4, we show how varying n and other parameters affects both sides of the equation.

statistics, it is unnecessary to solve the Markov equilibrium to analyze the effects of monetary shocks.

If these statistics are not observed, we need to solve the Markov equilibrium. Equation (9) gives a function $B = B(\mu, \omega, \epsilon, n, \lambda/\rho)$. In particular, taking ϵ and n as given, to solve for the markup μ, and thus B, we need another relationship between B and μ. The next section provides this relationship.

4 Market Concentration and other Comparative Statics

The sufficient statistic approach from the previous section answers the question: given the observed markups, concentration and demand elasticities, what is the aggregate price stickiness?

In this section we seek to answer how aggregate stickiness would *change* when market concentration and other parameters change.[20] To do so, we take a more structural approach: instead of using the observed equilibrium markup as a sufficient statistic, we need to solve for it. This allows us to perform counterfactual analyses, and investigate in depth which factors cause the oligopolistic model to depart from the standard monopolistic model.

4.1 Preliminaries: Method, Elasticity and Superelasticity

For a small number of firms, the Markov equilibrium can be easily solved numerically using standard methods, such as value function iteration. We employ this method, but since we want a solution for any n, the state space can become very large. Thus, we develop an alternative solution method, detailed in Appendix D.[21]

Our method selects Kimball preferences that generate an equilibrium that can be solved analytically locally around the steady state. Crucially, we can match any desired elasticities, superelasticities, and higher order elasticities of the demand function d^i, up to any desired order m. We employ $m = 2$ on the grounds that this is

[20] To simplify we analyze the effect of changes in market concentration, that have occurred slowly over time relative to business cycle frequencies, as a one-time comparative static experiment.

[21] The IO literature also acknowledges this challenge and employs approximate solution concepts such as "oblivious equilibria" (Weintraub, Benkard and Van Roy, 2008). Our method relates to the algorithm in Krusell, Kuruscu and Smith (2002) and Levintal (2018). Their solution approximates the Markov policy and value functions using polynomials of order m. Instead, we exhibit primitives such that locally the equilibrium is indeed polynomial of order m.

sufficient to flexibly match the first two derivatives, which are the only ones indirectly estimated in practice (using pass-through regressions, as we discuss later).[22] Equivalently we are matching the elasticity ϵ and the "superelasticity" Σ of demand at the steady state:[23]

$$\epsilon = -\frac{\partial \log d^i}{\partial \log p_i}(\bar{p}) \qquad \text{and} \qquad \Sigma = \frac{\partial^2 \log d^i}{\partial \log p_i^2}(\bar{p}) \,/\, \frac{\partial \log d^i}{\partial \log p_i}(\bar{p}).$$

Our method can be summed up as follows. The sufficient statistic formula (9) provides one equation

$$B = B(\mu, \omega, \epsilon, n, \lambda/\rho).$$

In Appendix E we derive an additional equation

$$\mu = \mu(B, \omega, \epsilon, \Sigma, n, \lambda/\rho).$$

We then solve these two equations in two unknowns (B, μ) for given $(\omega, \epsilon, \Sigma, n, \lambda/\rho)$.

Elasticity and Superelasticity. Our analysis allows for general sectoral demand (i.e. general aggregator H_s) taking only as inputs the local ϵ and Σ.

One way to flexibly parameterize demand locally is to use the preference construct from Kimball (1995). For our local analysis, this turns out to be without loss of generality. Define H_s implicitly as the unique solution for C_s to $\frac{1}{n_s}\sum_{i\in I_s}\phi_s\left(\frac{c_{i,s}}{C_s}\right) = 1$ for some increasing, concave function ϕ_s with $\phi_s(1) = 1$. If ϕ_s is a power function, we obtain the standard CES aggregator across firms. Letting $\Phi_s(x) = -\frac{\phi_s'(x)}{x\phi_s''(x)}$ it is standard to define

$$\eta_s = \Phi_s(1) \qquad \text{and} \qquad \theta_s = -\Phi_s'(1),$$

so that η_s is the (local) elasticity of substitution and CES corresponds to $\theta_s = 0$.[24]

[22] Another interpretation (as in Krusell et al. 2002) is that the infinite sequence of elasticities is given, for instance if preferences are exactly CES, and our method approximates preferences to match the same elasticities up to order $m < \infty$. We show in Appendix D in the context of a duopoly with CES preferences that using a higher order $m = 3$ yields very close results to $m = 2$ and that the results are also very close to the solution obtained using value function iteration.

[23] Given homotheticity these own-price elasticities also pin down cross-price elasticities, as proved in Appendix C.

[24] Klenow and Willis (2016) propose a functional form defined so that $\theta = -\Phi_s'(x)$ holds globally for all x, but this is not needed for our local analysis.

Then we have (dropping the s subscript):

$$\epsilon = \left(1 - \frac{1}{n}\right)\eta + \frac{1}{n}\omega, \tag{10}$$

$$\Sigma = \frac{n-1}{n} \cdot \frac{(n-2)\theta\eta + \eta^2 - (1+\omega)\,\eta + \omega}{(n-1)\,\eta + \omega}. \tag{11}$$

Equation (11) shows that as n goes to infinity, Σ converges to θ hence the superelasticity under monopolistic competition is non-zero only if preferences are not CES. But with finite n, Σ generally differs from zero even with CES preferences $\theta = 0$. As equation (10) shows, elasticities depend on market shares and, thus, n as is well known in the CES case studied by Atkeson and Burstein (2008). Our expressions above generalize this to any Kimball aggregator and derive new expressions for the superelasticity (Appendix C provides cross-elasticities).

Note that for the special case $n = 2$, equation (11) reveals that θ plays no role and the superelasticity Σ is restricted to being that of the CES case $\theta = 0$. This restriction with $n = 2$ is not special to Kimball preferences and reflects a more general property of symmetric homothetic demand systems that we prove in Appendix C.

4.2 Market Concentration

Our main counterfactual exercise is to study how changes in market concentration (the number of firms n in a sector) affect the transmission of monetary policy. If we knew how the sufficient statistics entering (9) changed with concentration, it would not be necessary to solve the model further. Absent this information, we need to make assumptions on how these statistics depend on n, for instance by taking a stand on what parameters to keep fixed when changing n. We start by holding "preferences" fixed, and exogenously shifting the number of firms; the implicit assumption is that each firm offers a fraction $1/n$ of varieties. We then explore an alternative, calibrating these preferences to the available evidence on pass-through from costs to prices.

Exogenous Changes in Number of Firms. We first hold preferences, embedded in the Kimball aggregator ϕ_s, fixed when changing n. Therefore the parameters η and θ are fixed, but the elasticities ϵ and Σ will change according to (10)-(11). One interpretation is that the set of varieties demanded by consumers is unchanged, but concentration increases due to mergers and acquisitions: firms expand and get to set prices for more varieties. We assume constant returns to scale in production, $f(\ell) = \ell$

Table 1: Parameter values.

Parameter	Description	Value
ρ	Annual discount rate	0.05
λ	Price changes per year	1
ω	Cross-sector elasticity	1
η	Within-sector elasticity	10

to focus on complementarities coming from the demand side. The remaining parameters are described in Table 1.

When we restrict preferences to be CES ($\theta = 0$), higher market concentration in the sense of lower n increases monetary non-neutrality. The maximal half-life, attained under duopoly $n = 2$, is approximately 40% higher than under monopolistic competition. This is a substantial effect: from Proposition 3 this corresponds to dividing the slope of the Phillips curve by $1.4^2 \approx 2$. The amplification from oligopoly decreases rapidly with n, however: with $n = 10$ firms the half-life is only 10% higher than under monopolistic competition. Allowing for an arbitrary number of firms n is thus crucial to understand the effect of a realistic increase in concentration.

Once we consider more general preferences than CES, Figure 2 shows that for high values of θ that generate strong demand complementarities, and thus large effects of monetary policy even under monopolistic competition, oligopoly can *dampen* monetary policy. For instance, for $\theta = 15$, going from monopolistic competition to duopoly decreases the half-life by 20%. The reason is that a high superelasticity Σ increases B, and for low θ (e.g., $\theta = 0$) Σ decreases with n, but the opposite holds for high θ. We discuss this further in Section 5.

In principle, this dampening effect of oligopoly can be arbitrarily large: the half-life with high n increases without bounds with θ, but interestingly, in the special case of a duopoly $n = 2$ the half-life is always the same as under CES. This can be seen from equation (11), where θ is irrelevant for Σ when $n = 2$.

There is therefore no guarantee that concentration increases non-neutrality: the direction of the effect depends on finer properties of demand, e.g., how ϵ and Σ depend on n. Next, we offer a different calibration strategy that infers these properties from available pass-through estimates.

A Calibration Based on Pass-Through. Previously, we fixed the "preference parameters" η, θ when changing the number of firms n. We now provide an alternative that

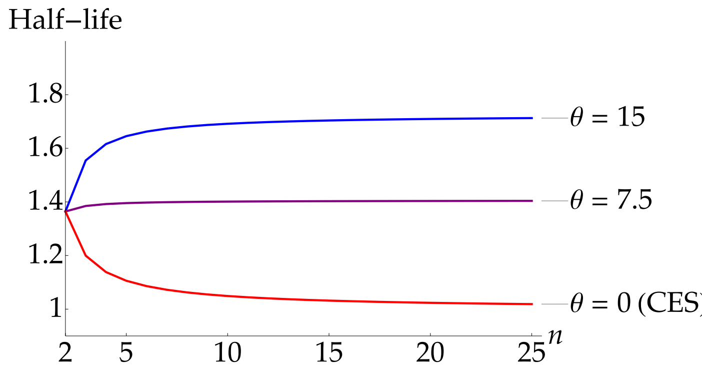

Figure 2: Half-life as a function of n for different values of θ.

does not hold preferences ϕ_s fixed as we change the number of firms. We have seen that the shape of demand is crucial to understand how market structure impacts the transmission of aggregate monetary shocks. As Atkeson and Burstein (2008) emphasized in a static setting, market structure also affects the pass-through of firm-level cost shocks, hereafter simply "pass-through". We now argue that calibrating the model to match the empirical relation between market share and pass-through implies that concentration significantly amplifies monetary non-neutrality, even more than under CES.[25]

Amiti et al. (2019) estimate pass-through regressions

$$\Delta \log p_{it} = \hat{\alpha} \Delta \log mc_{it} + \hat{B} \frac{\sum_{j \neq i} \Delta \log p_{jt}}{n-1} + u_{it} \tag{12}$$

separately for small and large firms, and find considerable heterogeneity in pass-through. Small firms behave as under a CES monopolistic competition benchmark, passing through own marginal cost shocks fully (and thus maintaining a constant markup) while not reacting to competitors' price changes orthogonal to their own cost. Large firms exhibit substantial strategic complementarities: they only pass through around half of their own cost shocks, thus letting their markup decline to absorb the other half.[26] Amiti et al. (2019) show that $\hat{\alpha}$ as a function of market share s

[25] This alternative calibration strategy can be interpreted in two ways. First, we can assume that if a sector becomes concentrated due to its firms growing larger, then these firms' idiosyncratic cost pass-through becomes similar to the pass-through of large firms currently observed in other concentrated sectors. Second, we can assume that aggregate concentration increases due to already concentrated sectors becoming larger in a way that preserves within-sector demand (and thus pass-through).

[26] Other papers such as Berman, Martin and Mayer (2012) and Chatterjee, Dix-Carneiro and Vichyanond (2013) also find lower pass-through for larger firms.

is well approximated by

$$\hat{\alpha} \approx \frac{1}{1 + \frac{(\eta-1)(1-s)s(\eta-\omega)}{\omega(\eta-1)-s(\eta-\omega)}} \tag{13}$$

with $\eta = 10$ and $\omega = 1$ (equation 12 and calibration p. 2398 in their paper).[27] Empirical variation in s captures both differences in concentration across sectors and heterogeneity across firms within sectors. We will associate the share under our symmetric model to concentration: $s = 1/n$.[28]

Appendix F details how to calibrate our model to pass-through estimates. We sketch our approach here. First we generalize the reaction function (3) to allow for shocks to marginal costs mc_j. When firm i adjusts its price it sets

$$\tilde{p}_i = \alpha \widetilde{mc}_i + B \frac{\sum_{j \neq i} \tilde{p}_j}{n-1} + \gamma \sum_{j \neq i} \widetilde{mc}_j \tag{14}$$

where tildes denote log-deviations from steady state values.[29] Equation (14) describes the reaction function, while (12) is a relation between equilibrium changes. The following result, proved in Appendix F, describes the mapping from the model parameters α, B in (14) to empirical estimates $\hat{\alpha}, \hat{B}$ from pass-through regressions (12):

Proposition 5. *There exist unique scalars*

$$\hat{\alpha} = \frac{n\alpha + B - 1}{\alpha + B + n - 2} \qquad\qquad \hat{B} = \frac{(n-1)(1-\alpha)}{\alpha + B + n - 2} \tag{15}$$

such that for any vector of cost shocks $[\Delta mc_i]'_{i=1..n}$, *equation* (12) *holds with* $u_i = 0$.

Therefore in a sector with n firms we set as target $\hat{\alpha}$ from (13) with $s = 1/n$. Then, fixing other parameters (i.e., η, λ, ρ), for each (n, θ) we compute α and B and solve for θ_n that satisfies (15).

Remark 2. The mapping $(\alpha, B) \mapsto (\hat{\alpha}, \hat{B})$ given by Proposition 5 cannot be inverted to obtain directly α, B as functions of empirical estimates $\hat{\alpha}, \hat{B}$. That is, for any $\hat{\alpha}, \hat{B}$ such

[27] As they show in Table 7 and discuss in Appendix D, this calibration implies that a firm with a market share of 12.5% has a cost pass-through of around 0.5, which matches their empirical pass-through estimates in Table 3 for large firms (defined by employment or sales share).

[28] Pass-through as a function of market share is essentially the same, whether variation in market share comes from varying the number n of symmetric firms, or from within-sector heterogeneity among a fixed number of firms. This equivalence is exact in a static model, and holds approximately in our dynamic model as shown in Figure A.6. In Section 6 we show that the equivalence also holds for the response to aggregate monetary shocks.

[29] The coefficients α, B, γ can be computed as before using envelope conditions applied to a generalization of the Bellman equation (equation (A.27)).

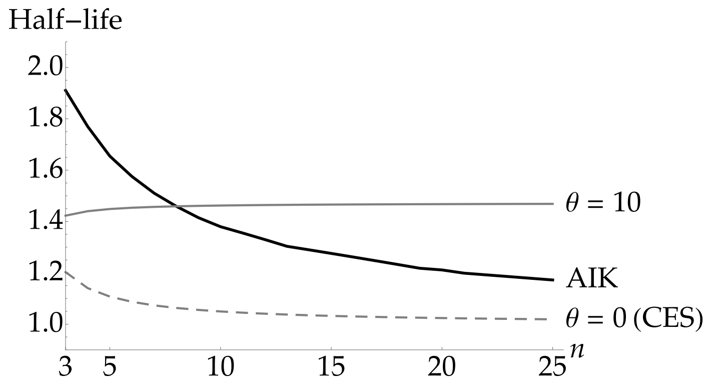

Figure 3: Half-life as a function of the number of firms n. AIK: variable θ_n to match pass-through estimates from Amiti et al. (2019).

that $\hat{\alpha} + \hat{B} = 1$ (a condition that $\hat{\alpha}$ and $\hat{B}$ in (15) must satisfy), the system (15) does not identify α, B.

This non-invertibility is the reason why we need to solve the full dynamic model to map pass-through estimates to the aggregate effects of monetary policy, as the model provides additional restrictions on α and B. Note that static oligopoly (i.e., the limit $\lambda/\rho \to 0$) also yields an additional restriction $\alpha + B = 1$ or equivalently $\gamma = 0$ on the coefficients in (14): in a static model firm i does not respond to a rival's cost shock mc_j directly because competitors' prices are sufficient statistics for payoffs (and thus competitors' costs are irrelevant conditional on their prices). Under that additional restriction we can recover uniquely $\alpha = \hat{\alpha}$, $B = \hat{B}$. The same holds for $n \to \infty$. However, under dynamic oligopoly, B differs from $\hat{B}$ in general. Indeed, quantitatively B is close to $1 - \sqrt{1 - \hat{B}}$, which can be much lower than $\hat{B}$.

Results. We fix η at 10, a common benchmark in the literature since Atkeson and Burstein (2008). We hold η fixed to focus the discussion on how pass-through and hence the superelasticity Σ changes with concentration.[30]

Figure 3 shows our results. "AIK" is our calibration with a variable parameter θ_n as explained above. Concentration amplifies stickiness substantially, much more

[30]Ideally, one would obtain non-parametric estimates of both $\epsilon(n)$ and $\Sigma(n)$ from matching jointly the relation of markups and pass-through with market shares, but at the time of writing there was no direct counterpart to Amiti et al. (2019) for markups. In recent work, Burstein, Carvalho and Grassi (2020) examine the relation between market shares and markups and find in French data that a linear regression of the inverse markup against the sectoral HHI yields a coefficient of -0.44. In our dynamic model, the corresponding coefficient is -0.24 and gets closer to their estimate than a CES model, which would yield -0.15. Allowing η to increase with n instead of fixing $\eta = 10$ would improve the fit further.

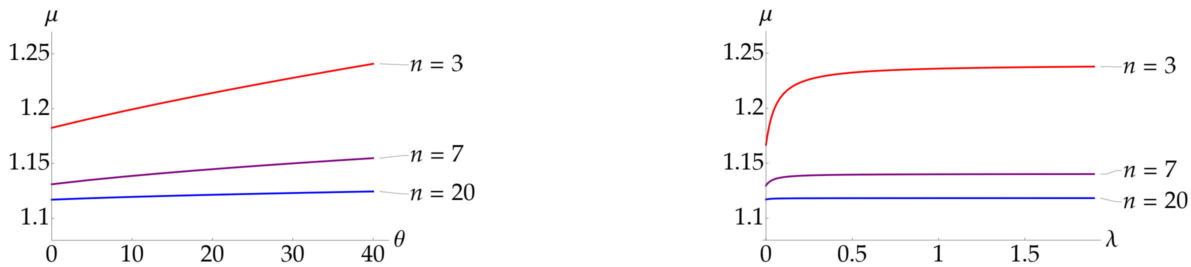

Figure 4: Steady state markup μ as a function of θ (left panel) and λ (right panel).

than under CES. When going from monopolistic competition to $n = 3$ firms, the half-life doubles; equivalently, the slope of the Phillips curve is divided by four. For comparison we include calibrations that hold θ fixed.

Unlike under CES, oligopoly matters even for realistic levels of concentration. Consider a rise in national concentration from an average Herfindahl index $1/n$ of 0.05 to 0.1, reflecting the trends observed since 1990 by, e.g., Gutiérrez and Philippon (2017). Under the "AIK" calibration, the half-life is 20% higher than under monopolistic competition when $n = 20$, and 40% higher when $n = 10$. This 15% increase in non-neutrality in terms of half lives is equivalent to a one-third reduction in the slope of the Phillips curve.[31]

4.3 Other Comparative Statics: the Determinants of Markups

Our sufficient statistics formula (9) highlights the role of markups. Other parameters such as preferences and price stickiness also affect markups and monetary policy transmission, holding concentration (i.e., n) fixed. Here we summarize the main findings and refer interested readers to Appendix G for more discussion and numerical explorations.

As under monopolistic competition, a lower elasticity of substitution η increases the markup. But the effect on stickiness is ambiguous (and depends on θ), because a lower η also decreases the demand elasticity ϵ hence multiple terms in (9) are changing.

We argued that under dynamic oligopoly, markups are not fully determined by

[31] Rossi-Hansberg et al. (2020) show a decline in *local* concentration, in particular in the retail sector. An interesting open question is then which level of aggregation (what we call "sectors" s) is most relevant for consumer price inflation. The answer depends in part on the prevalence of "uniform pricing" policies (DellaVigna and Gentzkow, 2019) and how concentration varies at different points of supply chains (e.g., pass-through is lower for wholesale prices than retail prices).

demand elasticities. Figure 4 illustrates this point. A higher superelasticity parameter θ increases the markup. Since this leaves ϵ unchanged, this experiment is the most transparent application of our sufficient statistic result: B and thus the half-life increase. Finally, markups increase with the frequency of price changes λ and decrease with discount rates ρ: more patient firms can sustain a higher markup, as in the literature on collusion.

5 Inspecting the Mechanism: Strategic vs. Naive Firms

The presence of a finite number of firms has two distinct effects on competition and pricing incentives: "feedback effects" capture the fact that each firm cares about its rivals' current and future prices when setting its price; "strategic effects" capture instead the fact that each firm realizes its current pricing decision can affect how its rivals will set their prices in the future. Feedback effects are present even under monopolistic competition ($n = \infty$) with Kimball demand or decreasing returns, however strategic effects are not.

We disentangle the two effects through the lens of a "naive" model, in which firms are naive in the following sense: when resetting their price, they form correct expectations about the stochastic process governing their competitors' future prices, but incorrectly assume that their own price-setting will have no effect on those competitors' future prices. The naive model captures all the feedback effects, while suppressing strategic effects. We have the following equivalence result.

Proposition 6. *The time paths for aggregates in the naive model with finite firms $n < \infty$ and parameters (η, θ) are identical to those of an economy with monopolistic competition $n' = \infty$ and modified Kimball preferences (η', θ') set to match the demand elasticity ϵ and superelasticity Σ of the model with n firms, using* (10)-(11).

Therefore the naive model provides a behavioral foundation for the notion of a "properly recalibrated" monopolistic economy.[32] We compute the half-life of the price level h in the strategic model and h^{Naive} in the naive model, and then define

[32] We could also compare oligopoly to a model with monopolistic competition that matches the same own-cost pass-through (the naive model matches elasticities, not pass-through). While such an economy lacks the behavioral interpretation of the naive model, it is a natural benchmark when thinking about recalibration. Quantitatively, this alternative is almost identical to the naive model, because matching elasticities is very close (and equivalent when $n \to \infty$ or $\lambda \to 0$) to matching pass-through.

strategic effects as follows:

$$h = \frac{1}{\lambda} \times \underbrace{\frac{h^{\text{Naive}}}{1/\lambda}}_{\text{feedback effect}} \times \underbrace{\frac{h}{h^{\text{Naive}}}}_{\text{strategic effect}}.$$

As n goes to infinity, h/h^{Naive} goes to 1 and the strategic effect disappears; what is left is the standard feedback effect that can stem from non-CES demand or decreasing returns to scale.

The Naive Reaction Slope. The solution of the naive model is in Appendix H. The key difference with the strategic equilibrium is that here, when setting a price firm i treats the evolution of rivals' prices as exogenous to its choice p_i. The steady state price of the naive model is the static Bertrand-Nash price p^{Nash}, that solves $\Pi_i^i \left(p^{\text{Nash}}\right) = 0$. To first order, each resetting firm i sets

$$\log p_{i,s}(t) = \log p_s^{\text{Nash}} + B_s^{\text{Naive}} \frac{\sum_{j \neq i} \left(\log p_{j,s}(t) - \log p_s^{\text{Nash}}\right)}{n_s - 1}.$$

Following the same steps as for Proposition 2, the price level in the naive model evolves according to (6) with B_s^{Naive} instead of B_s.

Denote B_s^{Nash} the slope of the *static* best response of a firm to a simultaneous price change by all its competitors[33]

$$B_s^{\text{Nash}} = \frac{(n_s - 1)\, \Pi_{ij}^{i,s}\left(p_s^{\text{Nash}}\right)}{-\Pi_{ii}^{i,s}\left(p_s^{\text{Nash}}\right)}.$$

The following result shows that the slope B_s^{Naive} is a simple increasing function of B_s^{Nash}.

Proposition 7. *Let* $\varphi(x, y) = 1 + 1/(2y) - \sqrt{(1 + 1/(2y))^2 - (1 + 1/y)\, x}$. *Then*

$$B_s^{Naive} = \varphi\left(B_s^{Nash}, \frac{\lambda_s}{\rho}\right).$$

[33] We can reexpress B^{Nash} in terms of the demand elasticities as $B^{\text{Nash}} = \frac{\Gamma}{1+\Gamma}$ where $\Gamma = \frac{\Sigma}{\epsilon - 1}$ is also known as the markup elasticity (Gopinath and Itskhoki, 2010) or responsiveness (Berger and Vavra, 2019). In the limit of monopolistic competition, $B^{\text{Nash}} \to \frac{\theta}{\theta + \eta - 1}$.

is increasing in B_s^{Nash} and decreasing in λ_s/ρ and

$$1 - \sqrt{1 - B_s^{Nash}} \leq B_s^{Naive} \leq B_s^{Nash}. \tag{16}$$

Note that if $B_s^{\text{Nash}} = 0$ (as under monopolistic competition $n = \infty$ with CES demand) then $B_s^{\text{Naive}} = 0$. With a finite number of firms, $B_s^{\text{Nash}} > 0$ and thus $B_s^{\text{Naive}} > 0$ even with CES demand.

The naive price-setting strategy is not completely naive: it is still forward-looking and differs from the static best-response, indeed $B_s^{\text{Naive}} < B_s^{\text{Nash}}$ for $\lambda_s/\rho > 0$. The lower and upper bounds in (16) come from the limits $\lambda_s/\rho \to \infty$ and $\lambda_s/\rho \to 0$, respectively.

In practice, since ρ is small relative to λ, B_s^{Naive} is closer to its lower bound $1 - \sqrt{1 - B_s^{\text{Nash}}}$. Numerically, we also find that B_s is close to B_s^{Naive}, as we show below.

Strategic, Naive, and Static Models. How well does the naive model approximate the strategic model? We start with some familiar limiting benchmarks. Intuitively, the strategic effects vanish if λ/ρ is small or n is large. Formally,

$$\lim_{\lambda/\rho \to 0} B_s = \lim_{\lambda/\rho \to 0} B_s^{\text{Naive}} = B_s^{\text{Nash}}.$$

Also for $\lambda/\rho > 0$,

$$\lim_{n \to \infty} B_s = \lim_{n \to \infty} B_s^{\text{Naive}} = \varphi\left(\lim_{n \to \infty} B_s^{\text{Nash}}, \frac{\lambda_s}{\rho}\right) = \varphi\left(\frac{\theta_s}{\theta_s + \eta_s - 1}, \frac{\lambda_s}{\rho}\right).$$

This latter limit for B_s together with Proposition 2 summarizes the standard monopolistic competitive model, with CES ($\theta = 0$) or Kimball preferences ($\theta > 0$).

Away from these two limit cases, we need to evaluate numerically the distance between the naive and strategic models. Figure 5 displays the strategic effect h/h^{Naive} as n varies from 3 to 25. We find that quantitatively, strategic effects do not explain much of the aggregate price stickiness under oligopoly. Strategic effects are considerably stronger in the "AIK" calibration, which also features stronger feedback effects. This interaction is intuitive: the reason a firm acts strategically is that its price will have a feedback effect on competitors when they get to reset their prices. Yet in all specifications, strategic effects are small: the half-life is always less than 3% higher than the naive half-life. Consistent with their definition, strategic effects vanish as n grows and the economy approaches monopolistic competition: they fall below 1%

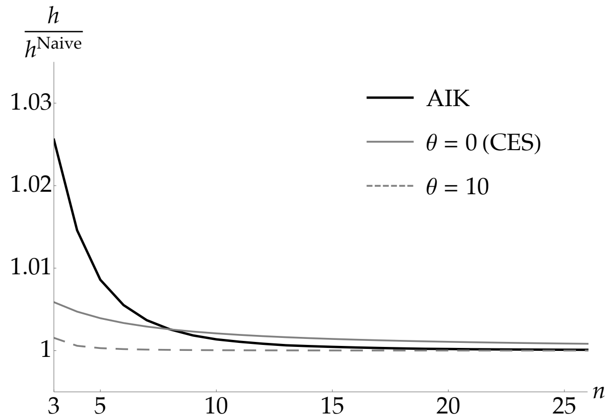

Figure 5: Strategic effect h/h^{Naive} as a function of n. AIK: variable θ_n to match pass-through estimates from Amiti et al. (2019).

when n exceeds 5.

Comparative Statics in the Naive Model. The effect of oligopoly on monetary policy transmission is transparent in the naive model because changes in n, θ, η affect B^{Naive} only through B^{Nash}. Going back to the findings from Section 4.3, the naive model helps understand, for instance, when concentration amplifies or dampens aggregate price stickiness. Figure 2 shows that the effect of n in the strategic model depends on the value of θ; in the naive model we can prove this property analytically and understand better the underlying intuition.

Holding fixed θ and η, B^{Nash} is decreasing in n if and only if

$$\theta < \frac{(\eta - 1)^2}{\eta + 1}. \tag{17}$$

Therefore, with CES preferences $\theta = 0$, concentration amplifies aggregate price stickiness, while θ above around 7.5 (for $\eta = 10$) implies that concentration dampens stickiness, which matches closely Figure 2.

Intuitively, there are two opposite forces. Recall that with CES preferences, the demand elasticity of a firm i with market share s_i is simply $\epsilon = \eta (1 - s_i) + \omega s_i$; a higher price p_i decreases s_i hence increases ϵ; in other words, $\Sigma > 0$. A smaller number of firms n strengthens this source of superelasticity because the impact of a given price change on the market share is larger. The opposite force arises only with non-CES preferences: $\theta > 0$ increases a firm's incentives to set a price close to the average price of other varieties. A smaller n weakens this source of complementarity, because each remaining firm controls prices for a larger share of varieties and thus becomes less sensitive to other firms' prices. Condition (17) characterizes precisely when the first force dominates.[34]

6 Heterogeneity Across and Within Sectors

We now explore the role of heterogeneity across and within sectors. Across sectors we focus on heterogeneity in the frequency of price changes λ_s, and discuss when it interacts with oligopolistic competition to further amplify monetary non-neutrality. Within sectors we allow firms to differ in their productivity or the demand they face, which results in heterogeneous firm size. Our main finding is that the baseline model with symmetric firms is a very good approximation of a richer model with heterogeneous firms, once we reinterpret the number of firms n in a sector as an "effective number of firms" equal to the inverse Herfindahl index.

6.1 Heterogeneous Price Stickiness across Sectors

The effect of the frequency of price changes on markups and therefore reaction functions is magnified in the presence of sector heterogeneity in λ. Several papers have documented a link between frequency of price changes and market structure.[35] Models with menu costs provide a microfoundation for the effect of concentration on price flexibility. Although our Calvo framework does not endogenize the frequency, interesting insights still arise from taking observed correlations as given, by letting λ_s comove with n_s. From (8), the cumulative output effect for a monetary shock of size δ is:

$$\frac{\delta}{\sigma} \times \left\{ \mathbf{E}\left[\frac{1}{\lambda_s}\right] \mathbf{E}\left[\frac{1}{1-B_s}\right] + \mathbf{Cov}\left(\frac{1}{\lambda_s}, \frac{1}{1-B_s}\right) \right\} \tag{18}$$

where $\mathbf{E}\left[x_s\right] = \int_s \zeta_s x_s ds$ denotes the average of a variable x across sectors. In Section 4.2 we saw that $\frac{1}{1-B_s}$ is higher in more concentrated sectors. If these sectors are also characterized by a lower frequency λ_s, then the covariance term is positive, which

[34] Similarly, holding n fixed, B^{Nash} decreases with the elasticity of substitution η (and thus the observed markup) if and only if $\theta < \frac{n}{n-2} \times \frac{(\eta-1)^2}{1+(n-1)\eta^2}$, which explains why, in Figure A.3, the half-life decreases with the markup μ under CES ($\theta = 0$) but not when θ is high enough.

[35] See, e.g., Carlton (1986). Most recently, Mongey (2018) shows that price changes are less frequent in more concentrated wholesale markets. Given that market shares and pass-through are negatively correlated, this fact is also consistent with Gopinath and Itskhoki (2010), who show price changes are less frequent for goods with a lower long-run exchange rate pass-through.

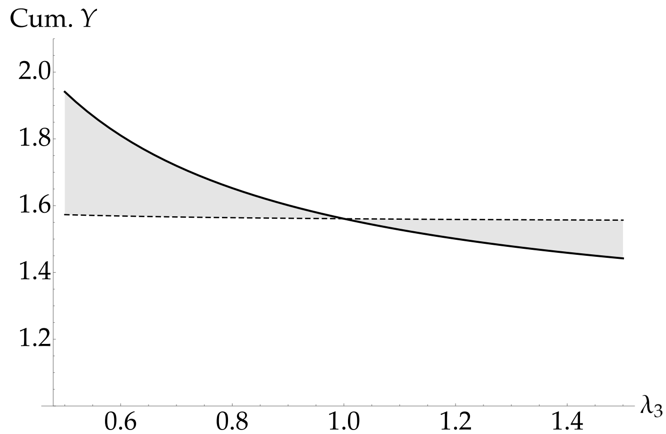

Figure 6: Effects of heterogeneous frequency across sectors. The black line shows the cumulative output effect and the gray area shows the covariance term in (18).

Note: Example with two sectors, one with $n = 3$ firms and one with $n = 20$ firms, with "AIK" calibration. λ_3 is the frequency in the sector with 3 firms, and we set the frequency λ_{20} in the other sector to keep the average duration $\frac{1}{2}\left(\frac{1}{\lambda_3} + \frac{1}{\lambda_{20}}\right)$ fixed at 1.

increases aggregate non-neutrality. This channel differs from the role of heterogeneity in, e.g., Carvalho (2006) under Calvo pricing or Nakamura and Steinsson (2010) under menu costs. Oligopoly is a very natural reason to have heterogeneity in B_s, but note that even under monopolistic competition $n_s \to \infty$, heterogeneous B_s could arise from differences in Kimball demand or the degree of decreasing returns to scale across sectors.

Figure 6 shows the magnitude of this channel in an example with one concentrated sector ($n = 3$) and one competitive sector ($n = 20$) under the "AIK" calibration. If the two sectors have the same price duration of 12 months, then the cumulative output effect is 56% higher than in a standard New Keynesian model without complementarities. If, instead, the durations are 18 months in the concentrated sector and 6 months in the competitive sector, the average duration is unchanged at 12 months, but tthe cumulative output effect is now 75% higher than in the standard model.

6.2 Heterogeneous Firm Productivity within Sectors

We now extend our baseline model to allow for permanent heterogeneity between firms *within* sectors, in terms of tastes and productivity. Focusing on one sector s, suppose that firms differ in their productivity z_i while consumers have different tastes

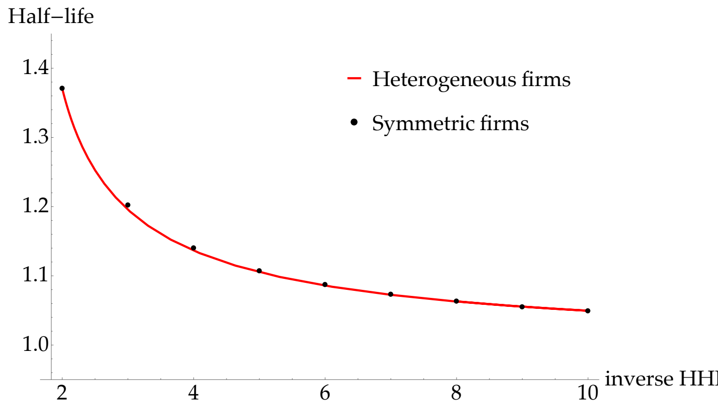

Figure 7: Heterogeneous vs. symmetric firms.

Note: Red line: Half-life with 10 heterogeneous firms ($n_a = 2, n_b = 8$) when varying relative productivity. Black dots: Half-life with $n = 2, 3, \ldots, 10$ symmetric firms. All cases feature nested CES preferences with $\eta = 10$ and $\omega = 1$.

captured by multiplicative demand shifters ξ_i.[36]

We solve the heterogeneous firms in Appendix E.2 using the same method as in the symmetric case. We consider a simple form of heterogeneity: in each sector, there are n_a firms of type a and n_b firms of type b, with the convention $n_a \leq n_b$. The two types of firms allow us to capture the case of small and large firms in a sector.[37]

The takeaway is that our baseline model with n symmetric firms is a good approximation to a model with heterogeneous firms, once we reinterpret n as the inverse HHI of the heterogeneous firms model. The red line in Figure 7 shows the half-life as a function of the inverse HHI of type-a firms as we vary continuously their relative productivity. Each black dot represents the half-life of a model with $n = n_a, n_a + 1, \ldots, n_a + n_b$ symmetric firms. The black dots remain extremely close to the red line. The same conclusion holds for different choices of n_a, n_b. Therefore, even though we assume symmetry in our baseline model for simplicity, our results extend to more realistic firm distributions once reinterpreted properly.

[36] Sectoral consumption C_s solves $\frac{1}{n} \sum_{i \in I_s} \phi\left(\frac{\xi_i c_i}{C_s}\right) = 1$.

[37] With heterogeneity we need to solve for an asymmetric steady state price vector and a *matrix* of strategy slopes $\beta^i_j = \frac{\partial g^i}{\partial p_j}$. We only assume two types for computational simplicity. The same solution method works with any $k \leq n$ types of firms, which would require solving for k prices and k^2 slopes.

7 The Oligopolistic Phillips Curve

We focused so far on the dynamics following a permanent money supply shock, under parametric restrictions that allowed to go from the stationary industry equilibrium to general equilibrium. In this section we generalize our analysis considerably in terms of preferences and shocks.

Qualitatively, our main result is an oligopolistic Phillips curve that features extra terms relative to the basic monopolistic competitive model, but is still tractable enough for computation. We then use a three-equation oligopolistic New Keynesian model to study the response to a variety of shocks. Quantitatively, our main finding is that although this extended model can display larger strategic effects than in the previous experiments, the standard monopolistic Phillips curve obtained with naive firms provides a very good approximation to inflation dynamics.

7.1 The Phillips Curve

We now relax the restrictions on preferences and the type of shock from Section 3. In Appendix I we derive the following Phillips curve, first expressed in integral form.

Proposition 8. *There exists $q \leq 7$ and a $q \times q$ matrix $\mathbf{A}$ described in Appendix I, that depends on steady state demand elasticities, with eigenvalues $\{\nu_j\}_{j=1}^{q}$, such that*

$$\pi(t) = \int_0^\infty \gamma^{mc}(s)\, mc\,(t+s)\, ds + \int_0^\infty \gamma^{c}(s)\, c\,(t+s)\, ds + \int_0^\infty \gamma^{R}(s)\, (R\,(t+s) - \rho)\, ds \tag{19}$$

where $R(t)$ is the nominal interest rate, $mc(t)$ and $c(t)$ are the log-deviations of the real marginal cost and consumption, respectively, and for each variable $x \in \{mc, c, R\}$, $\gamma^x(s)$ is a linear combination of $\{e^{-\nu_j s}\}_{j=1}^{q}$.

The Phillips curve provides a general mapping from the paths of future marginal costs, aggregate consumption and interest rates to current inflation. For the shocks we considered so far, equilibrium marginal costs and consumptions are in fixed proportions and $R(t) = \rho$ is fixed, but (19) is more general, allowing to incorporate, e.g., interest rate and productivity shocks. The content of Proposition 8 is not that there exist generic coefficients $\gamma^x(s)$ satisfying (19), but that they have a very specific and solvable structure tied to the oligopoly game and demand elasticities.

Under monopolistic competition, (19) simplifies drastically to

$$\pi(t) = \kappa \int_0^\infty e^{-\rho s} mc\,(t+s)\, ds \tag{20}$$

or equivalently $\dot{\pi} = \rho\pi - \kappa\, mc$ for some coefficient κ, that is, the slope of the Phillips curve. In particular, when firms are naive, the Phillips curve is simply

$$\dot{\pi} = \rho\pi - \kappa^{\text{Naive}}(n)\, mc \tag{21}$$

with $\kappa^{\text{Naive}}(n) = (1 - B^{\text{Nash}}(n))\lambda(\lambda + \rho)$.

Under strategic oligopoly, inflation is also primarily determined by a weighted average of future marginal costs captured by the first term in (19), but oligopoly is not isomorphic to a lower λ due to two qualitative differences. First, the multiplicity of eigenvalues induces higher-order terms in the dynamical system that alter the shape of $\gamma^{mc}(s)$. Second, inflation depends on other variables than future marginal costs through the other terms in (19). In the standard New Keynesian model, real marginal costs capture all the forces that influence price setting. Here, consumption and interest rates have an independent first-order effect because they alter the strategic complementarities between firms. For instance the coefficients $\gamma^{c}(s)$ capture the Rotemberg and Saloner (1986)-like aggregate demand effects absent when $\omega\sigma = 1$.[38]

We can transform (19) into a high-order scalar ordinary differential equation for inflation. Focusing on an example with few firms, $n = 3$, to highlight the differences with monopolistic competition, the integral Phillips curve (19) is approximately equivalent to[39]

$$\dot{\pi} = 0.07\pi - 0.27 mc + 1.33\ddot{\pi} + \underbrace{0.44\dot{mc} + 0.03\,(R-\rho)}_{=u} \tag{22}$$

under the AIK calibration. Turning to the naive Phillips curve (21), we have

$$\kappa^{\text{Naive}}(3) = 0.25,$$
$$\kappa^{\text{Naive}}(\infty) = 1.05.$$

Going from $n = \infty$ to $n = 3$ reduces κ^{Naive} by a factor of *four*; in this sense the

[38]Figure A.7 illustrates the coefficients $\gamma^{mc}(s), \gamma^{c}(s), \gamma^{R}(s)$ for different n.

[39]In general there are $q = 7$ eigenvalues, but q can be reduced to 3 under a simplifying condition (A.36) given in Appendix I, which we assume to ease the exposition of (22); inflation dynamics are almost unchanged when we use $q = 7$, as we do in all the figures.

amplification from oligopoly appears very large. This result is consistent with Figure 3 in which the half-life h doubles when going from $n = \infty$ to $n = 3$, given the relation $\kappa \propto 1/h^2$ from Proposition 3.[40]

Relative to the naive Phillips curve, the strategic Phillips curve (22) features a similarly low coefficient on mc, but also (i) more discounting, (ii) a higher-order term $1.33\ddot{\pi}$, and (iii) a term u that resembles an endogenous "cost-push" shock. Although the Phillips curve (22) is qualitatively different, in practice we show next that the naive equilibrium continues to provide a great fit. By implication, the standard NK Phillips curve with slope κ^{Naive} (21) provides a very good fit to the dynamics of inflation in response to various shocks.

7.2 Three-Equation Model

We can now analyze a three-equation New Keynesian model that combines the Phillips curve with an Euler equation

$$\dot{c} = \sigma^{-1}\left(R - \pi - \rho - \varepsilon^r\right),$$

and a monetary policy interest rate rule

$$R = \max\left\{0, \rho + \phi_\pi \pi + \varepsilon^m\right\},$$

where $\varepsilon^r(t)$ and $\varepsilon^m(t)$ are real and monetary shocks, respectively. The rest of the model is standard. Wages are flexible, technology is linear in labor $Y = \ell$ and households have preferences $\frac{C^{1-\sigma}}{1-\sigma} - \frac{\ell^{1+\psi}}{1+\psi}$, hence $mc = (\psi + \sigma)\, c$. We set standard values $\sigma^{-1} = 0.5$ for the elasticity of intertemporal substitution, $\psi^{-1} = 0.5$ for the Frisch elasticity of labor supply, and $\phi_\pi = 1.5$ for the Taylor rule coefficient on inflation.

Date-0 Monetary Policy Shocks. We first consider unanticipated date-0 interest rate shocks that decay geometrically, $\varepsilon^m(t) = \varepsilon_0^m e^{-\xi t}$, while shutting other shocks, $\varepsilon^r = 0$, so that the zero lower bound remains slack. The solution is detailed in Appendix I.1. Under both monopolistic and oligopolistic competition, all the equilibrium variables $x \in \{c, \pi, mc, R - \rho\}$ are proportional to $e^{-\xi t}$, e.g., $x(t) = x(0)\, e^{-\xi t}$, hence differ-

[40] Another measure (e.g., Mongey 2018) is the standard deviation of output when the economy is hit by recurring monetary shocks hence at a yearly frequency $c_t = e^{-\frac{1}{h}} c_{t-1} + m_t$. Under that metric the doubling of the half-life is equivalent to a 17% increase in the standard deviation of output $\sigma_c = \frac{\sigma_m}{\sqrt{1-e^{-2/h}}}$.

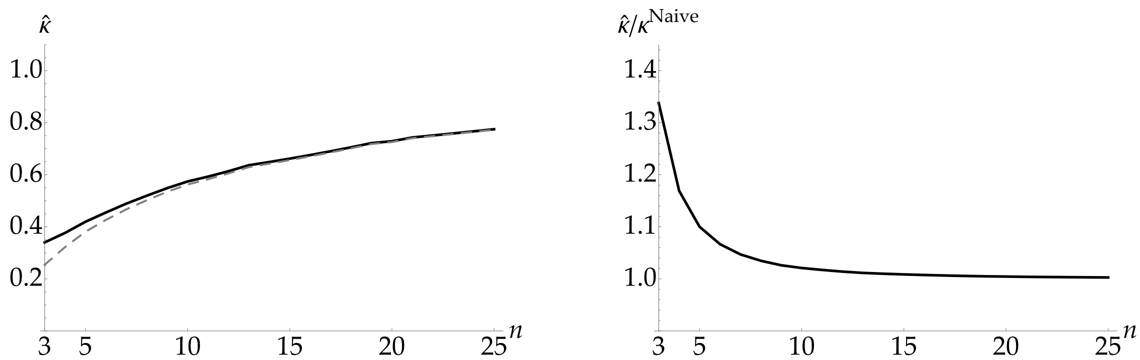

Figure 8: Effective slope of the Phillips curve $\hat{\kappa}$, strategic vs. naive oligopoly.

Note: Left panel: $\hat{\kappa} = \frac{\xi+\rho}{\psi+\sigma} \cdot \frac{\pi(0)}{c(0)}$ as a function of n following an interest rate shock ε_0^m with decay $\xi = 10$ under AIK calibration. Solid black line: (strategic) oligopolistic Phillips curve (19). Dashed gray line: κ^{Naive}. Right panel: Ratio $\hat{\kappa}/\kappa^{\text{Naive}}$ as a function of n.

ences across models are summarized by the impact effects on consumption $c(0)$ and inflation $\pi(0)$.[41] Since $mc(0) = (\psi+\sigma)c(0)$ then

$$\hat{\kappa} = \frac{\xi+\rho}{\psi+\sigma} \cdot \frac{\pi(0)}{c(0)}$$

is defined to reveal the actual slope κ in the special case of a first-order Phillips curve; in particular $\hat{\kappa} = \kappa^{\text{Naive}}$ in the naive economy. More generally it captures the trade-off between inflation and output, even in the more complex strategic oligopoly model.

The left panel of Figure 8 compares $\hat{\kappa}$ to κ^{Naive} as a function of n (under the AIK calibration); the right panel shows the ratio $\hat{\kappa}/\kappa^{\text{Naive}}$ as a measure of strategic effects. The message is consistent with what we found for permanent money shocks: concentration amplifies monetary non-neutrality by a significant amount. The left panel shows that a large part of the amplification can again be explained by feedback effects, that is, through the lens of the naive model. For low n strategic effects are more substantial than we found earlier: the naive model actually underestimates the effective slope $\hat{\kappa}$ by around 30% when $n = 3$. But strategic effects vanish rapidly as n increases.[42]

[41]One can recover the permanent money shock case from Proposition 3 by setting $\xi = \lambda(1-B)$ since then $\phi_\pi \pi + \varepsilon^m = 0$ so $R(t)$ is unchanged.

[42]The shock is very transitory as the exponential decay ξ is set at 10; more persistent shocks bring $\hat{\kappa}/\kappa^{\text{Naive}}$ even closer to 1.

Other Shocks: News Shocks and Liquidity Traps. In the Appendix we consider two more sophisticated policy experiments. First, we assume the previous monetary shock is realized at some date $t_{\text{shock}} > 0$:

$$\varepsilon^m(t) = \begin{cases} 0 & t < t_{\text{shock}} \\ \varepsilon_0^m e^{-\xi(t-t_{\text{shock}})} & t \geq t_{\text{shock}}. \end{cases}$$

This captures a news shock about monetary policy (the same results obtain with news about real shocks ε^r). Figure A.8 shows the similarity of the impulse responses in the strategic and naive models, for different shock dates $t_{\text{shock}} = 1, 2, 3$.

Finally we consider a liquidity trap scenario, in which the nominal interest rate R is stuck at zero from $t = 0$ to $t = T$ while the natural real rate $\rho + \varepsilon^r$ turns negative to -1%, i.e., $\varepsilon^m = 0$ and

$$\varepsilon^r(t) = \begin{cases} -\rho - 0.01 & t < T \\ 0 & t \geq T. \end{cases}$$

At $t = T$ the economy reverts to the steady state with $c(T) = \pi(T) = 0$ (for instance because the central bank lacks commitment). Figure A.9 shows the impact effects $c(0)$ and $\pi(0)$ as a function of the length of the trap T, for two economies, $n = 3$ and $n = \infty$. Just like higher price flexibility λ leads to more deflation and deeper recession (e.g.,Werning 2012), for given λ the stickiness due to oligopoly significantly dampens the severity of the trap by weakening the deflationary response. Figure A.10 compares the impact effects $c(0)$ and $\pi(0)$ for the strategic and naive models as a function of the length of the trap T: the responses are almost identical for short traps, but start diverging as T increases.

Overall, for a variety of shocks we find that the naive model captures most of the dynamics, except for very small n, in which strategic effects can cause up to a 30% increase in the effective slope of the Phillips curve. By implication, despite the new terms in our exact oligopolistic Phillips curve, a standard first order NK Phillips curve, appropriately parameterized using $\kappa^{\text{Naive}}(n)$, provides a very good approximation.

8 Conclusion

We conclude by collecting some takeaways and directions for future work suggested by our analysis.

Quantitatively we find that market concentration has a large effect on monetary policy transmission. The direction and magnitude of the effect depend on how concentration affects the superelasticity of demand. Under a calibration that matches pass-through estimates, going from monopolistic competition $n = \infty$ to an oligopoly with $n = 3$ firms doubles the half-life of the price level following monetary shocks, or equivalently divides the slope of the Phillips curve by four.

A central insight of our paper is that simple game-theoretic, partial equilibrium, objects—the slopes B_s—encapsulate the relevant pricing interactions and the general equilibrium response to standard monetary shocks. These slopes B_s can be computed from sufficient statistics; our model predicts that measures of sectoral non-neutrality are positively related to markups, after controlling for elasticities, frequency and concentration.[43] The markup, in turn, is not simply a function of the demand elasticity as in a static model. For example, a higher superelasticity of demand increases the markup.

We show that a simpler model with monopolistic competition and non-CES (e.g., Kimball) demand, when properly recalibrated, goes a long way towards approximating the dynamic responses to shocks. We propose a way to calibrate the relevant properties of demand using pass-through estimates and applied it using one particular study. More generally, our results highlight the importance of understanding how demand elasticities and superelasticities depend on market concentration, an important avenue for further empirical work.

Menu costs introduce several additional effects which we abstracted from here. Higher concentration may affect the frequency of price changes by reducing the profit-losses from failing to adjust prices (e.g., Rotemberg and Saloner 1987). Moreover, for a given average frequency of price changes, concentration interacts with menu costs through two effects: the selection effect (i.e., which firms are more likely to adjust), and possibly by coordinating the price changes (i.e., increasing the correlation between different firms' price changes). In Mongey (2018) the first effect dominates and

[43] Menu costs models may also imply a positive relation between markups and non-neutrality: lower demand elasticity can increase both markups and monetary non-neutrality, but it does so by lowering the frequency of price changes (e.g., Alvarez and Lippi 2014). The effect we describe is different as it is conditional on frequency.

generates additional amplification from oligopoly relative to Calvo pricing. The second effect, on the other hand, works towards price flexibility. In extreme cases this is what Nirei and Scheinkman (2021) call "repricing avalanches".

Klenow and Willis (2016) have pointed out that high strategic complementarities from Kimball demand are difficult to reconcile with the observed large price changes. An interesting possibility in an oligopolistic model with firm heterogeneity may help explain this fact. Small firms adjust their prices in response to cost shocks, whereas larger firms may have more market power, and only pass through a fraction of their cost shocks, but drive most of the aggregate price stickiness. This suggests distinguishing empirically the distributions of price changes by small and large firms.

We focused on Markov Perfect equilibria, the dominant equilibrium concept in Industrial Organization. Extending the analysis to non-Markov equilibria with "trigger strategies" seems feasible. Indeed, Proposition 1 applies to any equilibrium of the dynamic game, and can be used to study the aggregate response to a monetary shock for non-Markov equilibria.

To solve for the set of subgame perfect equilibria (e.g., the best and worst) one can employ the methods in Abreu, Pearce and Stacchetti (1990) adapted to the staggered price-setting structure. Although a full analysis is beyond the scope of this paper, when discounting is low enough, the best equilibrium achieves "perfect collusion": on the equilibrium path, firms set prices to maximize the present value of total sectoral profits. One can show that this implies greater aggregate price stickiness than under the Markov equilibrium. Understanding the extent to which this conclusion extends to other situations (e.g. low discounting) or other equilibria (not the best equilibrium) is a promising avenue for future research.

There is much more to investigate. We emphasized that the shape of demand is crucial to understand the transmission of shocks, and affected by trends in concentration. Another possibility is that macroeconomic shocks also affect the shape of demand, creating a force for inflation independent of the marginal cost changes we focused on.

References

Abreu, Dilip, David Pearce, and Ennio Stacchetti, "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 1990, *58* (5), 1041–1063.

Afrouzi, Hassan, "Strategic Inattention, Inflation Dynamics, and the Non-Neutrality of Money," working paper 2020.

Alvarez, Fernando and Francesco Lippi, "Price Setting With Menu Cost for Multi-product Firms," *Econometrica*, 2014, *82* (1), 89–135.

__ , __ , and Juan Passadore, "Are State- and Time-Dependent Models Really Different?," in "NBER Macroeconomics Annual 2016, Volume 31" University of Chicago Press November 2016, pp. 379–457.

__ , Hervé Le Bihan, and Francesco Lippi, "The Real Effects of Monetary Shocks in Sticky Price Models: A Sufficient Statistic Approach," *American Economic Review*, October 2016, *106* (10), 2817–51.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings, "International Shocks, Variable Markups, and Domestic Prices," *The Review of Economic Studies*, 02 2019.

Andrés, Javier and Pablo Burriel, "Inflation and optimal monetary policy in a model with firm heterogeneity and Bertrand competition," *European Economic Review*, 2018, *103*, 18 – 38.

Atkeson, Andrew and Ariel Burstein, "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, December 2008, *98* (5), 1998–2031.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen, "Concentrating on the Fall of the Labor Share," *American Economic Review*, May 2017, *107* (5), 180–85.

Barkai, Simcha, "Declining Labor and Capital Shares," *The Journal of Finance*, 2020.

Benigno, Pierpaolo and Ester Faia, "Globalization, Pass-Through, and Inflation Dynamics," *International Journal of Central Banking*, 2016, *12* (4).

Berger, David and Joseph Vavra, "Shocks versus Responsiveness: What Drives Time-Varying Dispersion?," *Journal of Political Economy*, 2019, *127* (5), 2104–2142.

Bergin, Paul R. and Robert C. Feenstra, "Staggered price setting, translog preferences, and endogenous persistence," *Journal of Monetary Economics*, 2000, *45* (3), 657–680.

Berman, Nicolas, Philippe Martin, and Thierry Mayer, " How do Different Exporters React to Exchange Rate Changes? *," *The Quarterly Journal of Economics*, 01 2012, *127* (1), 437–492.

Bilbiie, Florin O, Fabio Ghironi, and Marc J Melitz, "Monetary Policy and Business Cycles with Endogenous Entry and Product Variety," Working Paper 13199, National Bureau of Economic Research June 2007.

Brown, Zach Y. and Alexander MacKay, "Competition in Pricing Algorithms," working paper 2021.

Burstein, Ariel, Vasco Carvalho, and Basile Grassi, "Bottom-Up Markup Fluctuations," working paper 2020.

Carlton, Dennis W., "The Rigidity of Prices," *The American Economic Review*, 1986, *76* (4), 637–658.

Carvalho, Carlos, "Heterogeneity in Price Stickiness and the Real Effects of Monetary Shocks," *The B.E. Journal of Macroeconomics*, December 2006, *6* (3), 1–58.

Chatterjee, Arpita, Rafael Dix-Carneiro, and Jade Vichyanond, "Multi-Product Firms and Exchange Rate Fluctuations," *American Economic Journal: Economic Policy*, May 2013, *5* (2), 77–110.

Corhay, Alexandre, Howard Kung, and Lukas Schmid, "Q: Risk, Rents, or Growth?," Working Paper 2020.

DellaVigna, Stefano and Matthew Gentzkow, "Uniform Pricing in U.S. Retail Chains*," *The Quarterly Journal of Economics*, 06 2019, *134* (4), 2011–2084.

Etro, Federico and Lorenza Rossi, "New-Keynesian Phillips curve with Bertrand competition and endogenous entry," *Journal of Economic Dynamics and Control*, 2015, *51*, 318 – 340.

Gertler, Mark and John Leahy, "A Phillips Curve with an Ss Foundation," *Journal of Political Economy*, 2008, *116* (3), 533–572.

Goldberg, Pinelopi Koujianou and Frank Verboven, "The Evolution of Price Dispersion in the European Car Market," *The Review of Economic Studies*, 2001, *68* (4), 811–848.

Golosov, Mikhail and Robert E. Lucas, "Menu Costs and Phillips Curves," *Journal of Political Economy*, 2007, *115* (2), 171–199.

Gopinath, Gita and Oleg Itskhoki, "Frequency of Price Adjustment and Pass-Through*," *The Quarterly Journal of Economics*, 05 2010, *125* (2), 675–727.

Gutiérrez, Germán and Thomas Philippon, "Declining Competition and Investment in the U.S.," Working Paper 23583, National Bureau of Economic Research July 2017.

Jun, Byoung and Xavier Vives, "Strategic incentives in dynamic duopoly," *Journal of Economic Theory*, 2004, *116* (2), 249 – 281.

Kimball, Miles S., "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 1995, *27* (4), 1241–1277.

Klenow, Peter J. and Jonathan L. Willis, "Real Rigidities and Nominal Price Changes," *Economica*, 2016, *83* (331), 443–472.

Krusell, Per, Burhanettin Kuruscu, and Anthony A. Smith, "Equilibrium Welfare and Government Policy with Quasi-geometric Discounting," *Journal of Economic Theory*, 2002, *105* (1), 42 – 72.

Levintal, Oren, "Taylor Projection: A New Solution Method for Dynamic General Equilibrium Models," *International Economic Review*, 2018, *59* (3), 1345–1373.

Maskin, Eric and Jean Tirole, "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles," *Econometrica*, 1988, *56* (3), 571–599.

Matsuyama, Kiminori and Philip Uschchev, "Beyond CES: Three Alternative Classes of Flexible Homothetic Demand Systems," working paper 2017.

Midrigan, Virgiliu, "Menu Costs, Multiproduct Firms, and Aggregate Fluctuations," *Econometrica*, 2011, *79* (4), 1139–1180.

Mongey, Simon, “Market structure and monetary non-neutrality,” working paper 2018.

Nakamura, Emi and Dawit Zerom, “Accounting for Incomplete Pass-Through,” *The Review of Economic Studies*, 07 2010, *77* (3), 1192–1230.

_ and Jón Steinsson, “Monetary Non-neutrality in a Multisector Menu Cost Model*,” *The Quarterly Journal of Economics*, 08 2010, *125* (3), 961–1013.

_ and _ , “Price Rigidity: Microeconomic Evidence and Macroeconomic Implications,” *Annual Review of Economics*, 2013, *5* (1), 133–163.

Neiman, Brent, “A state-dependent model of intermediate goods pricing,” *Journal of International Economics*, 2011, *85* (1), 1 – 13.

Nirei, Makoto and José Scheinkman, “Repricing Avalanches,” working paper 2021.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter, “Diverging Trends in National and Local Concentration,” *NBER Macro Annual*, 2020.

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

_ and _ , “The Relative Rigidity of Monopoly Pricing,” *American Economic Review*, December 1987, *77* (5), 917–926.

_ and Michael Woodford, “Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity,” *Journal of Political Economy*, 1992, *100* (6), 1153–1207.

Slade, Margaret E., “Optimal Pricing with Costly Adjustment: Evidence from Retail-Grocery Prices,” *The Review of Economic Studies*, 01 1998, *65* (1), 87–107.

Weintraub, Gabriel Y., C. Lanier Benkard, and Benjamin Van Roy, “Markov Perfect Industry Dynamics With Many Firms,” *Econometrica*, 2008, *76* (6), 1375–1411.

Werning, Iván, “Managing a Liquidity Trap: Monetary and Fiscal Policy,” working paper 2012.

Appendix
For Online Publication

A Dynamics after a Monetary Shock

A.1 Exact Dynamics

Proof of Proposition 1. If the consumer maximizes

$$\int e^{-\rho t}\left[\frac{C(t)^{1-\sigma}}{1-\sigma}-\frac{N(t)^{1+\psi}}{1+\psi}+\frac{m(t)^{1-\chi}}{1-\chi}\right]dt$$

we have

$$\frac{\dot{C(t)}}{C(t)}=\frac{1}{\sigma}\left(R(t)-\pi(t)-\rho\right)$$

$$N(t)^{\psi}C(t)^{\sigma}=\frac{W(t)}{P(t)}\Rightarrow\psi\frac{\dot{N(t)}}{N(t)}=\frac{\dot{W(t)}}{W(t)}-R(t)+\rho$$

$$M(t)^{-\chi}P(t)^{\chi}C(t)^{\sigma}=R(t)$$

We look for an equilibrium with constant nominal interest rate $R(t)=R$ and nominal wage $W(t)=W$ following a permanent shock to M. Suppose $\psi=0$ then we get

$$\frac{\dot{W(t)}}{W(t)}=R-\rho$$

To get constant wage $W(t)=W$ we need $R=\rho$ (this is necessary, otherwise we would get permanent wage inflation). The constant wage implies

$$P(t)C(t)^{\sigma}=W$$

Then the third equation gives

$$\rho M^{\chi}=P(t)^{\chi}C(t)^{\sigma}$$

So we need $\chi=1$ for our guess to be indeed an equilibrium.

The representative consumer's expenditure in sector s at time t is

$$E_s(t)=P_s(t)^{1-\omega}\left[C(t)P(t)^{\omega}\right]$$

where $P(t)$ is the aggregate price level $\left(\int_s P_s(t)^{1-\omega} ds\right)^{\frac{1}{1-\omega}}$ hence the real demand vector in sector s is (given our within-sector CRS assumption as in Kimball)

$$D\left(\left\{p_{j,s}(t)\right\}, E_s(t)\right) = D\left(\left\{p_{j,s}(t)\right\}, 1\right) P_s(t)^{1-\omega} C(t) P(t)^{\omega}$$

where P_s is the sectoral price index.Denote the function of prices in sector s only

$$d\left(\left\{p_{j,s}\right\}\right) = D\left(\left\{p_{j,s}\right\}, 1\right) P_s^{1-\omega}$$

The nominal profit of firm i in sector s given all the other prices in the economy is

$$d^i\left(p_{i,s}, p_{-i,s}\right) C(t) P(t)^{\omega} \left[p_{i,s} - W(t) \frac{f^{-1}\left(d^i\left(p_{i,s}, p_{-i,s}\right) C(t) P(t)^{\omega}\right)}{d^i\left(p_{i,s}, p_{-i,s}\right) C(t) P(t)^{\omega}}\right]$$

where $p_{-i,s} = \left\{p_{j,s}\right\}_{j \neq i}$. Thus the real profit is

$$d^i\left(p_{i,s}, p_{-i,s}\right) C(t) P(t)^{\omega-1} \left[p_{i,s} - W(t) \frac{f^{-1}\left(d^i\left(p_{i,s}, p_{-i,s}\right) C(t) P(t)^{\omega}\right)}{d^i\left(p_{i,s}, p_{-i,s}\right) C(t) P(t)^{\omega}}\right]$$

Firm i maximizes the present value of real profits discounted using the SDF $e^{-\rho t} C(t)^{-\sigma}$, that is

$$\int e^{-\rho t} C(t)^{1-\sigma} P(t)^{\omega-1} d^i\left(p_{i,s}, p_{-i,s}\right) \left[p_{i,s} - W(t) \frac{f^{-1}\left(d^i\left(p_{i,s}, p_{-i,s}\right) C(t) P(t)^{\omega}\right)}{d^i\left(p_{i,s}, p_{-i,s}\right) C(t) P(t)^{\omega}}\right] dt.$$

With general σ (but linear disutility of labor and log-utility of real balances, that are needed to obtain constant nominal interest rate and wage) we have that

$$P(t) C(t)^{\sigma} = W,$$

therefore if

$$\omega \sigma = 1$$

then the terms

$$\begin{aligned} C(t)^{1-\sigma} P(t)^{\omega-1} &= W^{\frac{1}{\sigma}-1} P(t)^{\omega-\frac{1}{\sigma}} \\ C(t) P(t)^{\omega} &= W^{\frac{1}{\sigma}} P(t)^{\omega-\frac{1}{\sigma}} \end{aligned}$$

are constant. Denote $\hat{p}_s = \left(\frac{p_{1,s}}{1+\delta}, \ldots, \frac{p_{n,s}}{1+\delta}\right)$ the vector of normalized prices. The present discounted value of real profits is

$$W^{1/\sigma-1}\int e^{-\rho t}d^{i}\left(\hat{p}_{s}\right)\left[p_{i,s}-W\frac{f^{-1}\left(d^{i}\left(p_{i,s},p_{-i,s}\right)W^{1/\sigma}\right)}{d^{i}\left(p_{i,s},p_{-i,s}\right)W^{1/\sigma}}\right]dt$$

$$=W_{-}^{1/\sigma-1}\left(1+\delta\right)^{1/\sigma-\omega}\int e^{-\rho t}d^{i}\left(\hat{p}_{s}\right)\left[\hat{p}_{i,s}-W_{-}\frac{f^{-1}\left(d^{i}\left(\hat{p}_{s}\right)W_{-}^{1/\sigma}\left(1+\delta\right)^{1/\sigma-\omega}\right)}{d^{i}\left(\hat{p}_{s}\right)W_{-}^{1/\sigma}\left(1+\delta\right)^{1/\sigma-\omega}}\right]dt$$

$$=W_{-}^{1/\sigma-1}\int e^{-\rho t}d^{i}\left(\hat{p}_{s}\right)\left[\hat{p}_{i,s}-W_{-}\frac{f^{-1}\left(d^{i}\left(\hat{p}_{s}\right)W_{-}^{1/\sigma}\right)}{d^{i}\left(\hat{p}_{s}\right)W_{-}^{1/\sigma}}\right]dt$$

which is exactly the same as before the shock up to the change of variables $p \to \hat{p}$.

A.2 Approximate Dynamics

Proof of Proposition 2. Fix n and a sector $s \in [0,1]$. Define the state $v_s(t)$ as

$$v_s = (z_1, \ldots, z_n)'$$

where $z_i = \log p_i - \log \bar{p}$. Denote the first-order expansion of the best response $p_i' = g\left(p_{-i}, P\right)$ by

$$z_i' = \alpha Z + \beta\left(\sum_{j\neq i} z_j\right)$$

where $Z(t) = \log P(t) - \log \bar{p}$ is the log deviation of the aggregate price level. Proposition 1 shows that $\alpha = 0$ if $\omega\sigma = 1$; otherwise α will be non-zero and we derive the aggregation in the general case.

When firm i adjusts its price, the state of sector s changes to

$$v_s'(t) = \alpha Z(t) u_i + M_i v_s(t)$$

where u_i is the vector $(0, \ldots, 0, \underset{\substack{\uparrow \\ i}}{0}, 0, \ldots, 0)$ M_i is the identity matrix except for row i which is equal to $(\beta, \ldots, \beta, \underset{\substack{\uparrow \\ i}}{0}, \beta, \ldots, \beta)$.

First suppose that all sectors are identical. Define the aggregate state variable

$$V(t) = \int_{s\in[0,1]} v_s(t)ds \in \mathbb{R}^n$$

Between t and $t+\Delta t$, a mass $n\lambda\Delta t$ of firms adjusts prices so V evolves as

$$\begin{aligned}
V(t+\Delta t) &= (1-n\lambda\Delta t)V(t) + \int_{\text{a firm in } s \text{ adjusts}} v_s(t+\Delta t)ds \\
&= (1-n\lambda\Delta t)V(t) + (\lambda n\Delta t)\left[\alpha Z(t)\frac{\sum_i u_i}{n} + \frac{\sum_i M_i}{n}V(t)\right]
\end{aligned}$$

therefore in the limit $\Delta t \to 0$

$$\dot{V}(t) = \lambda\alpha Z(t)U + n\lambda\left(\frac{\sum_i M_i}{n} - I_n\right)V(t)$$

where $U = \sum_i u_i = (1,\ldots,1)'$ and

$$\frac{\sum_i M_i}{n} - I_n = \begin{pmatrix} \frac{-1}{n} & \frac{\beta}{n} & \cdots & \frac{\beta}{n} \\ \frac{\beta}{n} & \frac{-1}{n} & \cdots & \frac{\beta}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\beta}{n} & \frac{\beta}{n} & \cdots & \frac{-1}{n} \end{pmatrix}$$

The aggregate price level is then $Z(t) = LV_t$ where $L = \frac{1}{n}(1,\ldots,1)$. The eigenvalues of $n\lambda\left(\frac{\sum_i M_i}{n} - I_n\right)$ are:

- $\mu_1 = -\lambda(1+\beta)$ with multiplicity $n-1$,
- $\mu_2 = -\lambda[1-(n-1)\beta]$ with multiplicity 1.

The vector U is an eigenvector associated with μ_2, so if we start from symmetric initial conditions $V(0) = (\log p_0 - \log \bar{p})\, U$ we have

$$V(t) = V\,(0)\, e^{(\lambda\alpha+\mu_2)t}$$

hence finally, the price index evolves to first order in δ as:

$$\begin{aligned}
\log\left(\frac{P(t)}{\bar{P}}\right) &= \log\left(\frac{P(0)}{\bar{P}}\right) e^{-\lambda[1-\alpha-(n-1)\beta]t} \\
&= -\delta e^{-\lambda[1-\alpha-(n-1)\beta]t}
\end{aligned}$$

With heterogeneous sectors s the aggregation across sectors yields

$$\log\left(\frac{P(t)}{\bar{P}}\right) = -\delta\int_s \zeta_s e^{-\lambda_s[1-\alpha_s-(n_s-1)\beta_s]t}ds$$

where ζ_s is the steady state expenditure share of sector s.

B Markov Equilibrium and Sufficient Statistics

Let $V^{i,s}(p;t)$ denote the value function for firm i, where p is the vector of n_s prices. We focus on equilibria with differentiable g and V satisfying the Bellman equation

$$R(t)V^{i,s}(p;t) = \Pi^{i,s}(p;t) + \lambda_s \sum_{j \in I_s} \Big(V^{i,s}(g^{j,s}(p_{-j};t), p_{-j};t) - V^{i,s}(p;t) \Big) + \frac{\partial V^{i,s}}{\partial t}(p;t) \tag{A.1}$$

where $g^{j,s}(p_{-j};t)$ satisfies the optimality condition $g^{j,s}(p_{-j};t) \in \arg\max_{p_j} V^{j,s}(p_j, p_{-j};t)$ with first-order necessary condition

$$V^{j,s}_{p_j}(g^{j,s}(p_{-j};t), p_{-j};t) = 0 \tag{A.2}$$

for all j.

Proof of Proposition 4. Differentiating the Bellman equation (A.1) and making use of symmetry, we obtain at the steady state $\bar{p}$ of a symmetric equilibrium:

$$0 = \Pi^i_{p_i}(\bar{p}) + \lambda \sum_{j \neq i} \left[V^i_{p_j}(\bar{p}) \frac{\partial g^j}{\partial p_i}(\bar{p}) \right]$$

$$V^i_{p_j}(\bar{p}) = \frac{\Pi^i_{p_j}(\bar{p})}{\rho + \lambda} + \frac{\lambda}{\rho + \lambda} \sum_{k \neq i,j} \left[V^i_{p_k}(\bar{p}) \frac{\partial g^k}{\partial p_j}(\bar{p}) \right] \quad \forall j \neq i$$

Using $\sum_j \sum_{k \neq i,j} V^i_{p_k}(\bar{p}) = (n-2) \sum_{j \neq i} V^i_{p_j}(\bar{p})$, the second condition becomes

$$\sum_{k \neq i} V^i_{p_k}(\bar{p}) = \frac{\sum_{k \neq i} \frac{\Pi_{p_k}(\bar{p})}{\rho + \lambda}}{1 - \frac{\lambda (n-2) \beta_n}{\rho + \lambda}}$$

Hence the first condition becomes

$$0 = \Pi_{p_i}(\bar{p}) + \frac{\lambda \beta_n}{\rho + \lambda \left[1 - (n-2)\beta_n\right]} \sum_{k \neq i} \Pi_{p_k}(\bar{p})$$

and the symmetry of $\Pi^i_{p_j}$ across $j \neq i$, we obtain

$$0 = \Pi^i_i(\bar{p}) + \frac{\lambda \,(n-1)\, \beta}{\rho + \lambda \left[1 - (n-2)\beta\right]} \Pi^i_j(\bar{p})$$

thus the formula for $B = (n-1)\,\beta$ is

$$B = \frac{\rho+\lambda}{\lambda}\frac{1}{\frac{n-2}{n-1}+\left(\frac{\Pi_j^i}{-\Pi_i^i}\right)}. \tag{A.3}$$

We can reexpress

$$\frac{\Pi_j^i}{-\Pi_i^i} = \frac{\epsilon_j^i\,(1-1/\mu)}{-\epsilon_i^i\,(1-1/\mu)-1}$$

where $\mu = \frac{\bar{p}}{W/f'\left(f^{-1}\left(d^i(\bar{p})\right)\right)}$ is the steady state markup (the denominator is the marginal cost) to rewrite (A.3) in terms of demand own-elasticity $\epsilon_i^i = \frac{\partial \log d^i}{\partial \log p_i}$ and cross-elasticity $\epsilon_j^i = \frac{\partial \log d^i}{\partial \log p_j}$:

$$B = \frac{\rho+\lambda}{\lambda}\frac{1}{\frac{n-2}{n-1}+\frac{\epsilon_j^i}{-\epsilon_i^i-\frac{\mu}{\mu-1}}}.$$

Homothetic preferences imply that the cross-elasticity is related to the own-elasticity through $(n-1)\epsilon_j^i = -(\omega+\epsilon_i^i)$.

$$B = \frac{\lambda+\rho}{\lambda}\frac{1}{1+\frac{1-(\mu-1)(\omega-1)}{(n-1)[(\epsilon-1)(\mu-1)-1]}}$$

where $\epsilon = \left|\epsilon_i^i\right|$.

C Demand Elasticities

C.1 General non-parametric results

We first assume an outer elasticity $\omega = 1$. Differentiating the budget constraint, we have for any i and p

$$c^i + \sum_j p_j \frac{\partial c^j}{\partial p_i} = 0 \tag{A.4}$$

Then Slutsky symmetry and constant returns to scale imply

$$\epsilon_i^i + \sum_{j\neq i} \epsilon_j^i = -1 \tag{A.5}$$

where $\epsilon^i_j = \frac{\partial \log c^i}{\partial \log p_J}$. At a symmetric price, this becomes

$$\epsilon^i_j = -\frac{1+\epsilon^i_i}{n-1} \tag{A.6}$$

so the convergence to Nash holds as long as the own elasticity ϵ^i_i is bounded. Call for any pair j, k

$$\epsilon^i_{jk} = \frac{\partial^2 \log d_i}{\partial \log p_k \partial \log p_j}$$

We can differentiate (A.5) with respect to $\log p_i$ to get

$$\epsilon^i_{ii} + \sum_{j \neq i} \epsilon^i_{ij} = 0$$

hence at a symmetric price,

$$\epsilon^i_{ii} + (n-1)\epsilon^i_{ij} = 0 \tag{A.7}$$

Differentiating once more the budget constraint with respect to p_i

$$2\frac{\partial c^i}{\partial p_i} + \sum_j \frac{\partial^2 c^j}{\partial p_i^2} = 0 \tag{A.8}$$

Elasticities and second-derivatives are related by

$$\frac{\partial^2 c^i}{\partial p_k \partial p_j} = \frac{c^i}{p_k p_j}\left[\epsilon^i_{jk} + \epsilon^i_j \epsilon^i_k\right] \text{ for any } j \neq k$$

$$\frac{\partial^2 c^i}{\partial p_j^2} = \frac{c^i}{p_j^2}\left[\epsilon^i_{jj} - \epsilon^i_j + \left(\epsilon^i_j\right)^2\right] \text{ for any } j$$

At a symmetric price (using $\epsilon^j_{ii} = \epsilon^i_{jj}$), we have from (A.8)

$$\epsilon^i_{jj} = \epsilon^i_j\left(1-\epsilon^i_j\right) - \frac{1}{n-1}\left[\epsilon^i_{ii} + \epsilon^i_i\left(1+\epsilon^i_i\right)\right] \tag{A.9}$$

Finally, differentiating (A.4) with respect to p_k for some $k \neq i$ gives

$$\frac{\partial c^i}{\partial p_k} + \frac{\partial c^k}{\partial p_i} + \sum_{j \neq i,k} p_j \frac{\partial^2 c^j}{\partial p_k \partial p_i} + p_i \frac{\partial^2 c^i}{\partial p_k \partial p_i} + p_k \frac{\partial^2 c^k}{\partial p_k \partial p_i} = 0$$

and at a symmetric price p

$$\frac{2}{p}\frac{\partial c^i}{\partial p_k} + (n-2)\frac{\partial^2 c^i}{\partial p_k \partial p_j} + 2\frac{\partial^2 c^i}{\partial p_k \partial p_i} = 0$$

Therefore, in elasticities at a symmetric price,

$$2\epsilon^i_j + (n-2)\left[\epsilon^i_{jk} + \left(\epsilon^i_j\right)^2\right] + 2\left[\epsilon^i_{ij} + \epsilon^i_j \epsilon^i_i\right] = 0 \tag{A.10}$$

for $k \neq j, i$, $j \neq i$. The own-superelasticity is defined as the elasticity of (minus the) elasticity:

$$\Sigma = \frac{\partial \log(-\epsilon^i_i)}{\partial \log p_i} = \frac{\epsilon^i_{ii}}{\epsilon^i_i}$$

So in the end we have two degrees of freedom: $\{\epsilon^i_i, \epsilon^i_{ii}\}$ or equivalently $\{\epsilon, \Sigma\}$ to parametrize a symmetric steady state.

In the non-Cobb-Douglas case $\omega \neq 1$, all the steps are almost the same except that we start from the sectoral budget constraint

$$\sum_{i \in I_s} p_i d^i = \mathcal{P}_s^{1-\omega}$$

where $\mathcal{P}_s$ is the sectoral price index. As a result the elasticities at a symmetric price satisfy (A.7), (A.10) as before, but (A.6) and (A.9) become respectively

$$\epsilon^i_j = -\frac{\omega + \epsilon^i_i}{n-1}$$
$$\epsilon^i_{jj} = \epsilon^i_j\left(1 + \epsilon^i_j\right) - \frac{1}{n-1}\left[\epsilon^i_{ii} + \epsilon^i_i\left(\omega + \epsilon^i_i\right)\right].$$

Special case: $n = 2$. If $n = 2$ there is only 1 degree of freedom, so CES is without loss of generality (locally), even when the outer aggregation is not Cobb-Douglas (i.e., $\omega \neq 1$). From (A.10), the cross-superelasticity ϵ^i_{ij} is determined by elasticities, hence so is $\epsilon^i_{ii} = -(n-1)\epsilon^i_{ij}$.

C.2 Closed-form elasticities with Kimball Demand

Here again we outline the steps under Cobb-Douglas preferences across sectors, $\omega = 1$, but give the general expressions with $\omega \neq 1$ below.

Start with a general Kimball (1995) aggregator that defines C as

$$\frac{1}{n}\sum_i \Psi\left(\frac{c_i}{C}\right) = 1 \tag{A.11}$$

where Ψ is increasing, concave, and $\Psi(1) = 1$ which ensures the convention that at a symmetric basket $c_i = c$, we have $C = c$. The consumer's problem is

$$\min_{\{c_i\}} \sum_i p_i c_i \text{ s.t. } \frac{1}{n}\sum_i \Psi\left(\frac{c_i}{C}\right) = 1$$

There exists a Lagrange multiplier $\lambda > 0$ such that for all i

$$p_i = \lambda \Psi'\left(\frac{c_i}{C}\right)\frac{1}{C} \tag{A.12}$$

If we define the Kimball sectoral price index P (which differs from the ideal price index except under CES) by

$$\frac{1}{n}\sum_i \varphi\left(\Psi'(1)\frac{p_i}{P}\right) = 1$$

where

$$\varphi = \Psi \circ (\Psi')^{-1}$$

then at a symmetric price $p_i = p$ we have $P = p$, and $\lambda \Psi'(1) = PC$ so we can rewrite (A.12) as

$$\frac{p_i}{P}\Psi'(1) = \Psi'\left(\frac{c_i}{C}\right)$$

Taking logs and differentating (A.12) with respect to $\log p_i$ yields

$$1 = \frac{\partial \log P}{\partial \log p_i} + \frac{\Psi''\left(\frac{c_i}{C}\right)}{\Psi'\left(\frac{c_i}{C}\right)}\frac{c_i}{C}\left[\epsilon_i^i - \frac{\partial \log C}{\partial \log p_i}\right]$$

Differentiating (A.11) yields

$$\sum_j \Psi'\left(\frac{c_j}{C}\right)\frac{c_j}{C}\left[\frac{\partial \log c_j}{\partial \log p_i} - \frac{\partial \log C}{\partial \log p_i}\right] = 0$$

hence

$$\frac{\partial \log C}{\partial \log p_i} = \frac{\sum_j \Psi'\left(\frac{c_j}{C}\right)\frac{c_j}{C}\epsilon_i^j}{\sum_j \Psi'\left(\frac{c_j}{C}\right)\frac{c_j}{C}}$$

Using Slutsky symmetry $p_j\epsilon_i^j = p_i\epsilon_j^i$ to express this using demand elasticities for good i only, we can reexpress as

$$\frac{\partial \log C}{\partial \log p_i} = \frac{\sum_j \Psi'\left(\frac{c_j}{C}\right)\frac{c_j}{C}\frac{p_i}{p_j}\epsilon_j^i}{\sum_j \Psi'\left(\frac{c_j}{C}\right)\frac{c_j}{C}}$$

At a symmetric price, budget exhaustion with constant returns implies

$$\frac{\partial \log C}{\partial \log p_i} = \frac{1}{n}\sum_j \epsilon_j^i = \frac{-1}{n}$$

For any $k \neq i$ we can differentiate

$$\log \Psi'\left(\frac{c^i}{C}\right) - \log \Psi'\left(\frac{c^k}{C}\right) = \log p_i - \log p_k$$

with respect to $\log p_i$ to get

$$\frac{\Psi''\left(\frac{c^i}{C}\right)}{\Psi'\left(\frac{c^i}{C}\right)}\left(\frac{c^i}{C}\right)\frac{\partial}{\partial \log p_i}\left[\log c^i - \log C\right] - \frac{\Psi''\left(\frac{c^k}{C}\right)}{\Psi'\left(\frac{c^k}{C}\right)}\left(\frac{c^k}{C}\right)\frac{\partial}{\partial \log p_i}\left[\log c^k - \log C\right] = 1$$

or, defining

$$R(x) = -\frac{x\Psi''(x)}{\Psi'(x)}$$

We have

$$R\left(\frac{c^k}{C}\right)\left[\epsilon_i^k - \frac{\partial \log C}{\partial \log p_i}\right] - R\left(\frac{c^i}{C}\right)\left[\epsilon_i^i - \frac{\partial \log C}{\partial \log p_i}\right] = 1 \tag{A.13}$$

Hence at a symmetric steady state, using $\epsilon_i^k = \epsilon_k^i = -\frac{1+\epsilon_i^i}{n-1}$ we have

$$\epsilon_i^i = -\left(\frac{n-1}{n}\frac{1}{R(1)} + \frac{1}{n}\right)$$

Differentiating once more with respect to $\log p_i$,

$$-R'\left(\frac{c^i}{C}\right)\left[\epsilon_i^i - \frac{\partial \log C}{\partial \log p_i}\right]^2 + R'\left(\frac{c^k}{C}\right)\left[\epsilon_i^k - \frac{\partial \log C}{\partial \log p_i}\right]^2 - R\left(\frac{c^i}{C}\right)\left[\epsilon_{ii}^i - \frac{\partial^2 \log C}{\partial^2 \log p_i}\right] + R\left(\frac{c^k}{C}\right)\left[\epsilon_{ii}^k - \frac{\partial^2 \log C}{\partial^2 \log p_i}\right] = 0$$

At a symmetric steady state,

$$-R'(1)\left[\epsilon_i^i + \frac{1}{n}\right]^2 + R'(1)\left[\epsilon_i^k + \frac{1}{n}\right]^2 - R(1)\left[\epsilon_{ii}^i - \epsilon_{ii}^k\right] = 0$$

$$-R'(1)\left[\epsilon_i^i+\frac{1}{n}\right]^2+R'(1)\left[\epsilon_i^k+\frac{1}{n}\right]^2-R(1)\left[\epsilon_{ii}^i-\epsilon_{jj}^i\right]=0$$

Using (A.9) we get

$$-R'(1)\left[\frac{n-1}{n}\frac{1}{R(1)}\right]^2+R'(1)\left[-\frac{1+\epsilon_i^i}{n-1}+\frac{1}{n}\right]^2-R(1)\left[\epsilon_{ii}^i\frac{n}{n-1}-\epsilon_j^i\left(1-\epsilon_j^i\right)+\frac{1}{n-1}\left[\epsilon_i^i\left(1+\epsilon_i^i\right)\right]\right]=0$$

Now differentiating (A.13) with respect to $\log p_j$ for some $j\neq i,k$

$$R'\left(\frac{c^i}{C}\right)\left[\epsilon_j^i-\frac{\partial\log C}{\partial\log p_j}\right]\left[\epsilon_i^i-\frac{\partial\log C}{\partial\log p_i}\right]+R\left(\frac{c^i}{C}\right)\left[\epsilon_{ij}^i-\frac{\partial^2\log C}{\partial\log p_i\partial\log p_j}\right]$$
$$-R'\left(\frac{c^k}{C}\right)\left[\epsilon_i^k-\frac{\partial\log C}{\partial\log p_i}\right]\left[\epsilon_j^k-\frac{\partial\log C}{\partial\log p_j}\right]-R\left(\frac{c^k}{C}\right)\left[\epsilon_{ij}^k-\frac{\partial^2\log C}{\partial\log p_i\partial\log p_j}\right]=0$$

At a symmetric price,

$$R'(1)\left[\epsilon_j^i+\frac{1}{n}\right]\left[\epsilon_i^i+\frac{1}{n}\right]+R(1)\,\epsilon_{ij}^i=R'(1)\left[\epsilon_j^i+\frac{1}{n}\right]^2+R(1)\,\epsilon_{jk}^i$$

Therefore, using (A.10) we have

$$\epsilon_i^i=-\left[\left(\frac{n-1}{n}\right)\frac{1}{R(1)}+\frac{1}{n}\right] \tag{A.14}$$

$$\epsilon_j^i=\frac{\frac{1}{R(1)}-1}{n}$$

$$\epsilon_{ii}^i=-\frac{n-1}{n^2}\left[\frac{R(1)\left[1-R(1)\right]^2+(n-2)R'(1)}{R(1)^3}\right]$$

$$\epsilon_{ij}^i=\frac{R(1)\left[1-R(1)\right]^2+(n-2)R'(1)}{n^2R(1)^3}\quad(j\neq i)$$

$$\epsilon_{jj}^i=\frac{-(n-1)R(1)\left[1-R(1)\right]^2+(n-2)R'(1)}{n^2R(1)^3}\quad(j\neq i)$$

$$\epsilon_{jk}^i=\frac{R(1)\left[1-R(1)\right]^2-2R'(1)}{n^2R(1)^3}\quad(j\neq k,n\geq 3)$$

In the general case $\omega \neq 1$, following similar steps these expressions generalize to

$$\begin{aligned}
\epsilon_i^i &= -\left[\left(\frac{n-1}{n}\right)\frac{1}{R(1)} + \frac{1}{n}\omega\right] \\
\epsilon_j^i &= \frac{\frac{1}{R(1)} - \omega}{n} \\
\epsilon_{ii}^i &= -\frac{n-1}{n^2}\left[\frac{R(1)\left[1-R(1)\right]\left[1-R(1)\omega\right] + (n-2)R'(1)}{R(1)^3}\right] \\
\epsilon_{ij}^i &= \frac{R(1)\left[1-R(1)\right]\left[1-R(1)\omega\right] + (n-2)R'(1)}{n^2 R(1)^3} \qquad (j \neq i) \\
\epsilon_{jj}^i &= \frac{-(n-1)R(1)\left[1-R(1)\right]\left[1-R(1)\omega\right] + (n-2)R'(1)}{n^2 R(1)^3} \qquad (j \neq i) \\
\epsilon_{jk}^i &= \frac{R(1)\left[1-R(1)\right]\left[1-R(1)\omega\right] - 2R'(1)}{n^2 R(1)^3} \qquad (j \neq k, n \geq 3)
\end{aligned}$$

Equations (10)-(11) are written using the more convenient $\varphi(x) = 1/R(x)$.

Klenow and Willis (2016) use the functional form

$$\Psi'(x) = \frac{\eta - 1}{\eta}\exp\left(\frac{1 - x^{\theta/\eta}}{\theta}\right)$$

$$\Psi''(x) = -\frac{x^{\frac{\theta}{\eta}-1}}{\eta}\Psi'(x)$$

$$\Psi'''(x) = \left[\left(\frac{x^{\frac{\theta}{\eta}-1}}{\eta}\right)^2 - \left(\frac{\theta - \eta}{\eta^2}\right)x^{\frac{\theta}{\eta}-2}\right]\Psi'(x)$$

Therefore

$$\begin{aligned}
R(1) &= \frac{1}{\eta} \\
R'(1) &= \frac{\theta}{\eta^2}
\end{aligned}$$

so that this nests CES with $\theta = 0$. We thus have

$$\epsilon_i^i = -\frac{\eta\,(n-1)+\omega}{n} \tag{A.15a}$$

$$\epsilon_j^i = \frac{\eta-\omega}{n} \tag{A.15b}$$

$$\epsilon_{ii}^i = -\frac{(n-1)}{n^2}\left[\eta^2-(1+\omega)\,\eta+\omega+(n-2)\theta\eta\right] \tag{A.15c}$$

$$\epsilon_{ij}^i = \frac{\eta^2-(1+\omega)\,\eta+\omega+(n-2)\,\theta\eta}{n^2} \tag{A.15d}$$

$$\epsilon_{jj}^i = \frac{(n-2)\theta\eta-(\eta-1)(n-1)\,(\eta-\omega)}{n^2} \tag{A.15e}$$

$$\epsilon_{jk}^i = \frac{\eta^2-(1+\omega)\,\eta+\omega-2\theta\eta}{n^2} \tag{A.15f}$$

With $\omega = 1$ as in the main text, the superelasticity, defined as $\Sigma = \frac{\epsilon_{ii}^i}{\epsilon_i^i}$, satisfies

$$\begin{aligned}\Sigma = &= \frac{1}{\frac{S}{1-S}+\eta}\left[\theta\eta+\left((\eta-1)^2-2\theta\eta\right)S\right]\\ &\approx \theta+\left[\frac{(\eta-1)^2}{\eta}-2\theta\right]S\end{aligned}$$

with $S = 1/n$ denoting the market share. The approximation in the second line holds if S is small relative to $\eta/\left(1+\eta\right)$, as is the case in a calibration with $\eta = 10$. With constant θ and η, the superelasticity is approximately linear in the Herfindahl index. If θ is lower than $\frac{(\eta-1)^2}{2\eta}$ which equals 4.05 when $\eta = 10$ (as in the CES case $\theta = 0$) then Σ increases with S. With high enough θ, it can actually decrease with S, but a high fixed θ is at odds with pass-through being larger for smaller firms.

D Solution Method

Iteratively differentiating the Bellman equation (A.1) and the optimality condition (A.2) generates a system of equations relating the derivatives of the reaction function g', g'', and so on, to the steady state markup, demand elasticity ϵ_i^i, superelasticity ϵ_{ii}^i, and so on. Our formula (9) is one of such equations.

The standard interpretation of this system treats the sequence of derivatives of g as unknowns, and the infinite sequence of higher-order elasticities as given structural parameters. Instead, we acknowledge that it is empirically impossible to know such fine properties of preferences or demand functions, since we can only estimate a finite number of elasticities. This leads us to take a dual view of the same system of equations: we still take low order elasticities as given, but choose the values of the unknown higher order elasticities to achieve some desired properties for the derivatives of g. In particular, we can find primitives such that the reaction function g is locally polynomial of order m, meaning that all its derivatives of higher order than m vanish when evaluated at the steady state.

Formally, let

$$\epsilon_{(1)} = \frac{\partial \log d^i}{\partial \log p_i}, \quad \epsilon_{(k)} = \frac{\partial \epsilon_{(k-1)}}{\partial \log p_i} \quad \forall k \geq 2.$$

Proposition 9. *For any order $m \geq 1$ and target elasticities $\left(\epsilon_{(1)}, \ldots, \epsilon_{(m)}\right)$, there exist Kimball within-sector preferences $\tilde{\phi}$ such that*

(i) the resulting elasticities up to order m match the target elasticities, and

(ii) any MPE of the game with within-sector preferences $\tilde{\phi}$, strategy $\tilde{g}$ and steady state $\tilde{p}$ satisfies $\tilde{g}^{(k)}(\tilde{p}) = 0$ for $k \geq m$.

Another interpretation is to view the infinite sequence of elasticities as structural: for instance, we could assume that preferences are exactly CES and compute the implied elasticities of any order. In this context our method is then an approximation of the exact solution given by the limit $m \to \infty$ where we can match all elasticities.

Under this interpretation we can evaluate the accuracy of the approximation by noting that for low n, we can compute the exact solution $m \to \infty$ using standard value function iteration. We then compare the resulting steady state price to what follows from our solution method with finite m. Figure A.1 plots the steady state markup with $m = 1, 2, 3$ in the case of a duopoly, showing that $m = 2$ already provides an excellent approximation (within 1%) to the exact solution $m \to \infty$ and going to a higher order $m = 3$ improves the fit but not by much. Note that low n allows us to check numerically the accuracy of the approximation, but we know theoretically that the approximation should be even better as n grows, since all the orders m of approximation coincide with monopolistic competition as $n \to \infty$.

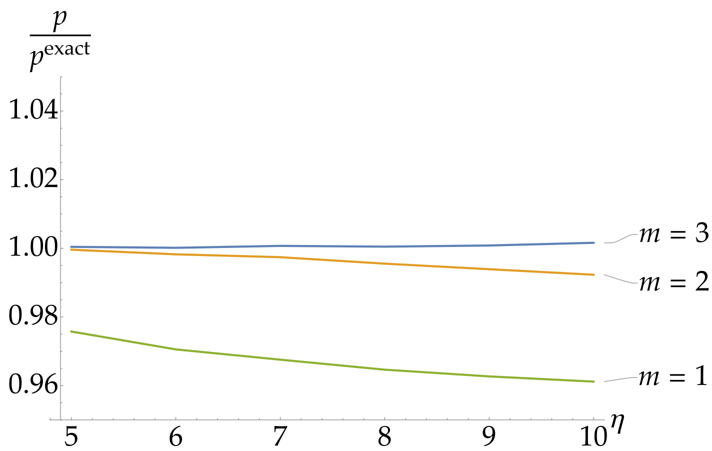

Figure A.1: Steady state markup p with $n = 2$ firms, under our solution method with $m = 1, 2, 3$, relative to exact solution p^{exact} (which corresponds to $m \to \infty$).

Proof of Proposition 9. We start from the system that defines an MPE:

$$(\rho + n\lambda) V(p) = \Pi(p) + \lambda \sum_j V\left(g(p_{-j}), p_{-j}\right) \tag{A.16}$$

$$V_p\left(g(p_{-i}), p_{-i}\right) = 0 \tag{A.17}$$

Differentiating k times the Bellman equation (A.16) gives us for each $k \geq 1$ a linear system in the kth-derivatives $\mathbf{V}^{(k)} = (V_{11\ldots11}, V_{11\ldots12}, V_{11\ldots22}, \ldots)$ of the value function V (evaluated at the symmetric steady state $\bar{p}$), which we can invert to obtain these derivatives as a function of the profit derivatives $\mathbf{\Pi}^{(k)} = (\Pi_{11\ldots11}, \ldots)$ and derivatives of the policy function (there are $k+1$ such equations in the case of $n = 2$ firms).

We can then compute $\mathbf{\Pi}^{(k)}$ as a function of $\bar{p}$ and own- and cross-superelasticities of the demand function d of order up to k.

Combining the solution $\mathbf{V}^{(k)}$ with the $k-1$th-derivative of the FOC (A.17) gives us a sequence of equations that must be satisfied at a steady state

$$F^k\left(\bar{p}, g'(\bar{p}), g''(\bar{p}), \ldots, g^{(k)}(\bar{p}); \epsilon_{(0)}, \epsilon_{(1)}, \epsilon_{(2)}, \ldots, \epsilon_{(k)}\right) = 0$$

where F^k is *linear* in $\tilde{\epsilon}_{(k)}$. Thus we can construct recursively a unique sequence $\tilde{\epsilon}_{(k)}$

starting from $k = m+1$, using

$$F^{m+1}\left(\bar{p}, g', \ldots g^{(m-1)}, 0, 0; \epsilon_{(1)}, \epsilon_{(2)}, \ldots, \tilde{\epsilon}_{(m+1)}\right) = 0$$
$$F^{m+2}\left(\bar{p}, g', \ldots g^{(m-1)}, 0, 0, 0; \epsilon_{(1)}, \epsilon_{(2)}, \ldots, \tilde{\epsilon}_{(m+1)}, \tilde{\epsilon}_{(m+2)}\right) = 0$$

and so on. Below we show that for $n \geq 3$ there are indeed enough degrees of freedom to make the equations $F^m, F^{m+1}, \ldots$ independent.

Define $\tilde{\varphi}$ as

$$\tilde{\varphi}(x) = \sum_{k=0}^{\infty} \frac{\tilde{\varphi}^{(k)}(1)}{k!} (x-1)^k$$

where $\tilde{\varphi}^{(k+1)}(1)$ is characterized by $\left(\epsilon_{(1)}, \ldots, \epsilon_{(m)}, \tilde{\epsilon}_{(m+1)}, \ldots, \tilde{\epsilon}_{(k)}\right)$ through the same computations as in Appendix C. Given this construction, $\bar{p}, g', \ldots, g^{(m-1)}$ are pinned down by $\left(\epsilon_{(1)}, \ldots, \epsilon_{(m)}\right)$ as the solution to the system of equations F^k for $k = 1, \ldots, m$.

The main potential impediment to the proof is that demand integrability (e.g., demand functions being generated by actual utility functions) imposes restrictions on higher-order elasticities that would prevent us from constructing the sequence $\tilde{\epsilon}$. Indeed, in Appendix C we saw that with $n = 2$ firms, general Kimball demand functions cannot generate superelasticities beyond those arising from CES demand. We now show that as long as $n \geq 3$, this is not the case, by proving that the number of elasticities exceeds the number of restrictions.

Formally, we want to compute $\#_n(m)$, the number of cross-elasticities of order m, that is derivatives

$$\frac{\partial^m \log d^1(p)}{\partial^{i_1} \log p_1 \partial^{i_2} \log p_2 \ldots \partial^{i_n} \log p_n}$$

where

$$0 \leq i_1, \ldots, i_n \leq m$$
$$i_1 + \cdots + i_n = m$$

as functions of the own-mth-elasticity $\epsilon^1_{\underbrace{11\ldots1}_{m \text{ times}}}$, and compare $\#_n(m)$ to the number of restrictions imposed by demand integrability and symmetry arguments.

By Schwarz symmetry, in a smooth MPE, we can always invert 2 indices in the derivatives. Moreover, from the viewpoint of firm 1 (whose demand d^1 we're differentiating), firms 2 and 3 are interchangeable. For instance, in the case of $n = 3$ firms and order of differentiation $m = 3$, these symmetries reduce the number of potential elasticities $n^m = 27$ to only 6 elasticities

$$\epsilon^1_{111}, \epsilon^1_{112}, \epsilon^1_{122}, \epsilon^1_{123}, \epsilon^1_{222}, \epsilon^1_{223}.$$

Denote

$$q_n(M)$$

the number of partitions of an integer M into n non-negative integers. For $M \geq n$ we have

$$q_n(M) = p_n(M+n)$$

where $p_n(M)$ is the number of partitions of an integer M into n positive integers. We can see this by writing, starting from a partition of M into n non-negative integers $i_1, \ldots, i_n$:

$$M + n = (i_1 + 1) + \cdots + (i_n + 1)$$

We can then compute $p_j(M)$ using the recurrence formula

$$p_j(M) = \underbrace{p_j(M-j)}_{\text{partitions for which } i_k \geq 2 \text{ for all } k} + \underbrace{p_{j-1}(M-1)}_{\text{partitions for which } i_k = 1 \text{ for some } k}$$

Lemma 1. *For any $n \geq 1$ and $m \geq 1$ the number of elasticities of order m is*

$$\#_n(m) = \sum_{k=0}^{m} q_{n-1}(m-k) \tag{A.18}$$

hence $\#_n(m+1) = \#_n(m) + q_{n-1}(m+1)$.

Proof. Firm 1 is special, so we need to count the number of times we differentiate with respect to $\log p_1$, which generates the sum over k. Then we get each term in the sum by counting partitions of $m-k$ into $n-1$ non-negative integers. □

Next, we want to count the reduction in the number of degrees of freedom imposed by economic restrictions. Our restrictions are

$$\Phi(p) = \sum_j p_j d^j(p) = 0 \quad \forall p \tag{A.19}$$

$$d^i_j(p) = d^j_i(p) \quad \forall p, \forall i, j \tag{A.20}$$

The first equation is the budget constraint. The second equation is the Slutsky symmetry condition (constant returns to scale allow to go from Hicksian to Marshallian elasticities). Note that Φ defined in (A.19) is symmetric, unlike the demand function d^1 we are using to compute elasticities. Therefore Φ's derivatives give us fewer restrictions than what we need in (A.18), leaving room for restrictions to come from the Slutsky equation.

We need to differentiate these two equations to obtain independent equations that relate the mth-cross-elasticities to the mth-own-elasticity. The number of restrictions coming from derivatives of Φ at order m is simply the number $q_n(m)$ of partitions of m into n non-negative integers. Denote $\flat_n(m)$ the number of restrictions we have from derivatives of the Slutsky equation. The initial equation $d_2^1 = d_1^2$ is irrelevant at a symmetric steady state; it only starts mattering once we differentiate it. We actually do not need to compute $\flat_n(m)$ exactly. The following lemma shows that there are always enough degrees of freedom $\#_n(m)$ to construct the Kimball aggregator in 9:

Lemma 2. *For $n \geq 3$ and any m we have*

$$q_n(m) + \flat_n(m) + 1 \leq \#_n(m) \tag{A.21}$$

Proof. We know by hand that (A.21) holds for $m = 1, 2$ so take $m \geq 3$. Then all the Slutsky conditions can be written as starting with

$$d^1_{12\ldots} = \ldots$$

hence we have

$$\flat_n(m) \leq \#_n(m-2) = \#_n(m) - p_{n-1}(n+m-1) - p_n(n+m-2)$$

hence the number of equations is bounded by

$$q_n(m) + \flat_n(m) \leq p_n(n+m) + \#_n(m) - p_{n-1}(n+m-1) - p_n(n+m-2)$$

Then we have (A.21) if

$$\begin{aligned}
& p_n(n+m) < p_{n-1}(n+m-1) + p_n(n+m-2) \\
\Leftrightarrow & p_{n-1}(n+m-1) + p_n(m) < p_{n-1}(n+m-1) + p_n(n+m-2) \\
\Leftrightarrow & p_n(m) < p_n(n+m-2)
\end{aligned}$$

which holds for $n \geq 3$. $\square$

Note that so far we have considered general CRS demand functions. Restricting attention to the Kimball class makes the inequality (A.21) bind, meaning that we can parametrize all the cross-elasticities of order m using the own-elasticity of order m.

What fails in the knife-edge case $n = 2$? Slutsky symmetry imposes too many restrictions: at $m = 2$ we only have 3 elasticities $\epsilon^1_{11}, \epsilon^1_{12}, \epsilon^1_{22}$ and also 3 restrictions, so we can solve out all the superelasticities as functions of ϵ^1_1, which prevents us from constructing the Kimball aggregator in Proposition 9.

E Model Solution

We apply the solution method described in Appendix D to derive analytical expressions in the case $m = 2$.

E.1 Symmetric Firms

We first solve the linear system in $\left\{V^i_j, V^i_{ii}, V^i_{ij}, V^i_{jj}, V^i_{jk}\right\}$ obtained from envelope conditions

$$\begin{aligned}
(\rho + \lambda)\, V^i_j &= \Pi^i_j + \lambda\,(n-2)\, V^i_j \beta \\
(\rho + \lambda)\, V^i_{ii} &= \Pi^i_{ii} + \lambda\,(n-1)\left(V^i_{jj}\beta^2 + 2V^i_{ij}\beta\right) \\
(\rho + 2\lambda)\, V^i_{ij} &= \Pi^i_{ij} + \lambda\,(n-2)\left(V^i_{jj}\beta^2 + V^i_{ij}\beta + V^i_{jk}\beta\right) \\
(\rho + \lambda)\, V^i_{jj} &= \Pi^i_{jj} + \lambda\,(n-2)\left(V^i_{jj}\beta^2 + 2V^i_{jk}\beta\right) + \lambda\left(V^i_{ii}\beta^2 + 2V^i_{ij}\beta\right) \\
(\rho + 2\lambda)\, V^i_{jk} &= \Pi^i_{jk} + \lambda\,(n-3)\left(V^i_{jj}\beta^2 + 2V^i_{jk}\beta\right) + \lambda\left(V^i_{ii}\beta^2 + 2V^i_{ij}\beta\right)
\end{aligned}$$

Injecting the solution into the derivative of the first-order condition

$$V^i_{ii}\beta + V^i_{ij} = 0$$

yields an equation

$$0 = A_{ii}\Pi^i_{ii}(\bar{p}) + A_{ij}\Pi^i_{ij}(\bar{p}) + A_{jj}\Pi^i_{jj}(\bar{p}) + A_{jk}\Pi^i_{jk}(\bar{p}) \tag{A.22}$$

with coefficients

$$
\begin{aligned}
A_{ii} =& \beta((\beta+1)\lambda^3\left(\beta^2\left(-2n^2+9n-10\right)+\beta^3(n-2)+6\beta(n-2)-4\right) \\
&-\lambda^2\rho\left(\beta^3\left(n^2-5n+6\right)+\beta^2\left(2n^2-15n+22\right)+\beta(24-9n)+8\right) \\
&+\lambda\rho^2\left(\beta^2(n-2)+\beta(3n-8)-5\right)-\rho^3)
\end{aligned} \tag{A.23a}
$$

$$
\begin{aligned}
A_{ij} =& -2(\beta+1)\lambda^3\left(-2\beta^3\left(n^2-3n+2\right)+\beta^4(n-1)+2\beta^2(n-1)-\beta(n-2)+1\right) \\
&+\lambda^2\rho\left(\beta^4\left(-2n^2+7n-5\right)-4\beta^3\left(n^2-4n+3\right)+3\beta^2 n-4\beta(n-3)+5\right) \\
&+\lambda\rho^2\left(\beta^2 n-2\beta(n-3)+4\right)+\rho^3
\end{aligned} \tag{A.23b}
$$

$$
\begin{aligned}
A_{jj} =& \beta^2\lambda((\beta+1)\lambda^2\left(2\left(\beta^2+3\beta+2\right)+\beta(\beta+1)n^2-\left(3\beta^2+7\beta+2\right)n\right) \\
&+\lambda\rho\left(4\beta^2+10\beta+\beta(\beta+1)n^2-\left(5\beta^2+9\beta+3\right)n+6\right) \\
&+\rho^2(\beta-(\beta+1)n+2))
\end{aligned} \tag{A.23c}
$$

$$
\begin{aligned}
A_{jk} =& -\beta\lambda(n-2)((\beta+1)\lambda^2\left(-\beta+\beta^3(n-1)+3\beta^2(n-1)+1\right) \\
&+\lambda\rho\left(2\beta^3(n-1)+\beta^2(3n-4)+2\right)+\rho^2)
\end{aligned} \tag{A.23d}
$$

Finally $\bar{p}^3\Pi^i_{ii}(\bar{p}), \bar{p}^3\Pi^i_{ij}(\bar{p}), \bar{p}^3\Pi^i_{jj}(\bar{p}), \bar{p}^3\Pi^i_{jk}(\bar{p})$ are all linear functions of $\bar{p}$ and W. Therefore, multiplying (A.22) by $\frac{\bar{p}^3}{W}$ we get a linear equation in μ which can be solved to obtain a function

$$
\mu = \mu\left(B, \omega, \epsilon, \Sigma, n, \lambda/\rho\right). \tag{A.24}
$$

Equation (A.24) together with the sufficient statistic formula (9)

$$
B = B\left(\mu, \omega, \epsilon, n, \lambda/\rho\right)
$$

form a system of two equations in the two unknowns μ and β.

E.2 Heterogeneous Firms

The demand faced by firm i is

$$
c_i = \frac{1}{\xi_i} d^i\left(\tilde{p}_i, \tilde{p}_{-i}\right)
$$

where d^i is the demand function from the symmetric case ($\xi_i = 1$ for all i) and $\tilde{p}_j = p_j/\xi_j$ is the normalized price of good j. As a result the nominal profit of firm i can be written as

$$\Pi^i(t) = \tilde{p}_i(t) d^i\left(\tilde{p}_i(t), \tilde{p}_{-i}(t)\right) - W(t) f^{-1}\left(\frac{d^i\left(\tilde{p}_i(t), \tilde{p}_{-i}(t)\right)}{\xi_i z_i}\right) \tag{A.25}$$

where d^i is the previous demand function from the symmetric firms model, and $\tilde{p}_j = p_j/\xi_j$ is the normalized price of good j. If $\xi_i z_i = 1$, the model with normalized prices is isomorphic to one with symmetric firms.

Suppose as in Section 6.2 that there are two types of firms, a and b, with $n = n_a + n_b$. a and b firms can differ permanently in their productivity z, their demand shifters ξ, or both. With two types we need to solve for six unknowns: two steady state prices $\{p_a, p_b\}$ and four slopes $\{\beta_a^a, \beta_b^a, \beta_a^b, \beta_b^b\}$ where β_j^i is the slope of the reaction of a firm of type i to the price change of a firm of type j. The envelope conditions for firms of type a are

$$\begin{aligned}
(\rho+\lambda) V_i^{i,a} &= \Pi_i^{i,a} + \lambda\left(n_a - 1\right) V_{j_a}^{i,a} \beta_a^a + \lambda n_b V_{j_b}^{i,a} \beta_a^b \\
(\rho+\lambda) V_{j_a}^{i,a} &= \Pi_{j_a}^{i,a} + \lambda\left(n_a - 2\right) V_{j_a}^{i,a} \beta_a^a + \lambda n_b V_{j_b}^{i,a} \beta_a^b \\
(\rho+\lambda) V_{j_b}^{i,a} &= \Pi_{j_b}^{i,a} + \lambda\left(n_a - 1\right) V_{j_a}^{i,a} \beta_b^a + \lambda\left(n_b - 1\right) V_{j_b}^{i,a} \beta_b^b
\end{aligned}$$

and, in the locally linear equilibrium:

$$\begin{aligned}
(\rho+\lambda) V_{ii}^{i,a} &= \Pi_{ii}^{i,a} + \lambda\left(n_a-1\right)\left[V_{j_a j_a}^{i,a}\left(\beta_a^a\right)^2 + 2V_{ij_a}^{i,a}\beta_a^a\right] + \lambda n_b\left[V_{j_b j_b}^{i,a}\left(\beta_a^b\right)^2 + 2V_{ij_b}^{i,a}\beta_a^b\right] \\
(\rho+2\lambda) V_{ij_a}^{i,a} &= \Pi_{ij_a}^{i,a} + \lambda\left(n_a-2\right)\left[V_{j_a j_a}^{i,a}\left(\beta_a^a\right)^2 + V_{j_a k_a}^{i,a}\beta_a^a + V_{ij_a}^{i,a}\beta_a^a\right] + \lambda n_b\left[V_{j_b j_b}^{i,a}\left(\beta_a^b\right)^2 + V_{j_a k_b}^{i,a}\beta_a^b + V_{ij_b}^{i,a}\beta_a^b\right] \\
(\rho+2\lambda) V_{ij_b}^{i,a} &= \Pi_{ij_b}^{i,a} + \lambda\left(n_a-1\right)\left[V_{j_a j_a}^{i,a}\beta_a^a\beta_b^a + V_{j_a k_b}^{i,a}\beta_a^a + V_{ij_a}^{i,a}\beta_b^a\right] + \lambda\left(n_b-1\right)\left[V_{j_b j_b}^{i,a}\beta_a^b\beta_b^b + V_{j_b k_b}^{i,a}\beta_a^b + V_{ij_b}^{i,a}\beta_b^b\right] \\
(\rho+\lambda) V_{j_a j_a}^{i,a} &= \Pi_{j_a j_a}^{i,a} + \lambda\left[V_{ii}^{i,a}\left(\beta_a^a\right)^2 + 2V_{ij_a}^{i,a}\beta_a^a\right] + \lambda\left(n_a-2\right)\left[V_{j_a j_a}^{i,a}\left(\beta_a^a\right)^2 + 2V_{j_a k_a}^{i,a}\beta_a^a\right] + \lambda n_b\left[V_{j_b j_b}^{i,a}\left(\beta_a^b\right)^2 + 2V_{j_a k_b}^{i,a}\beta_a^b\right] \\
(\rho+2\lambda) V_{j_a k_b}^{i,a} &= \Pi_{j_a k_b}^{i,a} + \lambda\left[V_{ii}^{i,a}\beta_a^a\beta_b^a + V_{ij_b}^{i,a}\beta_a^a + V_{ij_a}^{i,a}\beta_b^a\right] + \lambda\left(n_a-2\right)\left[V_{j_a j_a}^{i,a}\beta_a^a\beta_b^a + V_{j_a k_b}^{i,a}\beta_a^a + V_{j_a k_a}^{i,a}\beta_b^a\right] \\
&\quad + \lambda\left(n_b-1\right)\left[V_{j_b j_b}^{i,a}\beta_a^b\beta_b^b + V_{j_b k_b}^{i,a}\beta_a^b + V_{j_a k_b}^{i,a}\beta_b^b\right] \\
(\rho+2\lambda) V_{j_a k_a}^{i,a} &= \Pi_{j_a k_a}^{i,a} + \lambda\left[V_{ii}^{i,a}\left(\beta_a^a\right)^2 + 2V_{ij_a}^{i,a}\beta_a^a\right] + \lambda\left(n_a-3\right)\left[V_{j_a j_a}^{i,a}\left(\beta_a^a\right)^2 + 2V_{j_a k_a}^{i,a}\beta_a^a\right] + \lambda n_b\left[V_{j_b j_b}^{i,a}\left(\beta_a^b\right)^2 + 2V_{j_a k_b}^{i,a}\beta_a^b\right] \\
(\rho+\lambda) V_{j_b j_b}^{i,a} &= \Pi_{j_b j_b}^{i,a} + \lambda\left[V_{ii}^{i,a}\left(\beta_b^a\right)^2 + 2V_{ij_b}^{i,a}\beta_b^a\right] + \lambda\left(n_a-1\right)\left[V_{j_a j_a}^{i,a}\left(\beta_b^a\right)^2 + 2V_{j_a k_b}^{i,a}\beta_b^a\right] + \lambda\left(n_b-1\right)\left[V_{j_b j_b}^{i,a}\left(\beta_b^b\right)^2 + 2V_{j_b k_b}^{i,a}\beta_b^b\right] \\
(\rho+2\lambda) V_{j_b k_b}^{i,a} &= \Pi_{j_b k_b}^{i,a} + \lambda\left[V_{ii}^{i,a}\left(\beta_b^a\right)^2 + 2V_{ij_b}^{i,a}\beta_b^a\right] + \lambda\left(n_a-1\right)\left[V_{j_a j_a}^{i,a}\left(\beta_b^a\right)^2 + 2V_{j_a k_b}^{i,a}\beta_b^a\right] + \lambda\left(n_b-2\right)\left[V_{j_b j_b}^{i,a}\left(\beta_b^b\right)^2 + 2V_{j_b k_b}^{i,a}\beta_b^b\right]
\end{aligned}$$

We can use these 11 envelope conditions to solve linearly for $\left\{V_i^{i,a}, V_{j_a}^{i,a}, V_{j_b}^{i,a}, V_{ii}^{i,a}, \ldots\right\}$, and then inject the solution into the first-order conditions

$$\begin{aligned} V_i^{i,a} &= 0 \\ V_{ii}^{i,a}\beta_a^a + V_{ij_a}^{i,a} &= 0 \\ V_{ii}^{i,a}\beta_b^a + V_{ij_b}^{i,a} &= 0 \end{aligned}$$

The same steps for firms of type b give us 3 more equations.

F Calibration to Pass-Through Evidence

In Section 4 we use evidence on own-cost pass-through from Amiti et al. (2019) (henceforth AIK) to calibrate how the superelasticity Σ varies with concentration. We describe the procedure in more detail here.

In the presence of permanent shocks to marginal costs mc_j, when firm i adjusts its price it sets

$$\log p_i - \log \bar{p}_i = \alpha\left(\log mc_i - \log \bar{c}_i\right) + B\frac{\sum_{j\neq i}\log p_j - \log \bar{p}_j}{n-1} + \gamma\sum_{j\neq i}\left(\log c_j - \log \bar{c}_j\right) \tag{A.26}$$

where the coefficients

$$\alpha = \frac{\partial g^i}{\partial mc_i}, \quad B = (n-1)\frac{\partial g^i}{\partial p_j}, \quad \gamma = \frac{\partial g^i}{\partial mc_j}$$

can be computed as before using our envelope conditions applied to a generalization of the Bellman equation (A.1) that allows for permanent cost shocks:

$$(\rho + n\lambda)\, V^i(p, mc) = \Pi^i(p, mc_i) + \lambda\sum_j V^i\left(g^j\left(p_{-j}, mc\right), p_{-j}, mc\right). \tag{A.27}$$

Unlike in static models of oligopoly (see Remark 2 below) γ is non-zero in general: although competitor j's cost c_j does not affect firm i's current profits, it will affect how firm j sets its price p_j in the future, which is relevant for firm i's future payoffs. Anticipating this, firm i will already respond itself to c_j when it gets to reset its price. The coefficients must satisfy the homogeneity restriction

$$\alpha + B + (n-1)\gamma = 1,$$

which says that if all firms' marginal costs increase by 1% then all firms' prices also increase by 1%.

Rewrite (A.26) in vector form as

$$\Delta\tilde{\boldsymbol{p}} = (\alpha I + \gamma S)\,\Delta\widetilde{\boldsymbol{mc}} + \beta S \Delta\tilde{\boldsymbol{p}}$$

where $S = J - I$ and J is the matrix with 1's everywhere, $\Delta\tilde{\boldsymbol{p}} = [\log p_i - \log \bar{p}_i]'$, $\Delta\widetilde{\boldsymbol{mc}} = [\log mc_i - \log \bar{mc}_i]'$. The following result describes the mapping from the parameters α, B in (A.26) to the regression estimates $\hat{\alpha}, \hat{B}$ in (12).

Proposition 10. *There exist unique scalars $\hat{\alpha}$, $\hat{B}$ such that for all vectors $\Delta\boldsymbol{mc}$*

$$\Delta\tilde{p}_i = \hat{\alpha}\Delta\widetilde{mc}_i + \hat{B}\frac{\sum_{j\neq i}\Delta\tilde{p}_j}{n-1}$$

for all i, namely

$$\hat{\alpha} = \frac{n\alpha + B - 1}{\alpha + B + n - 2} \tag{A.28}$$

$$\hat{B} = \frac{(n-1)(1-\alpha)}{\alpha + B + n - 2} \tag{A.29}$$

thus they satisfy $\hat{\alpha} + \hat{B} = 1$.

Proof. We need for all $\Delta\boldsymbol{c}$

$$\Delta\tilde{\boldsymbol{p}} = \hat{\alpha}\Delta\widetilde{\boldsymbol{mc}} + \hat{\beta} S \Delta\tilde{\boldsymbol{p}}$$

that is

$$\left(I - \hat{\beta}S\right)(I - \beta S)^{-1}(\alpha I + \gamma S) = \hat{\alpha} I$$

where $\beta = \frac{B}{n-1}$, $\hat{\beta} = \frac{\hat{B}}{n-1}$. Using $M = (I - \beta S)^{-1} = \sum_{k\geq 0}\beta^k S^k$ this is equivalent to

$$\sum_{k\geq 0}\beta^k\left[S^k - \hat{\beta}S^{k+1}\right](\alpha I + \gamma S) = \hat{\alpha} I$$

$$\sum_{k\geq 0}\beta^k\left[\alpha S^k + \gamma S^{k+1} - \alpha\hat{\beta}S^{k+1} - \gamma\hat{\beta}S^{k+2}\right] = \hat{\alpha} I$$

$$\alpha M + \frac{\gamma}{\beta}(M - I) - \alpha\frac{\hat{\beta}}{\beta}(M - I) - \gamma\frac{\hat{\beta}}{\beta^2}(M - I - \beta S) = \hat{\alpha} I$$

$$\alpha\beta M + \left(\gamma - \alpha\hat{\beta} - \gamma\hat{\beta}/\beta\right)(M - I) + \gamma\hat{\beta}S = \hat{\alpha}\beta I$$

Multiplying by $I-\beta S$ this becomes

$$\alpha\beta I+\left(\gamma-\alpha\hat{\beta}-\gamma\hat{\beta}/\beta\right)\beta S+\gamma\hat{\beta}\left(S-\beta S^{2}\right)=\hat{\alpha}\beta\left(I-\beta S\right)$$

$$\left(\gamma\beta-\alpha\beta\hat{\beta}+\hat{\alpha}\beta^{2}\right)S-\gamma\hat{\beta}\beta S^{2}=\left(\hat{\alpha}-\alpha\right)\beta I$$

Using

$$J^{2}=nJ$$

(recall that J is the matrix with ones everywhere) we have

$$S^{2}=\left(n-1\right)I+\left(n-2\right)S$$

Therefore $\hat{\alpha},\hat{\beta}$ must satisfy

$$\left(\gamma\beta-\alpha\beta\hat{\beta}+\hat{\alpha}\beta^{2}-\gamma\hat{\beta}\beta(n-2)\right)S=\left[\left(\hat{\alpha}-\alpha\right)\beta+\gamma\hat{\beta}\beta\left(n-1\right)\right]I$$

which can only be true if both sides are zero, that is (after replacing γ using the homogeneity restriction):

$$\hat{\alpha}=\frac{n\alpha+B-1}{\alpha+B+n-2}$$
$$\hat{B}=\frac{\left(n-1\right)\left(1-\alpha\right)}{\alpha+B+n-2}$$

□

Amiti et al. (2019) show that the empirical behavior of $\hat{\alpha}$ as a function of market share is well approximated by

$$\hat{\alpha}\approx\frac{1}{1+\frac{(\eta-1)(1-s)s(\eta-\omega)}{\omega(\eta-1)-s(\eta-\omega)}}$$

with $\eta=10$ and $\omega=1$. Therefore in a sector with n firms we set as target the corresponding pass-through $\hat{\alpha}_{n}=\frac{1}{1+9/n}$. Then, fixing other parameters (e.g., η,λ,ρ), for each (θ,n) we can compute α and B and solve for θ_{n} that sets allows to match $\hat{\alpha}_{n}$.

Figure A.2 shows the resulting pass-through as a function of market share $1/n$ under this "AIK" calibration, contrasting with the case of fixed $\theta=0$ (CES) and fixed $\theta=10$.

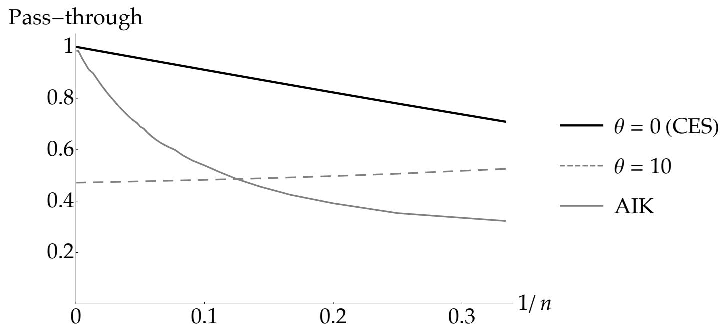

Figure A.2: Pass-through $\hat{\alpha}$ as a function of market share $1/n$.

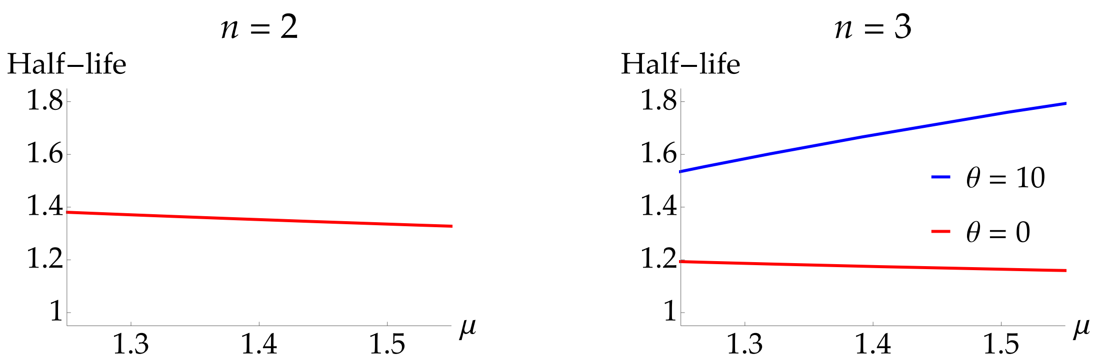

Figure A.3: Half-life as a function of steady state markup μ when η varies.

G Other Comparative Statics

Changes in Preference Parameters η and θ. Changes in η and θ affect both the steady state markup μ and the half-life of the price level following monetary shocks.

Figure A.3 shows the half-life as a function of the steady state markup, when variation in markups is produced through variation in the within-sector elasticity of substitution η; higher η implies lower markups. The effect on the half-life is ambiguous, however, except in the special case $n = 2$ in which there is always a negative relation between the markup and the half-life.[44] In particular, as soon as there are at least $n = 3$ firms, the value of θ matters. When $\theta = 0$ (CES), we have the same negative relation as in the duopoly case, but with a high enough value of θ, the half-life becomes negatively related to the steady state markup. We explain these patterns in Section 5.

[44] In Appendix C.1 we show that for any homothetic preferences, ϵ and Σ are the same as under CES when there are only $n = 2$ symmetric firms, whether the cross-sector aggregator has unit elasticity ω or not. This means that θ is irrelevant when $n = 2$, as can be seen in Figure 2, where all the curves coincide when $n = 2$. When n is above 2, however, knowing the markup is not enough to infer the slope, which is why formula (9) also requires information on demand elasticities.

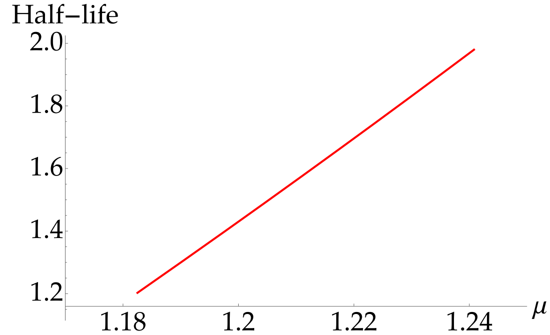

Figure A.4: Half-life as a function of steady state markup μ when θ varies.

We argued that under dynamic oligopoly, markups are not fully determined by demand elasticities. Figure A.4 shows the half-life as a function of the steady state markup, when variation in markups is produced through the superelasticity parameter θ. Higher θ implies higher markups, even though the demand elasticity ϵ is unchanged throughout. As we vary θ, all the objects appearing in the right-hand side of (9) remain fixed except μ, hence this experiment yields a transparent application of the formula showing how B and the half-life increase with μ.

Changes in Discount Rates and Price Stickiness. The discount rate ρ and the frequency of price changes λ can also affect the steady state markup (and therefore the slope B). These two parameters only enter through the ratio ρ/λ, so a higher frequency is isomorphic to a lower discount rate and we focus the discussion on λ.

Figure A.5 shows that markups increase with λ, especially when n is low. This shows once again that equilibrium markups are complex objects that depend on many features of the environment beyond demand elasticities. In the limit of infinitely sticky prices $\lambda \to 0$, firms play the one-shot best-response, and so the Markov equilibrium coincides with the static Bertrand-Nash equilibrium, both in terms of steady state markup and reaction functions, which is apparent in Figure A.5. Interestingly, the limit of infinitely frequent price changes $\lambda/\rho \to \infty$ does *not* equal the frictionless (flexible price) model, in which firms would play the static Bertrand-Nash equilibrium at each instant. For instance, when $n = 3$ (in red), the static markup is $\mu^{\text{Nash}} = 1.17$ while the steady state markup converges to $\mu = 1.24$ as $\lambda \to \infty$. For higher n, the gap between the Nash markup and the $\lambda \to \infty$ limit becomes negligible.[45]

[45] This discontinuity in markups in the limit of flexible prices or very patient firms has been noted in

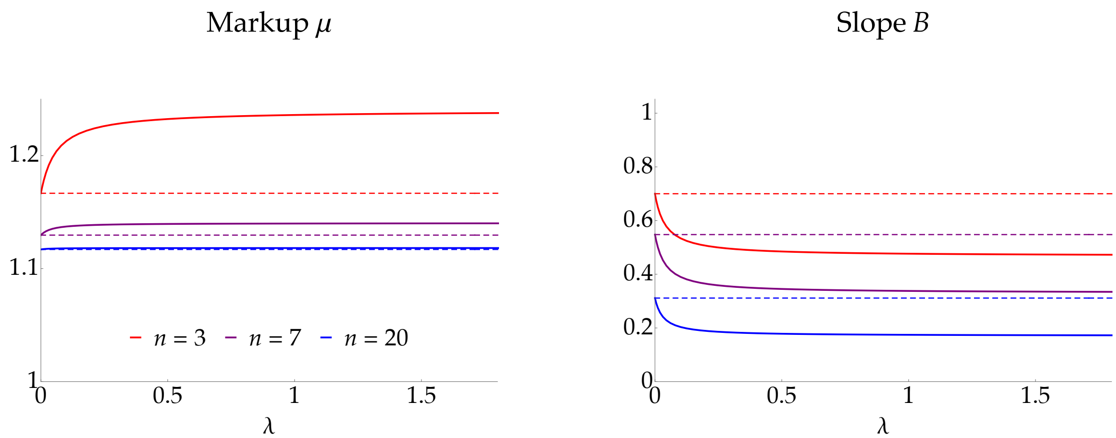

Figure A.5: Steady state markup μ and slope B as a function of frequency of price changes λ under "AIK" calibration. Dashed horizontal lines correspond to the static Bertrand-Nash equilibrium (μ^{Nash} and B^{Nash}).

H Solution of the Naive Model

The quadratic approximation of profit Π^i of firm i around the naive steady state which is the static Nash p^{Nash} writes

$$\pi^i(z_i, z_{-i}) = BQ_i + CQ_i^2 + Dz_iQ_i + Ez_i^2 + FR_i$$

where $z_j = \log p_j - \log p^{\text{Nash}}$ for each j and

$$Q_i = \sum_{j \neq i} z_j$$

$$R_i = \sum_{j \neq i} z_j^2$$

There is no term Az_i because we are approximate around the Nash price p^{Nash} where $\Pi_i^i = 0$ for all i. The most important coefficients D and E are

$$D = \Pi_{ij}\left(p^{\text{Nash}}\right)$$

$$E = \frac{\Pi_{ii}}{2}\left(p^{\text{Nash}}\right)$$

other contexts, such as the alternating moves model of Maskin and Tirole (1988) and the model with quadratic Rotemberg adjustment costs in Jun and Vives (2004). A recent empirical IO literature, e.g., Brown and MacKay (2021), finds that algorithms allowing for fast repricing do lead to higher markups.

We look for a symmetric equilibrium where each resetting firm j sets

$$z_j^* = \beta Q_j.$$

Then between s and $s + \Delta s$ we have

$$\mathbf{E}_t Q_i(s+\Delta s) = (1-(n-1)\lambda\Delta)\,\mathbf{E}_t Q_i(s) + \lambda\Delta\mathbf{E}_t \sum_{j\neq i}\left[Q_i(s) - p_j(s) + \beta Q_j(s)\right]$$

hence taking the limit $\Delta s \to 0$

$$\frac{d}{ds}\mathbf{E}_t Q_i(s) = \lambda\left\{\beta\sum_{j\neq i}\mathbf{E}_t Q_j(s) - \mathbf{E}_t Q_i(s)\right\}$$

thus the variable $Z(s) = \sum_i \mathbf{E}_t Q_i(s)$ follows

$$\frac{d}{ds}Z(s) = -\lambda\left[1-\beta(n-1)\right]Z(s)$$

Therefore, by symmetry

$$\mathbf{E}_t Q_i(s) = Q_i(t)e^{-\lambda[1-\beta(n-1)](s-t)}$$

When it resets, firm i chooses $z_i^*(t)$ such that

$$\max_{z_i^*(t)} \mathbf{E}_t\left[\int_t^\infty e^{-(\lambda+\rho)(s-t)}\pi^i(z_i^*(t), z_i(t+s))ds\right]$$

The FOC is

$$\begin{aligned}
z_i^*(t) &= -\frac{\int_t^\infty e^{-(\lambda+\rho)(s-t)}D\mathbf{E}_t\left[Q_i(s)\right]ds}{\int_t^\infty e^{-(\lambda+\rho)s}2Eds} \\
&= -\frac{\int_t^\infty e^{-(\lambda+\rho)(s-t)}\left(DQ_i(t)e^{-\lambda(1-(n-1)\beta)(s-t)}\right)ds}{\int_t^\infty e^{-(\lambda+\rho)(s-t)}2Eds} \\
&= -\frac{D(\lambda+\rho)}{2E\left[\lambda+\rho+\lambda(1-(n-1)\beta)\right]}Q_i(t)
\end{aligned}$$

Therefore $B = (n-1)\,\beta$ solves

$$B = \frac{B^{\text{Nash}}}{1+\frac{\lambda}{\rho+\lambda}\left[1-B\right]} \tag{A.30}$$

where the ratio $B^{\text{Nash}} = \frac{(n-1)\Pi_{ij}}{-\Pi_{ii}}$ is the slope of the static best response to a simultaneous price change by all firms $j \neq i$ in a static model. We need B^{Nash} to be strictly lower than 1 for a static symmetric Nash equilibrium to exist. (A.30) shows that the slope of the dynamic naive best response at a stable steady state is always smaller than the slope of the static best response B^{Nash} and is decreasing in λ/ρ. The stable root in $(0,1)$ is

$$B^{\text{Naive}} = \left(\frac{\rho + 2\lambda}{2\lambda}\right)\left[1 - \sqrt{1 - 4\frac{\lambda(\rho+\lambda)}{(\rho+2\lambda)^2}B^{\text{Nash}}}\right].$$

I Derivation of the Oligopolistic Phillips Curve

Consider the general non-stationary versions of the Bellman equation (A.1) and the first-order condition (A.2):

$$(R_t + n\lambda)\, V^i(p,t) = V_t^i(p,t) + \Pi^i(p, MC_t, Z_t) + \lambda \sum_j V^i\left(g^j(p_{-j},t), p_{-j}, t\right) \tag{A.31}$$

$$V_i^i\left(g^i(p_{-i},t), p_{-i}, t\right) = 0 \tag{A.32}$$

Nominal profits are given by

$$\Pi^i(p, MC, Z) = ZD^i(p)\,[p_i - MC]$$

where Z is an aggregate demand shifter that can depend arbitrarily on C_t and P_t.[46]

Define $\alpha(t)$ as the solution to

$$g^i(\alpha(t), \alpha(t), \ldots, \alpha(t), t) = \alpha(t).$$

This is the price that each firm would set if all the firms were resetting at the same time. α is the counterpart of the reset price in the standard New Keynesian model.

To obtain the dynamics of α from (A.31), we start by deriving time-varying envelope conditions evaluated at the symmetric price $p_1 = p_2 = \cdots = p_n = \alpha(t)$. After applying symmetry and using Proposition 9 to make the strategies approximately linear in the neighborhood of the steady state, the non-linear first-order and second-order envelope conditions of the non-stationary game imply the following partial

[46]In Section 3, conditions (5) ensured a constant Z_t.

differential equations (PDEs)

$$0 = V^i_{it} + \Pi^i_i + \lambda\,(n-1)\,V^i_j\beta \tag{A.33a}$$

$$(i_t+\lambda)\,V^i_j = V^i_{jt} + \Pi^i_j + \lambda\,(n-2)\,V^i_j\beta \tag{A.33b}$$

$$(i_t+\lambda)\,V^i_{ii} = V^i_{iit} + \Pi^i_{ii} + \lambda\,(n-1)\left(V^i_{jj}\beta^2 + 2V^i_{ij}\beta\right) \tag{A.33c}$$

$$(i_t+2\lambda)\,V^i_{ij} = V^i_{ijt} + \Pi^i_{ij} + \lambda\,(n-2)\left(V^i_{jj}\beta^2 + V^i_{jk}\beta + \beta V^i_{ij}\right) \tag{A.33d}$$

$$(i_t+\lambda)\,V^i_{jj} = V^i_{jjt} + \Pi^i_{jj} + \lambda\,(n-2)\left(V^i_{jj}\beta^2 + 2\beta V^i_{jk}\right) + \lambda\left(V^i_{ii}\beta^2 + 2\beta V^i_{ij}\right) \tag{A.33e}$$

$$(i_t+2\lambda)\,V^i_{jk} = V^i_{jkt} + \Pi^i_{jk} + \lambda\,(n-3)\left(V^i_{jj}\beta^2 + 2\beta V^i_{jk}\right) + \lambda\left(V^i_{ii}\beta^2 + 2\beta V^i_{ij}\right) \tag{A.33f}$$

Denote the functions

$$W^i_i(t) = V^i_i\left(\alpha(t),\ldots,\alpha(t),t\right),\; W^i_{ii}(t) = V^i_{ii}\left(\alpha(t),\ldots,\alpha(t),t\right)$$

and so on for all derivatives of the value function V^i. We can transform the system (A.33) into a system of ordinary differential equations in the functions $W^i_i(t), W^i_j(t)$, and so on. The partial derivatives with respect to time such as

$$V^i_{it} = \frac{\partial V^i_i}{\partial t}\left(\alpha(t),\ldots,\alpha(t),t\right)$$

in equations (A.33) can be mapped to corresponding total derivatives of W functions

$\dot{W}^i_{it} = \frac{dW^i_{it}}{dt}$ using

$$V^i_{it} = \dot{W}^i_i - \left[V^i_{ii} + \sum_{j \neq i} V^i_{ij} \right] \dot{\alpha}$$

$$V^i_{jt} = \dot{W}^i_j - \left[V^i_{ij} + V^i_{jj} + \sum_{k \neq i,j} V^i_{jk} \right] \dot{\alpha}$$

$$V^i_{iit} = \dot{W}^i_{ii} - \left[V^i_{iii} + \sum_{j \neq i} V^i_{iij} \right] \dot{\alpha}$$

$$V^i_{ijt} = \dot{W}^i_{ij} - \left[V^i_{iij} + V^i_{ijj} + \sum_{k \neq i,j} V^i_{ijk} \right] \dot{\alpha}$$

$$V^i_{jjt} = \dot{W}^i_{jj} - \left[V^i_{ijj} + V^i_{jjj} + \sum_{k \neq i,j} V^i_{jjk} \right] \dot{\alpha}$$

$$V^i_{jkt} = \dot{W}^i_{jk} - \left[V^i_{ijk} + V^i_{jjk} + V^i_{jkk} + \sum_{l \neq i,j,k} V^i_{jkl} \right] \dot{\alpha}$$

where the third derivatives of V at the steady state come from the third-order envelope conditions of the stationary game, solving the linear system:

$$\begin{aligned}
(\rho + \lambda) V^i_{iii} &= \Pi^i_{iii} + \lambda (n-1) \left\{ V^i_{jjj}\beta^3 + 3V^i_{ijj}\beta^2 + 3V^i_{iij}\beta \right\} \\
(\rho + 2\lambda) V^i_{iij} &= \Pi^i_{iij} + \lambda (n-2) \left\{ V^i_{jjj}\beta^3 + 2V^i_{ijj}\beta^2 + V^i_{jjk}\beta^2 + 2V^i_{ijk}\beta + V^i_{iij}\beta \right\} \\
(\rho + 2\lambda) V^i_{ijj} &= \Pi^i_{ijj} + \lambda (n-2) \left\{ V^i_{jjj}\beta^3 + 2\beta^2 V^i_{jjk} + \beta^2 V^i_{ijj} + 2\beta V^i_{ijk} + \beta V^i_{jjk} \right\} \\
(\rho + 3\lambda) V^i_{ijk} &= \Pi^i_{ijk} + \lambda (n-3) \left\{ V^i_{jjj}\beta^3 + 2\beta^2 V^i_{jjk} + \beta^2 V^i_{ijj} + 2\beta V^i_{ijk} + \beta V^i_{jkl} \right\} \\
(\rho + \lambda) V^i_{jjj} &= \Pi^i_{jjj} + \lambda (n-2) \left\{ \beta^3 V^i_{jjj} + 3\beta^2 V^i_{jjk} + 3\beta V^i_{jjk} \right\} \\
&\quad + \lambda \left\{ \beta^3 V^i_{iii} + 3\beta^2 V^i_{iij} + 3\beta V^i_{ijj} \right\} \\
(\rho + 2\lambda) V^i_{jjk} &= \Pi^i_{jjk} + \lambda (n-3) \left\{ \beta^3 V^i_{jjj} + 3\beta^2 V^i_{jjk} + \beta V^i_{jjk} + 2\beta V^i_{jkl} \right\} \\
&\quad + \lambda \left\{ \beta^3 V^i_{iii} + 3\beta^2 V^i_{iij} + \beta V^i_{ijj} + 2\beta V^i_{ijk} \right\} \\
(\rho + 3\lambda) V^i_{jkl} &= \Pi^i_{jkl} + \lambda (n-4) \left\{ \beta^3 V^i_{jjj} + 3\beta^2 V^i_{jjk} + 3\beta V^i_{jkl} \right\} \\
&\quad + \lambda \left\{ \beta^3 V^i_{iii} + 3\beta^2 V^i_{iij} + 3\beta V^i_{ijk} \right\}
\end{aligned}$$

Importantly, to approximate the second derivatives of V^i, we need to solve for the third derivatives of V^i around the steady state by applying the envelope theorem one more time.

Imposing symmetry again, the following non-linear system of ODEs in the functions $\left(\alpha, \beta, W^i_j, W^i_{ii}, W^i_{ij}, W^i_{jj}, W^i_{jk}\right)$ holds exactly (omitting the time argument):

$$0 = -\left[W^i_{ii} + (n-1)\, W^i_{ij}\right] \dot{\alpha} + \Pi^i_i + \lambda\, (n-1)\, W^i_j \beta \tag{A.35a}$$

$$(\imath_t + \lambda)\, W^i_j = \dot{W}^i_j - \left[W^i_{ij} + W^i_{jj} + (n-2)\, W^i_{jk}\right] \dot{\alpha} + \Pi^i_j + \lambda\, (n-2)\, W^i_j \beta \tag{A.35b}$$

$$0 = W^i_{ii} \beta + W^i_{ij} \tag{A.35c}$$

$$(\imath_t + \lambda)\, W^i_{ii} = \dot{W}^i_{ii} - \left[V^i_{iii} + (n-1)\, V^i_{iij}\right] \dot{\alpha} + \Pi^i_{ii} + \lambda\, (n-1) \left(W^i_{jj} \beta^2 + 2 W^i_{ij} \beta\right) \tag{A.35d}$$

$$(\imath_t + 2\lambda)\, W^i_{ij} = \dot{W}^i_{ij} - \left[V^i_{iij} + V^i_{ijj} + (n-2)\, V^i_{ijk}\right] \dot{\alpha} + \Pi^i_{ij} + \lambda\, (n-2) \left(W^i_{jj} \beta^2 + W^i_{jk} \beta + W^i_{ij} \beta\right) \tag{A.35e}$$

$$(\imath_t + \lambda)\, W^i_{jj} = \dot{W}^i_{jj} - \left[V^i_{ijj} + V^i_{jjj} + (n-2)\, V^i_{jjk}\right] \dot{\alpha} + \Pi^i_{jj} + \lambda\, (n-2) \left(W^i_{jj} \beta^2 + 2\beta W^i_{jk}\right) + \lambda \left(W^i_{ii} \beta^2 + 2\beta W^i_{ij}\right) \tag{A.35f}$$

$$(\imath_t + 2\lambda)\, W^i_{jk} = \dot{W}^i_{jk} - \left[V^i_{ijk} + V^i_{jjk} + V^i_{jkk} + (n-3)\, V^i_{jkl}\right] \dot{\alpha} + \Pi^i_{jk} + \lambda\, (n-3) \left(W^i_{jj} \beta^2 + 2\beta W^i_{jk}\right) + \lambda \left(W^i_{ii} \beta^2 + 2\beta W^i_{ij}\right) \tag{A.35g}$$

Next, we linearize system (A.35) around a symmetric steady state $\bar{\alpha} = \alpha(\infty)$ with zero inflation (and steady state values of aggregate variables $\bar{C}, \bar{P}$). Let lower case variables denote log-deviations, e.g., $a(t) = \log \alpha(t) - \log \bar{\alpha}$, and write nominal marginal cost as

$$p(t) + k(t)$$

where $k(t)$ is the log-deviation of the real marginal cost. Profit derivatives such as $\Pi^i_i(t)$ in (A.35) are evaluated at the moving price $\alpha(t)$, hence become once linearized[47]

$$\pi^i_i(t) = \bar{\alpha} \left[\Pi^i_{ii} + (n-1)\, \Pi^i_{ij}\right] a(t) + \bar{M}C\Pi^i_{i,MC} \left(p(t) + k(t)\right) + \Pi^i_i \left(z_c c(t) + z_p p(t)\right)$$
$$\pi^i_j(t) = \bar{\alpha} \left[\Pi^i_{ij} + \Pi^i_{jj} + (n-2)\, \Pi^i_{jk}\right] a(t) + \bar{M}C\Pi^i_{j,MC} \left(p(t) + k(t)\right) + \Pi^i_j \left(z_c c(t) + z_p p(t)\right)$$
$$\pi^i_{ii}(t) = \bar{\alpha} \left[\Pi^i_{iii} + (n-1)\, \Pi^i_{iij}\right] a(t) + \bar{M}C\Pi^i_{ii,MC} \left(p(t) + k(t)\right) + \Pi^i_{ii} \left(z_c c(t) + z_p p(t)\right)$$
$$\pi^i_{ij}(t) = \bar{\alpha} \left[\Pi^i_{iij} + \Pi^i_{ijj} + (n-2)\, \Pi^i_{ijk}\right] a(t) + \bar{M}C\Pi^i_{ij,MC} \left(p(t) + k(t)\right) + \Pi^i_{ij} \left(z_c c(t) + z_p p(t)\right)$$
$$\pi^i_{jj}(t) = \bar{\alpha} \left[\Pi^i_{ijj} + \Pi^i_{jjj} + (n-2)\, \Pi^i_{jjk}\right] a(t) + \bar{M}C\Pi^i_{jj,MC} \left(p(t) + k(t)\right) + \Pi^i_{jj} \left(z_c c(t) + z_p p(t)\right)$$
$$\pi^i_{jk}(t) = \bar{\alpha} \left[\Pi^i_{ijk} + 2\Pi^i_{jjk} + (n-3)\, \Pi^i_{jkl}\right] a(t) + \bar{M}C\Pi^i_{jk,MC} \left(p(t) + k(t)\right) + \Pi^i_{jk} \left(z_c c(t) + z_p p(t)\right)$$

where $\bar{\Pi}^i_i, \bar{\Pi}^i_{ii}$ etc. denote steady state values.

[47] It is more convenient to linearize and not log-linearize profit derivatives, but we use the notation $\pi^i_i(t) = \Pi^i_i(t) - \bar{\Pi}^i_i$.

This yields the system of 6 linear ODEs in $\left(a(t), w^i_j(t), w^i_{ii}(t), w^i_{ij}(t), w^i_{jj}(t), w^i_{jk}(t)\right)$

$$
\begin{aligned}
\left[V^i_{ii} + (n-1)\,V^i_{ij}\right]\dot{a}(t) &= \frac{1}{\bar{\alpha}}\pi^i_i(t) + \lambda\,(n-1)\,\frac{V^i_j\beta}{\bar{\alpha}}\left[w^i_j(t) + b(t)\right] \\
(\rho+\lambda)\,w^i_j(t) + R_t - \rho &= \dot{w}^i_j(t) - \bar{\alpha}\left[\frac{V^i_{ij} + V^i_{jj} + (n-2)\,V^i_{jk}}{V^i_j}\right]\dot{a}(t) + \frac{1}{V^i_j}\pi^i_j(t) + \lambda\,(n-2)\,\beta\left[w^i_j(t) + b(t)\right] \\
(\rho+\lambda)\,w^i_{ii}(t) + R_t - \rho &= \dot{w}^i_{ii}(t) - \frac{\bar{\alpha}}{V^i_{ii}}\left[V^i_{iii} + (n-1)\,V^i_{iij}\right]\dot{a}(t) + \frac{1}{V^i_{ii}}\pi^i_{ii}(t) \\
&\quad + \lambda\,(n-1)\left\{\frac{V^i_{jj}\beta^2}{V^i_{ii}}\left[w^i_{jj}(t) + 2b(t)\right] + \frac{2V^i_{ij}\beta}{V^i_{ii}}\left[w^i_{ij}(t) + b(t)\right]\right\} \\
(\rho+2\lambda)\,w^i_{ij}(t) + R_t - \rho &= \dot{w}^i_{ij}(t) - \frac{\bar{\alpha}}{V^i_{ij}}\left[V^i_{iij} + V^i_{ijj} + (n-2)\,V^i_{ijk}\right]\dot{a}(t) + \frac{1}{V^i_{ij}}\pi^i_{ij}(t) \\
&\quad + \lambda\,(n-2)\left\{\frac{V^i_{jj}\beta^2}{V^i_{ij}}\left[w^i_{jj}(t) + 2b(t)\right] + \frac{V^i_{jk}\beta}{V^i_{ij}}\left[w^i_{jk}(t) + b(t)\right] + \beta\left[w^i_{ij}(t) + b(t)\right]\right\} \\
(\rho+\lambda)\,w^i_{jj}(t) + R_t - \rho &= \dot{w}^i_{jj} - \frac{\bar{\alpha}}{V^i_{jj}}\left[V^i_{ijj} + V^i_{jjj} + (n-2)\,V^i_{jjk}\right]\dot{a}(t) + \frac{1}{V^i_{jj}}\pi^i_{jj}(t) \\
&\quad + \lambda\,(n-2)\left\{\frac{V^i_{jj}\beta^2}{V^i_{jj}}\left[w^i_{jj}(t) + 2b(t)\right] + \frac{2V^i_{jk}\beta}{V^i_{jj}}\left[w^i_{jk}(t) + b(t)\right]\right\} \\
&\quad + \lambda\left\{\frac{V^i_{ii}\beta^2}{V^i_{jj}}\left[w^i_{ii}(t) + 2b(t)\right] + \frac{2V^i_{ij}\beta}{V^i_{jj}}\left[w^i_{ij}(t) + b(t)\right]\right\} \\
(\rho+2\lambda)\,w^i_{jk}(t) + R_t - \rho &= \dot{w}^i_{jk} - \frac{\bar{\alpha}}{V^i_{jk}}\left[V^i_{ijk} + V^i_{jjk} + V^i_{jkk} + (n-3)\,V^i_{jkl}\right]\dot{a}(t) + \frac{1}{V^i_{jk}}\pi^i_{jk}(t) \\
&\quad + \lambda\,(n-3)\left\{\frac{V^i_{jj}\beta^2}{V^i_{jk}}\left[w^i_{jj}(t) + 2b(t)\right] + \frac{2V^i_{jk}\beta}{V^i_{jk}}\left[w^i_{jk}(t) + b(t)\right]\right\} \\
&\quad + \lambda\left\{\frac{V^i_{ii}\beta^2}{V^i_{jk}}\left[w^i_{ii}(t) + 2b(t)\right] + \frac{2V^i_{ij}\beta}{V^i_{jk}}\left[w^i_{ij}(t) + b(t)\right]\right\}
\end{aligned}
$$

In general there are thus 6 ODEs because β may be time-dependent hence $b(t) \neq 0$. But note that if $b(t) = 0$ then the system becomes block-recursive and we can solve separately the first two equations in a and w^i_j. From the optimality conditions we have

$$\dot{\beta} = -\dot{\alpha}\left[W^i_{iij}\left[1-(n-1)\,\beta\right] + (n-1)\,W^i_{ijj} - \beta W_{iii}\right]$$

Using our perturbation argument we can show that there exists a third-order cross-elasticity ϵ^i_{iij} such that at the steady state

$$V^i_{iij}\left[1-(n-1)\,\beta\right] + (n-1)\,V^i_{ijj} - \beta V_{iii} = 0 \tag{A.36}$$

where $V_{iij}, V_{ijj}, V_{iii}$ are solutions to the system (A.34). Thus in what follows we consider β as constant for the first-order dynamics to simplify expressions, although we could solve the larger system without this assumption.

The last step is to replace the single "reset price" variable $a(t)$ with two variables, the aggregate price level $p(t)$ and inflation $\pi(t) = \dot{p}(t)$ using our aggregation result that inflation follows

$$\pi(t) = \lambda \left[1 - (n-1)\,\beta(t)\right] \left[\log \alpha(t) - \log P(t)\right].$$

After log-linearization we have

$$a(t) = \frac{\pi(t)}{\lambda \left[1 - (n-1)\,\beta\right]} + p(t).$$

Therefore, we obtain in matrix form that the vector

$$\mathbf{Y}(t) = \left(\pi(t), p(t), w_j^i(t)\right)'$$

solves the linear differential equation

$$\dot{\mathbf{Y}}(t) = \mathbf{A}\mathbf{Y}(t) + \mathbf{Z}_k k(t) + \mathbf{Z}_c c(t) + \mathbf{Z}_R \left[R(t) - \rho\right]$$

where $\mathbf{A} \in \mathbb{R}^{3\times 3}$, $\mathbf{Z}_k, \mathbf{Z}_c, \mathbf{Z}_R \in \mathbb{R}^3$ collect the terms above (evaluated at the steady state), with boundary conditions $\lim_{t\to\infty} \mathbf{Y}(t) = 0$. The solution is given by

$$\mathbf{Y}(t) = -\int_0^\infty e^{s\mathbf{A}} \left\{\mathbf{Z}_k k\left(t+s\right) + \mathbf{Z}_c c\left(t+s\right) + \mathbf{Z}_R \left[R\left(t+s\right) - \rho\right]\right\} ds$$

where $e^{s\mathbf{A}} = \sum_{k=0}^{\infty} \frac{s^k \mathbf{A}^k}{k!}$ denotes the matrix exponential of $s\mathbf{A}$. Proposition 8 then follows by taking the first component of $\mathbf{Y}$.

To obtain the scalar higher-order ODE for π, let $[\mathbf{M}]_i$ and $[\mathbf{M}]_{xy}$ denote the ith line and the (x,y) element of a generic matrix $\mathbf{M}$ respectively. Let $\mathbf{B}(t) = \mathbf{Z}_k k(t) + \mathbf{Z}_c c(t) + \mathbf{Z}_r \left[r(t) - \rho\right]$. Iterating $\dot{\mathbf{Y}}(t) = \mathbf{A}\mathbf{Y}(t) + \mathbf{B}(t)$, we have for all $k \geq 1$

$$\mathbf{Y}^{(k)}(t) = \mathbf{A}^k \mathbf{Y}(t) + \sum_{j=0}^{k-1} \mathbf{A}^j \mathbf{B}^{(k-1-j)}(t).$$

Taking the first line for each $k = 1, \ldots, K = 3$, we have K equations

$$\frac{d^k \pi(t)}{dt^k} - \left[\sum_{j=0}^{k-1} \mathbf{A}^j \mathbf{B}^{(k-1-j)}(t)\right]_1 = \left[\mathbf{A}^k\right]_1 \mathbf{Y}(t)$$

which we can each rewrite as

$$\frac{d^k \pi(t)}{dt^k} - \left[\sum_{j=0}^{k-1} \mathbf{A}^j \mathbf{B}^{(k-1-j)}(t)\right]_1 - \left[\mathbf{A}^k\right]_{11} \pi(t) = \sum_{i=2}^{K} \left[\mathbf{A}^k\right]_{1i} y_i(t)$$

Let

$$\mathbf{M} = \begin{pmatrix} \mathbf{A}_{12} & \dots & \mathbf{A}_{1n} \\ \left[\mathbf{A}^2\right]_{12} & & \left[\mathbf{A}^2\right]_{1n} \\ \vdots & & \vdots \\ \left[\mathbf{A}^n\right]_{12} & \dots & \left[\mathbf{A}^n\right]_{1n} \end{pmatrix} \in \mathbb{R}^{K\times(K-1)}$$

Take any vector $\alpha^\pi = \left(\alpha^\pi_j\right)_{j=1}^{K}$ in $\ker \mathbf{M}'$ (whose dimension is at least 1), i.e., such that $\mathbf{M}'\gamma^\pi = 0 \in \mathbb{R}^{K-1}$. Then

$$\sum_{k=1}^{K} \alpha^\pi_k \left(\frac{d^k \pi(t)}{dt^k} - \left[\sum_{j=0}^{k-1} \mathbf{A}^j \mathbf{B}^{(k-1-j)}(t)\right]_1 - \left[\mathbf{A}^k\right]_{11} \pi(t) \right) = 0.$$

and we can define $\alpha^\pi_0 = -\sum_{k=1}^{K} \alpha^\pi_k \left[\mathbf{A}^k\right]_{11}$. This simplifies to

$$\begin{aligned} \dddot{\pi} &= \left(\mathbf{A}_{\pi\pi} + \mathbf{A}_{ww}\right) \ddot{\pi} \\ &+ \left(\mathbf{A}_{\pi p} + \mathbf{A}_{\pi w}\mathbf{A}_{w\pi} - \mathbf{A}_{\pi\pi}\mathbf{A}_{ww}\right) \dot{\pi} \\ &+ \left(\mathbf{A}_{\pi w}\mathbf{A}_{wp} - \mathbf{A}_{\pi p}\mathbf{A}_{ww}\right) \pi \\ &+ \mathbf{A}_{\pi w}\dot{\mathbf{B}}_w + \ddot{\mathbf{B}}_\pi - \mathbf{A}_{ww}\dot{\mathbf{B}}_\pi \end{aligned} \tag{A.37}$$

I.1 One-time shocks

Given (19) we can guess and verify that $x = \psi_x e^{-\xi t}$ for all variables $x \in \{\pi, k, c, R-\rho\}$ and solve for the coefficients ψ_x using the system

$$\begin{aligned} \psi_\pi \left(\gamma^\pi_0 - \gamma^\pi_1\xi + \gamma^\pi_2\xi^2 - \gamma^\pi_3\xi^3\right) &= \psi_k \left(\gamma^k_0 - \gamma^k_1\xi + \gamma^k_2\xi^2\right) \\ &+ \psi_c \left(\gamma^c_0 - \gamma^c_1\xi + \gamma^c_2\xi^2\right) \\ &+ \left(\psi_R - \psi_\pi\right) \left(\gamma^r_0 - \gamma^r_1\xi + \gamma^r_2\xi^2\right) \\ -\xi\psi_c &= \sigma^{-1}\left(\psi_R - \psi_\pi - \epsilon^r_0\right) \\ \psi_R &= \phi_\pi\psi_\pi + \epsilon^m_0 + (1-\kappa)\,\epsilon^r_0 \end{aligned}$$

Thus

$$\begin{aligned}
\psi_c &= \frac{1}{\sigma\xi}\left(\psi_\pi\left(1-\phi_\pi\right)+\kappa\epsilon_0^r-\epsilon_0^m\right)\\
\psi_k &= \psi_c\left(\chi+\sigma\right)
\end{aligned}$$

and

$$\begin{aligned}
\psi_\pi\left(\gamma_0^\pi-\gamma_1^\pi\xi+\gamma_2^\pi\xi^2-\gamma_3^\pi\xi^3\right) = \frac{1}{\sigma\xi}\left(\kappa\epsilon_0^r-\epsilon_0^m-\psi_\pi\left(\phi_\pi-1\right)\right)&\left[\left(\chi+\sigma\right)\left(\gamma_0^k-\gamma_1^k\xi+\gamma_2^k\xi^2\right)+\left(\gamma_0^c-\gamma_1^c\xi+\gamma_2^c\xi^2\right)\right]\\
&+\left(\epsilon_0^m+\left(1-\kappa\right)\epsilon_0^r+\psi_\pi\left(\phi_\pi-1\right)\right)\left(\gamma_0^r-\gamma_1^r\xi+\gamma_2^r\xi^2\right)
\end{aligned}$$

which yields

$$\psi_\pi = \frac{\frac{\kappa\epsilon_0^r-\epsilon_0^m}{\sigma\xi}\left[\left(\chi+\sigma\right)\left(\gamma_0^k-\gamma_1^k\xi+\gamma_2^k\xi^2\right)+\left(\gamma_0^c-\gamma_1^c\xi+\gamma_2^c\xi^2\right)\right]+\left(\epsilon_0^m+\left(1-\kappa\right)\epsilon_0^r\right)\left(\gamma_0^r-\gamma_1^r\xi+\gamma_2^r\xi^2\right)}{\gamma_0^\pi-\gamma_1^\pi\xi+\gamma_2^\pi\xi^2-\gamma_3^\pi\xi^3+\left(\phi_\pi-1\right)\left[\frac{\left(\chi+\sigma\right)\left(\gamma_0^k-\gamma_1^k\xi+\gamma_2^k\xi^2\right)+\left(\gamma_0^c-\gamma_1^c\xi+\gamma_2^c\xi^2\right)}{\sigma\xi}-\left(\gamma_0^r-\gamma_1^r\xi+\gamma_2^r\xi^2\right)\right]}$$

J Additional Figures

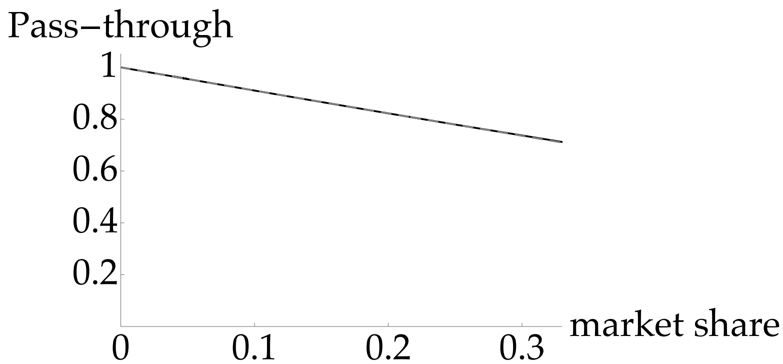

Figure A.6: Pass-through $\hat{\alpha}$ as a function of market share: symmetric firms (solid black line) vs. heterogeneous firms (dashed gray line). The two lines lie almost exactly on top of each other.

Note: Black line: market share varies through the number $n = 2, 3, \ldots$ of symmetric firms (black). Gray dashed line: market share varies through heterogeneity in productivity among a fixed number $n = 10$ of firms. Nested CES preferences with $\eta = 10$, $\omega = 1$.

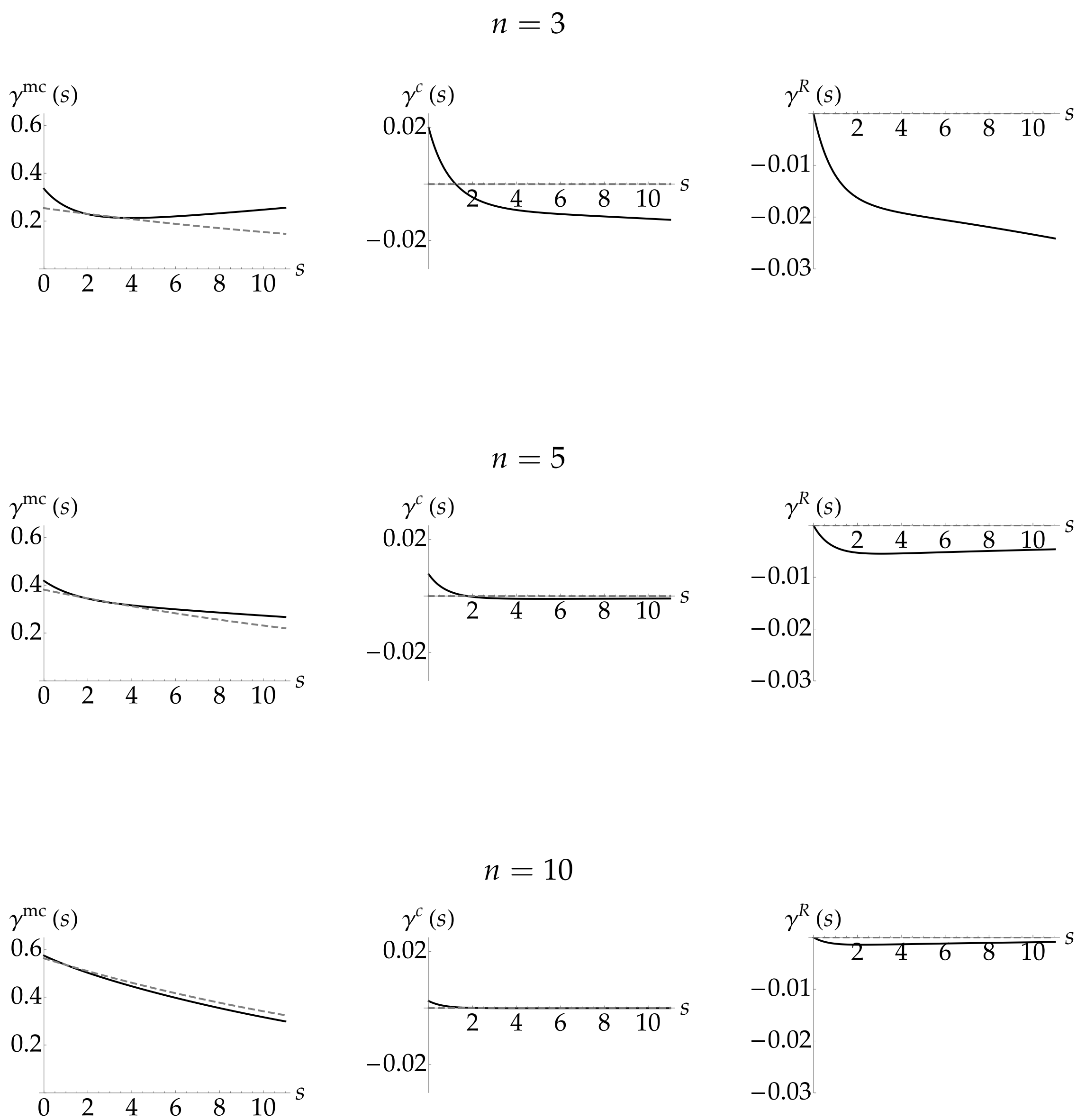

Figure A.7: Green functions $\gamma^{mc}(s), \gamma^{c}(s), \gamma^{R}(s)$ for different numbers of firms n.

Note: AIK calibration. Solid black line: Strategic oligopoly. Dashed gray line: Naive model.

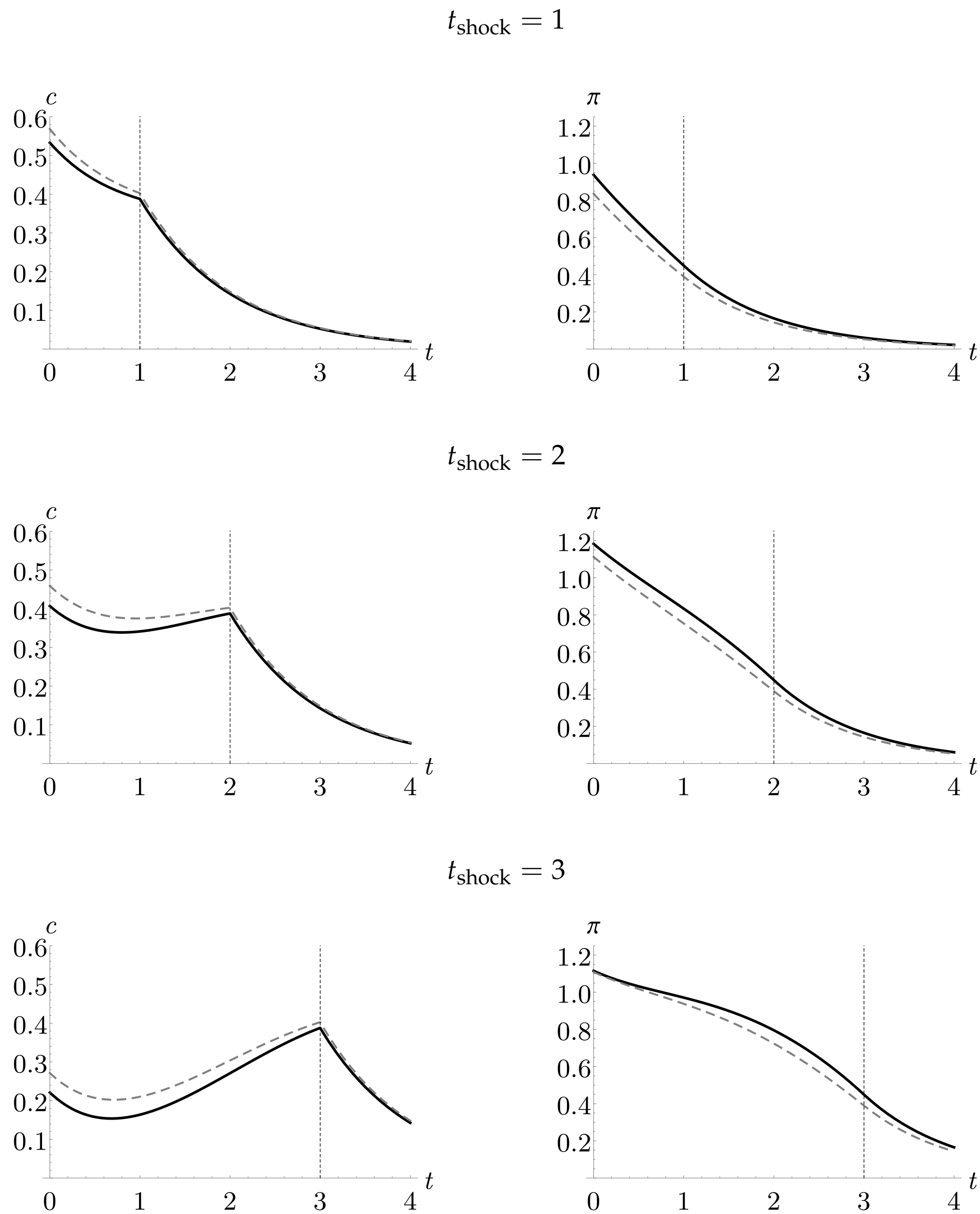

Figure A.8: Impulse responses for consumption and inflation following date-0 news about monetary policy shock happening at t_{shock} indicated by the vertical line.

Note: $n = 3$ firms with AIK calibration. Solid black line: Strategic oligopoly. Dashed gray line: Naive model. c and π denote log-deviations from steady state values in %.

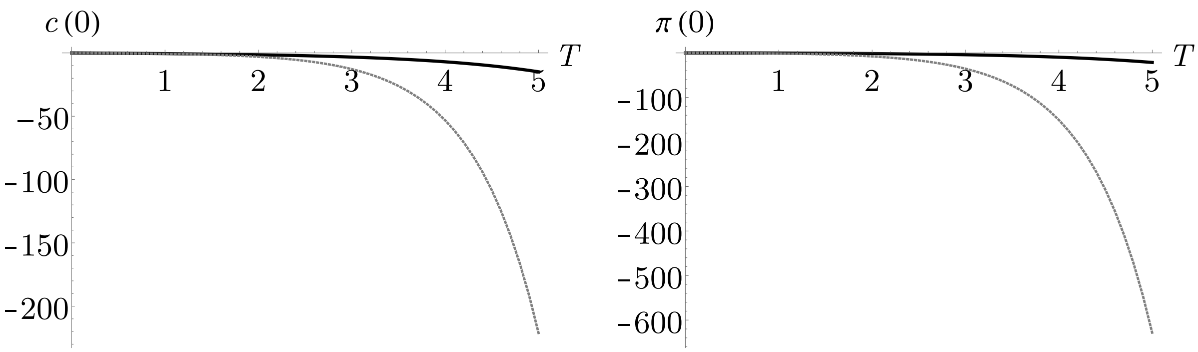

Figure A.9: Date-0 consumption and inflation in a liquidity trap lasting from $t = 0$ to $t = T$, for different values of T.

Note: Solid black line: $n = 3$. Dotted gray line: $n = \infty$. c and π denote log-deviations from steady state values in %.

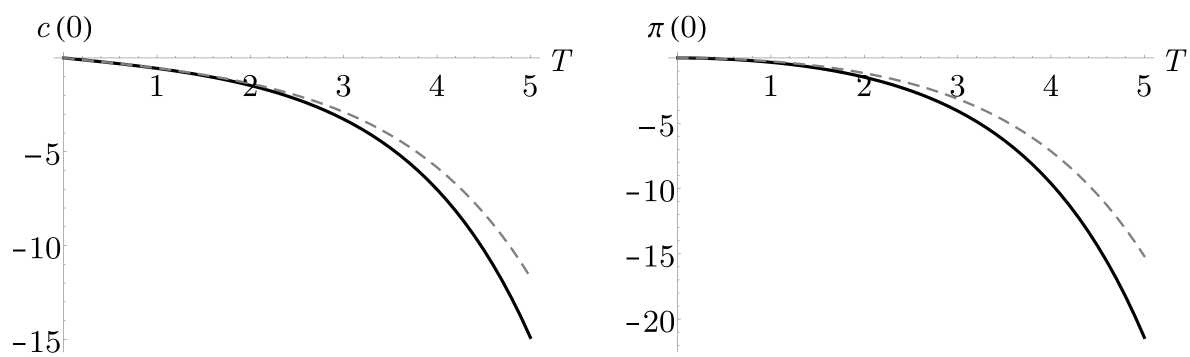

Figure A.10: Date-0 consumption and inflation in a liquidity trap lasting from $t = 0$ to $t = T$, for different values of T.

Note: $n = 3$ firms with AIK calibration. Solid black line: Strategic oligopoly. Dashed gray line: Naive model. c and π denote log-deviations from steady state values in %.

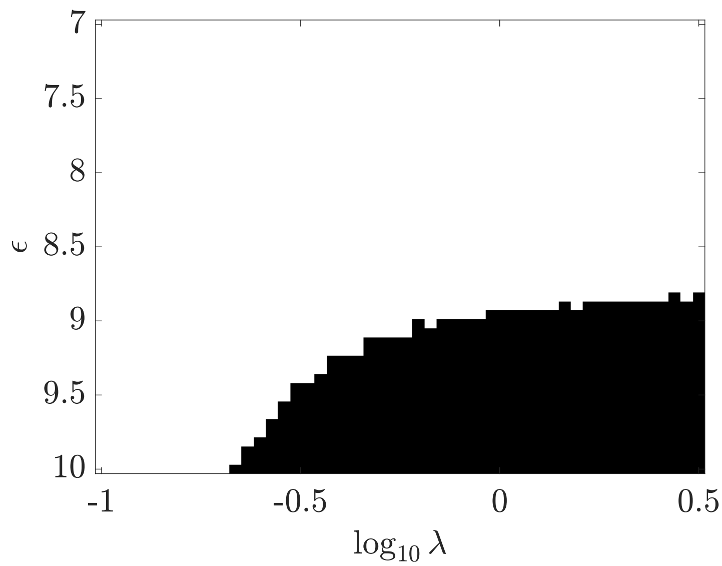

Figure A.11: In white: convergence of value function iteration algorithm towards a monotone MPE in (λ, ϵ) space, with $n = 2$ firms.