Dynamic Oligopoly and Price Stickiness*

Olivier Wang — NYU Stern
Iván Werning — MIT

September 2021

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 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 firms and modified consumer preferences that ensure firms face comparable residual demands. Finally, the Phillips curve for our model displays inflation persistence and endogenous cost-push shocks.

*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 and Juliette Fournier, 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 and Wisconsin. All remaining errors are ours.

1 Introduction

The recent rise in product-market concentration in the U.S. has been viewed as a driving force behind several macroeconomic trends. For instance, Gutiérrez and Philippon (2017) document an increase in the mean Herfindahl-Hirschman 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.[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 is not designed to address these questions. Following the recognition that some form of imperfect competition and pricing power is required to model nominal rigidities, the New Keynesian literature has been built on the tractable paradigm of monopolistic competition, pervasive in other areas of macroeconomics and international trade. Under monopolistic competition, markups only depend on tastes, through consumers' elasticity of substitution between competing goods, which leaves no room for changes in concentration to affect markups or monetary policy.

In this paper, we provide a new framework to study the link between market structure and monetary policy. We generalize the standard New Keynesian model by allowing for dynamic oligopolistic competition between any finite number of firms in each sector of the economy, also allowing for heterogeneity across and within sectors. In each sector, firms compete by setting their prices, but they do so in a staggered and infrequent manner due to nominal rigidities. We use this model to study the aggregate real effects of monetary shocks and highlight the restrictions imposed by monopolistic competition. Departing from monopolistic competition to study oligopoly poses new challenges, because it requires solving a dynamic game with strategic interactions at the sectoral level and embedding it into a general equilibrium macroeconomic model. We focus on 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.

[1] Rossi-Hansberg, Sarte and Trachter (2020) document, however, diverging trends in national and local measures of concentration. We will discuss how to interpret our results in light of these two views.

Despite these complexities, our first result derives a closed-form formula for the response of aggregate output to small monetary shocks. Our formula inputs the cross-sectoral distribution of three sufficient statistics: market concentration as captured by the effective number of firms within a sector, demand elasticities, and markups. The intuition is based on the link between the steady state markup that can be sustained in an oligopolistic equilibrium and the slope of the reaction function of each firm to the prices of its competitors. All else equal, steeper reaction functions lead to higher equilibrium markups: each firm has little incentives to cut prices if doing so would lead competing firms to cut prices as well for some time. Inverting the logic, we can infer from high observed markups that reaction functions are steep and therefore complementarities in pricing are strong, which in turn implies a slow pass-through of monetary shocks into prices and therefore large real effects on output. In this way, our formula encapsulates a tight restriction between endogenous markups and stickiness, conditional on demand elasticities.[2]

While our key sufficient statistics, demand elasticities and markups, can be estimated at any given point in time, they are endogenous objects that change in reaction to shifts in fundamentals. To perform counterfactual experiments, we take a more structural approach and solve numerically the oligopolistic equilibrium in terms of fundamentals. We use a flexible Kimball (1995) demand system that allows us to parametrize separately demand elasticities and superelasticities, as the latter can affect monetary policy transmission through variable markups even under monopolistic competition.

In our main exercise, we vary the number of firms n in each sector while keeping preference parameters fixed. We find that higher concentration (lower n) can significantly amplify or dampen the real effects of monetary policy, depending on how properties of the residual demand vary with n. When preferences are CES, higher concentration amplifies monetary policy transmission. Maximal effects are attained under duopoly, for which the half-life of the price level in reaction to monetary shocks is around 40% higher than under monopolistic competition. With Kimball preferences, higher concentration actually dampens monetary policy transmission if the superelasticity of demand is high enough, and the dampening can be arbitrary large. It is thus essential to first understand the link between concentration

[2]In the standard monopolistic competition model desired markups are constant and only a function of the demand elasticity. However, in a strategic environment the endogenous markup is no longer a simple function of the demand elasticity.

and demand functions.

We use evidence on the heterogeneity in idiosyncratic cost pass-through across small and large firms from Amiti, Itskhoki and Konings (2019) to calibrate how concentration affects the shape of demand functions, and find substantial amplification. The rise in the average Herfindahl index observed in the U.S. since 1990 increases the response of output (and decreases the response of inflation) to monetary shocks by around 15%.

What explains these results? The number of competitors in a market has an effect on firms' strategic incentives, but also on the residual demand faced by each firm. We disentangle these two ways through which oligopolistic competition differs from monopolistic competition. On the one hand, "feedback effects" make each firm care about its rivals' current and future prices when setting its price. 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 the number of firms is finite. To isolate these two effects for each n, we compare the oligopolistic model with n firms to a "non-strategic" scenario in which the n firms act myopically, ignoring the effect of their prices on competitors' future incentives. The non-strategic equilibrium is isomorphic to a model with monopolistic competition and Kimball preferences modified to match the elasticity and superelasticity of the residual demand in the finite n model. We find that departures from monopolistic competition are mostly working through feedback effects, that is changes in the shape of residual demand. 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 can only be reached after solving the full, strategic, model.

Standard models of monopolistic competition featuring non-CES demand can thus be viewed as simple approximations to an oligopolistic reality, and our framework provides a rigorous mapping from micro-evidence on pass-through and concentration, that may change over time and with competition policy, to the reduced-form Kimball parameter driving these models. It does not follow, however, that oligopoly is isomorphic to monopolistic competition up to some recalibration. 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 residual demand has a positive effect on both markups and the pass-through of monetary policy. Therefore, controlling for concentration and demand elasticity, our model predicts that monetary policy is transmitted relatively more through sectors or regions with higher markups, because they are the ones featuring the slowest price adjustment following monetary shocks.

Moreover, the quantitative near-equivalence between the oligopoly model and the recalibrated non-strategic economy depends on the specific processes for real and monetary shocks. In order to go beyond the permanent money supply shocks most commonly studied in the literature, we derive a three equations New Keynesian model with an oligopolistic Phillips curve that allows for more general shocks and non-stationary dynamics. We find that strategic effects are quantitatively important once we allow for richer dynamics. In particular, the oligopolistic Phillips curve features a form of endogenous inflation persistence (or equivalently, endogenous cost-push shocks) that can dampen fluctuations in inflation and output relative to the non-strategic model.

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.[3] 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

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

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.[4] This allows us to go beyond two firms and explore different questions, in particular by changing industry concentration and separating strategic complementarities from residual demand effects.[5] Modeling more than two firms also lets us incorporate recent evidence on cost pass-through and market shares from Amiti, Itskhoki and Konings (2019) to infer the relation between concentration and monetary non-neutrality. As we show, this evidence implies that even under Calvo pricing, oligopoly leads to significant amplification.

The literature on variable markups in international trade (e.g., Atkeson and Burstein 2008) highlights the importance of market structure and for cost (e.g., exchange rate) pass-through in static settings. We study a dynamic version of these models, as is needed to analyze monetary policy, and show which properties of residual demand functions matter in this context (see also Neiman (2011) for a partial equilibrium dynamic duopoly model of exchange rate pass-through with menu costs). In particular, we use the evidence from Amiti, Itskhoki and Konings (2019) on heterogeneous pass-through behavior across small and large firms to calibrate our oligopolistic model.

Kimball (1995) introduced non-CES aggregators that generate variable markups even under monopolistic competition.[6]As we show in section 7, 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,

[4]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.

[5]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.

[6]Translog preferences achieve the same purpose, as shown by Bergin and Feenstra (2000). We focus on Kimball preferences for concreteness, but our results are non-parametric and thus apply to other non-CES preferences, including translog.

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, while a large literature studies the feedback between the cyclicality of markups and entry and exit dynamics (e.g. Bilbiie, Ghironi and Melitz, 2007).

Our macroeconomic results 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).

2 A Macro Model with Oligopolies

In this section we first describe the economic environment, preferences, technology, and the market structure. We then define an equilibrium.

The household side of our model is standard. On the production side, we depart from the atomistic monopolistic competitive framework in favor of oligopolies, with a finite number of firms, producing differentiated varieties in each sector. These firms compete with each other by setting prices at random intervals of time, resulting in a staggered set of price changes.

Basics. Time is continuous with an infinite horizon $t \in [0, \infty)$.[7] We abstract from aggregate uncertainty. This suffices to study the impact and transitional dynamics induced by an unanticipated shock. Following much of the menu-cost literature,

[7]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).

we focus on such a monetary shock, and our goal is to understand the degree of monetary non-neutrality it induces.

There are three types of economic agents: households, firms and the government. Households are described by a continuum of infinitely lived agents that consumes nondurable goods and supplies labor to a competitive labor market.

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. Otherwise, there are strategic interactions across firms within a sector, but not across sectors (due to the continuum assumption). We study the dynamic game within a sector and focus on Markov equilibria.

The government controls the money supply, provides transfers and issues bonds, to satisfy its budget constraint.

Household Preferences. Utility is given by

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

with real money balances $m(t) = \frac{M(t)}{P(t)}$ and aggregate consumption

$$C(t) = G\left(\{C_s(t)\}_{s \in S}\right),$$
$$C_s(t) = H_s\left(\{c_{i,s}(t)\}_{i \in I_s}\right),$$

where $\{C_s\}$ and $\{c_{i,s}(t)\}$ describe sectoral consumption across sectors $s \in S$ and good varieties across firms $i \in I_s$ within each sector, respectively. G and H_s are aggregator functions homogeneous of degree one.

Following Golosov and Lucas (2007), in most of the paper we adopt the specification

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

As is well known, these preferences help simplify the aggregate equilibrium dynamics; we consider more general preferences in section 8. In addition, we assume

CES aggregation across sectors:[8]

$$G\left(\{C_s\}_{s\in S}\right) = \left(\int_S {C_s}^{1-\frac{1}{\omega}} ds\right)^{\frac{1}{1-\frac{1}{\omega}}}$$

but crucially, H_s can be more general than CES. We make no assumption on H_s beyond homotheticity until section 5.

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

$$y_{i,s}(t) = z_{i,s}\ell_{i,s}(t).$$

We first assume no sectoral or idiosyncratic differences in productivity for simplicity, setting $z_{i,s} = 1$. In section 6 we extend the model to heterogeneous firms.

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.

Individual firm nominal profits are

$$\Pi^{i,s}(t) = p_{i,s}(t)y_{i,s}(t) - W(t)\ell_{i,s}(t)$$

and aggregate firm nominal profits $\Pi(t) = \int \sum_{i\in I_s} \Pi^{i,s}(t)\, ds$. Firms seek to maximize the present value of profits,

$$\mathbb{E}_0 \int_0^\infty Q(t)\Pi^{i,s}(t)dt$$

where $Q(t) = e^{-\int_0^t R(s)ds}$ denotes the nominal price deflator between period t and 0.

Although there is no aggregate uncertainty, the expectation averages over the idiosyncratic uncertainty about the dates at which changes are allowed for each firm and its immediate competitors within a sector. (This firm objective can be justified in a number of ways, such as by introducing an asset market for the stock price of firms.)

[8]When $\omega = 1$ we set $G\left(\{C_s\}_s\right) = \exp \int_S \log C_s ds$.

Household Budget Constraints. The flow budget constraint can be summarized by

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

for all $t \geq 0$, where $B(t)$ are bonds paying nominal interest rate $R(t)$, $M(t)$ nominal money holdings, $W(t)$ the nominal wage, $T(t)$ nominal lump-sum transfers, and $P(t)$ the (ideal) price index given by

$$P(t) = \mathcal{P}(\{p_{i,s}(t)\}),$$

where $\mathcal{P}(\{p_{i,s}\}) \equiv \min_{\{c_{i,s}\}} \int \sum_{i\in I_s} p_{i,s}c_{i,s}ds$ s.t. $G\left(\left\{H_s\left(\{c_{i,s}\}_{i\in I_s}\right)\right\}_{s\in S}\right) = 1$. For $\omega \neq 1$, we can write $\mathcal{P}(\{p_{i,s}\}) \equiv \left(\int P_s^{1-\omega}ds\right)^{\frac{1}{1-\omega}}$ with $P_s = \mathcal{P}_s(p_{1,s}, p_{2,s}, \ldots, p_{n_s,s})$.[9]

Let $A(t) = B(t) + M(t)$ denote total nominal wealth. Households are also subject to the No Ponzi condition $\lim_{t\to\infty} Q(t)A(t) \geq 0$. This leads to the present value condition

$$\int_0^\infty Q(t)(P(t)C(t) + R(t)M(t) - T(t) - W(t)\ell(t) - \Pi(t))dt = A(0) = M(0) + B(0).$$

Demand. Define the vector of prices within a sector s as

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

and let $p_{-i,s}(t) = (p_{1,s}(t), \ldots, p_{i-1,s}(t), p_{i+1,s}(t), \ldots, p_{n,s}(t))$ denote the vector that excludes $p_{i,s}(t)$. The demand for firm $i \in I_s$ can be written as

$$c_{i,s}(t) = D^{i,s}(p_{i,s}(t), p_{-i,s}(t); C(t), P(t)).$$

Given symmetry, constant returns and the CES structure across sectors, we obtain

$$D^{i,s}(p_i, p_{-i}; C, P) = d^{i,s}(p_i, p_{-i})CP^{\omega}.$$

The demand faced by firm i is a stable function of the price vector $d^{i,s}(p_i, p_{-i})$. This demand captures within-sector substitution as well as across-sector substitution. Firms understand that they can switch expenditure in both ways by changing their price.

[9] We have $\mathcal{P}(\{p_{i,s}\}) \equiv \exp \int \log P_s ds$ when $\omega = 1$.

Discounted nominal profits are then

$$\int_0^\infty e^{-\int_0^t R(s)ds} C(t)P(t)^\omega \, d^{i,s}(p_{i,s}(t), p_{-i,s}(t))\,(p_{i,s}(t) - W(t))\,dt$$

Markov Equilibria. 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,

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

Given that sectors are symmetric and firms are symmetric within sectors, we consider strategies $g(p_-, t)$ that are symmetric, except in section 6.2.

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 $d(p_i, p_{-i};t)$ and strategy functions for firms $g(p_{-i};t)$ such that: (a) consumers optimize quantities taking as given the sequence of prices and interest rates; (b) the firm reset price strategy g is optimal, given the path for $P(t), C(t)$ and its rivals' strategies g and demand function of consumers 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, the supply of labor equals aggregate demand for labor

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

and the demand for money equals supply $M(t)$.

3 Stationary Oligopoly Game within a Sector

We first focus on the dynamics within a sector, assuming all conditions external to the sector are fixed and given: the wage, the nominal discount rate, aggregate consumption and price are assumed constant. These assumptions imply that the oligopoly game within an industry is stationary. This partial equilibrium analysis also characterizes a steady state in general equilibrium.

We shall later explore conditions under which we can use the sectoral dynamics we characterize here to study the aggregate macroeconomic adjustment to a monetary shock.

3.1 Prices, Demands and Profits

We now focus within a sector, suppressing the notation conditioning on $s \in S$ we collect prices within the sector in a vector

$$p = (p_1, \ldots, p_n)$$

and let $p_{-i} = (p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n)$ denote competitor prices for firm i. The profit function for firm i is then

$$\Pi^i(p) = d^i(p_i, p_{-i})(p_i - W).$$

Since $R(t) = \rho$ we have $Q(t) = e^{-\rho t}$ and firms maximize

$$\mathbb{E}_0 \int_0^\infty e^{-\rho t} d^i(p_i, p_{-i})(p_i - W) dt.$$

3.2 Markov Equilibria

In a Markov equilibrium firms i follow a strategy specifying the reset price

$$p_i^* = g^i(p_{-i})$$

they will chose in the event that they receive a price change opportunity. Together with an initial price vector and the Poisson arrival rate this fully describes the stochastic dynamics within the sector. We focus on differentiable symmetric Markov equilibria, where

$$g^i = g.$$

Let $V^i(p)$ denote the value function obtained by firm i, where the argument p is a vector of n prices. The Bellman equation is then

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

where Π^i is the profit function of firm i and for each j

$$g^j(p_{-j}) = \arg\max_{p'_j} V^j(p'_j, p_{-j})$$

satisfying the optimality condition

$$V^j_{p_j}(g^j(p_{-j}), p_{-j}) = 0. \tag{2}$$

The right-hand side of (1) states that with Poisson rate λ, one of the firms indexed by $j = 1, \ldots, n$ (including firm i) will adjust its price to $g_j\left(p_{-j}\right)$, which will make firm i's value jump to $V^i\left(g^j\left(p_{-j}\right), p_{-j}\right)$.[10]

Remark 1. There could be multiple equilibria even within the Markov class, but our main results apply for any differentiable selection. The differentiability assumption rules out "kinked demand curve" and "Edgeworth cycles" Markov equilibria studied by Maskin and Tirole (1988) in a Bertrand duopoly model with *perfectly substitutable* goods as firms become infinitely patient, which in our setting is isomorphic to the flexible prices limit $\lambda \to \infty$. Maskin and Tirole (1988) show that firms can "collude" around the joint monopoly price in this limit by using strategies that are nonmonotonic in the rival's price. For less extreme combinations of substitutability and price flexibility, value function iteration converges to a standard MPE with monotonic strategies corresponding to the ones we study locally.[11]

3.3 A Steady State Condition

We now provide a key expression for the slope of the reset price strategy at a steady state. Differentiating the Bellman equation (1) and making use of symmetry, we obtain at the steady state $\bar{p}$ of a symmetric equilibrium:

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

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

[11] Figure A.10 displays the locus of existence of these monotone equilibria in the (λ, ϵ)-space (where ϵ 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.

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

Denote $\frac{\partial g^j}{\partial p_k}(\bar{p}) = \beta$ for all $k \neq j$. Using $\sum_k \sum_{j\neq i,k} V^i_{p_j}(\bar{p}) = (n-2)\sum_{k\neq i} V^i_{p_k}(\bar{p})$, and the symmetry of Π_{p_k} across $k \neq i$, we obtain

$$0 = \Pi^i_{p_i}(\bar{p}) + \frac{\lambda\,(n-1)\,\beta}{\rho+\lambda\,[1-(n-2)\beta]}\Pi^i_{p_k}(\bar{p}) \tag{3}$$

With flexible prices, firms would continuously play the static Nash equilibrium price p^{NE} that solves $0 = \Pi^i_{p_i}(p^{NE})$. If the elasticity of substitution between goods is higher within sectors than across sectors (as in the literature and in all our applications), then $\Pi^i_{p_k} \geq 0$ hence the steady state price $\bar{p}$ of the dynamic oligopoly game is above the static Nash price p^{NE} if and only if $\beta > 0$. As n grows to infinity, the influence of any single rival $\Pi^i_{p_k}$ vanishes and the steady state price converges to the Nash price (i.e., monopolistic competition).

Sufficient Statistics: Markups and Elasticities. The main object of our analysis is the slope $(n-1)\,\beta$ of the reaction function, where the term $n-1$ scales the aggregate effect of the rivals. We can further simplify (3) to write $(n-1)\,\beta$ in terms of observable sufficient statistics. Rewrite (3) in terms of demand own-elasticity $\epsilon^i_i = \frac{\partial \log d^i}{\partial \log p_i}$ and cross-elasticity $\epsilon^i_j = \frac{\partial \log d^i}{\partial \log p_j}$:

$$(n-1)\,\beta = \frac{\rho+\lambda}{\lambda}\frac{1}{\frac{n-2}{n-1} + \frac{\epsilon^i_j}{-\epsilon^i_i - \frac{\bar{p}}{\bar{p}-W}}}.$$

Constant returns to scale imply that the cross-elasticity is related to the own-elasticity through $(n-1)\epsilon^i_j = -(1+\epsilon^i_i)$. For any n, we obtain the slope in terms of only two steady state objects, the own-elasticity and the markup:

Proposition 1. *In a sector with n firms, the slope of the reaction function around the steady state* $\beta = \frac{\partial g^j}{\partial p_k}(\bar{p})$ *satisfies*

$$(n-1)\beta = \frac{\lambda+\rho}{\lambda}\frac{1}{1+\frac{1}{(n-1)[(|\epsilon_i|-1)(\bar{\mu}-1)-1]}} \tag{4}$$

where $\epsilon_i = \frac{\partial \log d^i}{\partial \log p_i}$ *and* $\bar{\mu} = \frac{\bar{p}}{W}$.

Proposition 1 is our first main result, showing how to locally infer unobserved steady state strategies from a small number of potentially observed sufficient statistics. Taking as given market concentration n and the demand elasticity ϵ_i, a higher steady state markup $\bar{\mu}$ is associated with a higher slope β. The intuition behind this result is based on reverse causality. Suppose that β is high. Then, whenever firm i decreases its price below the steady state, its rivals will lower prices as well. This undermines firm i's incentives to cut prices. This threat of undercutting allows the sector to sustain a high equilibrium markup. Turning the argument on its head, for a given elasticity ϵ_i, a high equilibrium markup must then be a consequence of steep reaction functions.

(4) is a relation between endogenous objects. In particular, we cannot take limits in n or λ while holding $\bar{\mu}$ fixed. In section 5, we show how varying n and other parameters affects both sides of the equation. While in static oligopoly models, the simple Lerner formula relates the markup to the demand elasticity

$$\mu^{\text{Nash}} = \frac{|\epsilon_i|}{|\epsilon_i| - 1}$$

under dynamic oligopoly markups are much more complex and depend on many factors beyond the demand elasticity. We use (4) to express the slope given the markup, but another useful formulation relates the steady state markup to the static Bertrand-Nash markup and the slope through

$$\bar{\mu} = \mu^{\text{Nash}} + \frac{1}{(n-1)\left(|\epsilon_i| - 1\right)} \times \frac{\frac{\lambda(n-1)\beta}{\lambda+\rho}}{1 - \frac{\lambda(n-1)\beta}{\lambda+\rho}}.$$

The advantage of Proposition 1 is that in order to infer the slope, we don't need to know the factors behind an observed markup. This is the sense in which $\bar{\mu}$ is a sufficient statistic.

In the next section, we show that steep reaction functions around the steady state imply a slow aggregate price adjustment following monetary shocks and thus persistent real effects of monetary policy.

4 Aggregate Effects of Permanent Monetary Shocks: Sufficient Statistics

We now study an unanticipated permanent shock to money. In particular, suppose initial prices are all equal, $p_{i,s} = P_-$, and aggregates are at a steady state with constant M_-, C_-, ℓ_-, W_- and $R_- = \rho$. Consider a permanent monetary shock arriving at $t = 0$ so that $M(t) = M_+ = (1+\delta)M_-$ for all $t \geq 0$.

In general, firms would have to forecast the path of macroeconomic variables $P(t)$ and $C(t)$ when choosing their reset price strategies $g^{i,s}$. These strategies would in turn affect the evolution of $P(t)$ and $C(t)$. It is possible to accomodate this fixed-point problem as we do in section 8, but for now we want to focus on clear analytical results. In the spirit of Golosov and Lucas (2007), our assumptions on preferences lead to the following simplification:

Proposition 2. *Equilibrium aggregates satisfy*

$$W(t) = (1+\delta)W_-, \quad R(t) = \rho, \quad P(t)\,C(t)^{\sigma} = \rho M_+. \tag{5}$$

If in addition

$$\omega\sigma = 1 \tag{6}$$

then any equilibrium strategy g of the stationary oligopoly game remains an equilibrium strategy in the non-stationary game that follows the monetary shock.

Proposition 2, proved in Appendix A, is very useful: it shows when firms can ignore the transitional dynamics of macroeconomic variables following the monetary shock, and therefore allows us to extend results based on the partial equilibrium game in section 3 to general equilibrium. This is an exact result, not an approximation for small monetary shocks as in Alvarez and Lippi (2014).[12] From now on, we consider preferences that satisfy condition (6) except in section 8 which allows for arbitrary parameters.

[12] The classic paper by Rotemberg and Saloner (1986) analyzes (non-Markov) trigger strategies that sustain high "collusive" prices in bad times but lead to price wars during booms, because the latter are periods with higher temporary profits to compete over. However, we just showed why, in general equilibrium, treating the dynamic game as a repeated game can be misleading, because changes in real interest rates cancel out exactly fluctuations in aggregate demand when $\omega\sigma = 1$. Away from this benchmark, the incentives to cut prices could be higher or lower in booms, depending on how the elasticity of intertemporal substitution $1/\sigma$ compares to the elasticity of intratemporal substitution across sectors ω.

We are interested in the speed of convergence of the aggregate price level to its new steady state $\bar{P} = (1+\delta) P_{-}$, as summarized by the half-life of $\log P(t)$. From (5), $\log P(t) + \sigma \log C(t)$ is constant after the monetary shock so this also gives us the half-life of $\log C(t)$.

After the shock, the sectoral price level P_s follows a stochastic process in each sector, and unlike with monopolistic competition, there is no law of large numbers at the sectoral level when the number of firms is finite. However, aggregating across the continuum of sectors yields the following deterministic law of motion for the aggregate price level (which we prove in Appendix B):

Proposition 3 (Aggregation)**.** *To first-order in the size of the monetary shock δ, the aggregate price level follows for $t \geq 0$*

$$\log P(t) - \log \bar{P} = -\delta \int_s e^{-\lambda_s[1-(n_s-1)\beta_s]t} ds, \tag{7}$$

where β_s is the slope $\frac{\partial g^{i,s}}{\partial p_{j,s}}$ in sector s . Therefore the cumulative output effect of the shock is

$$\int_0^\infty \log\left(\frac{C(t)}{\bar{C}}\right) dt = \frac{\delta}{\sigma} \times \int_s \frac{ds}{\lambda_s [1-(n_s-1)\beta_s]}. \tag{8}$$

In the standard New Keynesian model with monopolistic competition and CES demand, the half-life of the log price level following a monetary shock (up to a factor $\ln 2$) is simply $1/\lambda$ (as in Woodford 2003) and the cumulative output effect is $\frac{\delta}{\sigma\lambda}$.[13] In the oligopolistic model, if $(n_s-1)\beta_s$ is low on average, then firms in each sector will reset prices close to the new steady state when given a chance, speeding up the convergence.

Combining the results from Propositions 1 to 3, we know the response of the aggregate price level and thus of output to a permanent monetary shock as a function of the distribution of three steady state statistics: markups, demand elasticities and industry concentration. If we can estimate these sufficient statistics, then it is not necessary to solve the full MPE to analyze the real effects of monetary policy. For instance, our formula tells us that all else equal, higher observed markups imply higher (unobserved) slopes $(n_s-1)\beta_s$.[14]

[13] With more general demand structures, for instance Kimball demand, the half-life can depart from $1/\lambda$ even under monopolistic competition, see Proposition 5 below.

[14] This holds conditional on the demand elasticity. If higher markups reflect a decline in the substitutability between competing varieties, and thus lower demand elasticities, then higher markups

5 The Effects of Rising Concentration and other Comparative Statics

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

In this section we seek to answer how stickiness would *change* when market concentration and other parameters change. 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.

From now on, we assume that within-sector aggregation H_s follows Kimball preferences: sectoral consumption C_s is the unique solution to

$$\frac{1}{n_s}\sum_{i\in I_s}\phi_s\left(\frac{c_{i,s}}{C_s}\right)=1 \tag{9}$$

for some increasing, concave, function ϕ_s such that $\phi_s(1)=1$. An important benchmark is the case where ϕ_s is a power function, in which case we obtain the standard CES aggregator across firms, i.e. $C_s=\left(\frac{1}{n_s}\sum_{i\in I_s}c_{i,s}^{1-\frac{1}{\eta}}\right)^{\frac{1}{1-\frac{1}{\eta}}}$.

5.1 Solution Method

In general, solving for the steady state markup requires solving the full MPE. Since we want a solution for any number of firms, the state space can become very large. Indeed, the IO literature also acknowledges this challenge and employs approximate solution concepts such as "oblivious equilibria" (Weintraub, Benkard and Van Roy, 2008) Here we avoid the computational burden by approximating consumer's utility in a way that generates an equilibrium that we can solve analytically. Crucially, our approximation leaves enough degrees of freedom to flexibly parametrize the elasticities of the demand system that can be estimated in practice.

Our construction is detailed in Appendix D, and the main idea is as follows. Iteratively differentiating (1) and (2) generates a system of equations relating the

may be associated with lower slopes instead, as we illustrate in section 5.3.

derivatives g', g'', and so on, to the steady state markup, demand elasticity ϵ_i^i, superelasticity ϵ_{ii}^i, and so on. Our formula (3) is one of such equations. The standard interpretation of this system treats the sequence of derivatives of g as unknowns, and the sequence of higher-order elasticities as given structural parameters. Instead, we acknowledge that it is empirically impossible to know such fine properties of utility 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 perturb the unknown higher order elasticities to achieve some desired properties for the derivatives of g. In particular, we seek to simplify the characterization of equilibrium by making g locally polynomial, meaning that all its derivatives higher than some order vanish when evaluated at the steady state.[15]

Formally, denote $\epsilon_{(k)}$ is the kth-own-superelasticity evaluated at a symmetric $\bar{p}$, i.e.,

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

Proposition 4. *For any order of approximation $m \geq 1$ 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 $k \geq m$.

The case $m = 1$ yields the static Bertrand-Nash equilibrium markup. In the remainder of the paper we will apply Proposition 4 in the case $m = 2$, which makes the game linear-quadratic. In a duopoly model, we can compute the full MPE using standard value function iteration, and compare the resulting steady state price to what follows from our solution method. Figure A.1 plots the steady state markup with $m = 1, 2, 3$, showing that $m = 2$ already provides an excellent approximation to the exact solution, which corresponds to $m = \infty$. The approximation should be

[15] Our approximation relates to the algorithm used in Krusell, Kuruscu and Smith (2002) and later called "Taylor projection" by Levintal (2018). In our context their method would still treat all elasticities as structural parameters but approximate the policy and value functions by polynomials of order m. Instead, we approximate the utility function.

even better as n increases, since all the orders m of approximation coincide with monopolistic competition as $n \to \infty$.

Parametrizing the Two Dimensions of Demand. In what follows, we use Klenow and Willis (2016)'s functional form for the Kimball aggregator ϕ_s, which is simpler to define through its derivative

$$\phi_s'(x) = \frac{\eta - 1}{\eta} \exp\left(\frac{1 - x^{\theta/\eta}}{\theta}\right). \tag{10}$$

η and θ control the elasticity and the superelasticity of demand, respectively: in the limit of monopolistic competition $n \to \infty$, the demand own-elasticity ϵ_i^i converges to $-\eta$ and the ratio $\frac{\epsilon_{ii}^i}{\epsilon_i^i}$, named the "superelasticity" of demand by Klenow and Willis (2016), converges to θ. The limit $\theta \to 0$ corresponds to a standard CES demand with $\phi_s(x) = x^{\frac{\eta-1}{\eta}}$.

With finite n, the perceived elasticities also depend on n because firms face a residual demand that depends on the number of rivals they have, as is well known in the CES case studied by Atkeson and Burstein (2008). We generalize their expressions to any Kimball aggregator in Appendix C, and also derive new expressions for the perceived superelasticities. In particular, with the functional form (10) we have:

$$\epsilon_{ii}^i = \frac{\partial^2 \log d^i}{\partial \log p_i^2} = -\frac{(n-1)}{n^2}\left[\eta^2 - (1+\omega)\,\eta + \omega + (n-2)\theta\eta\right]. \tag{11}$$

Equation (11) shows that as n goes to infinity, ϵ_{ii}^i converges to $-\theta\eta$ hence is non-zero if and only if θ is non-zero. But with finite n, ϵ_{ii}^i generally differs from zero even with $\theta = 0$.[16]

5.2 Market Concentration

We now turn to our main counterfactual exercise, in which we study how changes in market concentration (the number of firms n in a sector) affect the transmission of monetary policy. If we knew how our sufficient statistics changed with n, we could

[16] The precise relation between n and $\epsilon_i^i, \epsilon_{ii}^i$ that arises when η and θ are fixed in (11) has no particular empirical grounding. In section 5.2 we turn to a more general non-parametric model that controls $\epsilon_{ii}^i(n)$ directly (which is equivalent to letting θ depend on n in (11)) to match the heterogeneity in idiosyncratic cost pass-through observed in the data.

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

just plug them into (4) and 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 and varieties. We then explore an alternative, using available evidence on pass-through from costs to prices, calibrating these preferences to the number of firms to match the available evidence.

Exogenous Changes in Number of Firms. We first interpret η and θ in the Klenow and Willis (2016) functional form (10) as structural parameters that are robust to changes in the number of firms and varieties. The remaining parameters are described in Table 1.

With CES demand ($\theta = 0$), higher market concentration in the sense of lower n increases monetary non-neutrality. With $n = 2$ firms, the half-life under oligopoly is approximately 40% higher than under monopolistic competition. The amplification decreases rapidly with n, however: with $n = 10$ firms it is only around 10%.

The blue line in Figure 1 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. In principle, this dampening effect can be arbitrarily large: the half-life with high n is unbounded above when θ increases, but the half-life with $n = 2$ is always the same as under CES (we explain why in section 5.3). There is therefore no guarantee that oligopolistic competition generates more non-neutrality than monopolistic competition: the direction of the effect depends on finer properties of demand systems, in particular the superelasticity of demand and how it depends on concentration. Next, we describe a strategy to infer these properties from available pass-through estimates.

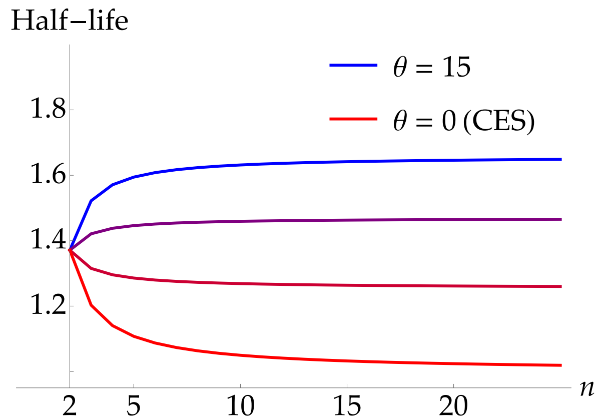

Figure 1: Half-life as a function of n for different values $\theta = 0$ (bottom red line), $5, 10, 15$ (top blue line), with $\eta = 10$.

A Calibration Based on Pass-Through. Previously, we fixed the "preference parameters" η, θ and changed the number of firms. We now provide an alternative that recalibrates other parameters as we change the number of firms. The shape of demand is crucial to understand how market structure impacts the transmission of monetary shocks, which affect all firms at the same time. As Atkeson and Burstein (2008) emphasized in a static setting, changes in residual demand also link market structure and the pass-through of *own cost* shocks, hereafter simply "pass-through". We now argue that the most recent and detailed pass-through estimates imply that market concentration significantly amplifies monetary non-neutrality.[17]

Amiti et al. (2019) 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.[18]

[17]Computing counterfactuals necessarily requires assumptions, and this recalibration strategy can be justified in two ways. First, we can assume that if a sector becomes more 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 sectors. Second, we can assume that aggregate concentration increases due to concentrated sectors becoming larger in a way that preserves within-sector demand (and thus pass-through).

[18]See also Berman, Martin and Mayer (2012), who show that the pass-through of exchange rates to export prices is lower for larger firms. Amiti et al. (2019) also provide direct estimates of strategic complementarities, defined as the regression coefficient γ of a firm i's price change on its competitors' price change, controlling for firm i's own cost change. In a static oligopoly model, the regression

We argued above that an increase in concentration (lower n) can dampen or amplify monetary policy transmission once we depart from nested CES systems. For the same reasons, in a static oligopolistic model with more general demand, an increase in market share holding industry concentration fixed could dampen or amplify pass-through. Reinterpreting Amiti et al. (2019)'s estimates within our dynamic model, we show that the empirical pattern of heterogeneity is consistent with a large superelasticity for large firms, and a small superelasticity for small firms.

Results. Figure 2 displays pass-through, computed in the dynamic model, under three specifications for within-sector demand. "AIK" is our baseline calibration: the superelasticity varies as a function of n through a variable parameter $\theta(n)$ (defined as in (11)) so as to match the relationship between market share and pass-through in a static Cournot model with $\eta = 10$ which, Amiti et al. (2019) argue, provides a good fit to their Belgian data. In "KW", θ is fixed at 10 as in Klenow and Willis (2016) and in standard DSGE calibration such as Smets and Wouters (2007). In "CES" θ is fixed at 0. In all cases, η equals 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 residual superelasticity of demand changes with concentration, but there is no reason for the residual elasticity itself to vary exactly as in (A.12). Ideally, one would obtain non-parametric estimates of $\epsilon_i^i(n)$ and $\epsilon_{ii}^i(n)$ from matching jointly the relation of markups and pass-through with market shares. However, there is no direct counterpart to Amiti et al. (2019), in part because markups are notably harder to estimate than pass-through. In the model with constant $\eta = 10$, going from $n = 4$ to 5 firms decreases prices by around 2%, which is broadly consistent with the evidence in Atkin, Faber and Gonzalez-Navarro (2018) and Busso and Galiani (2019).[19]

Figure 3 shows that under the calibration consistent with the micro evidence on

coefficient γ corresponds exactly to the slope of the firm's best response. Unfortunately, in a dynamic model, we cannot directly use estimates γ to obtain the slope β that governs transitional dynamics even when looking at long-run price changes. Intuitively, competitors' marginal costs mc_{-i} are an omitted variable affecting their future prices p_{-i}, hence influence firm i's decision when it gets to reset p_i.

[19] Recent work by Burstein, Carvalho and Grassi (2020) examines the relation between market shares and markups at the firm and sectoral levels. They 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 -0.27 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.

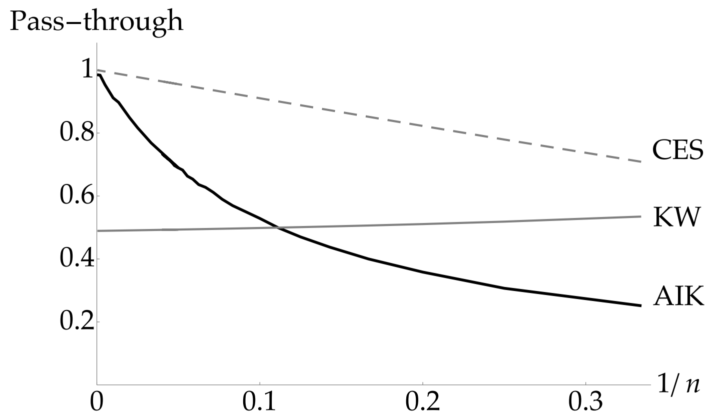

Figure 2: Pass-through as a function of market share $1/n$. AIK: variable superelasticity to match heterogeneity in pass-through from Amiti et al. (2019). KW: Fixed $\theta = 10$. CES: Fixed $\theta = 0$. In all cases, $\eta = 10$.

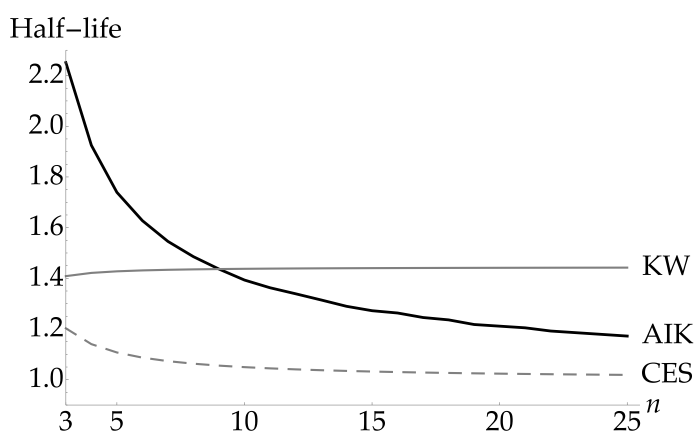

Figure 3: Half-life as a function of number of firms n. AIK: variable superelasticity to match heterogeneity in pass-through from Amiti et al. (2019). KW: Fixed $\theta = 10$. CES: Fixed $\theta = 0$. In all cases, $\eta = 10$.

pass-through, a rise in national concentration corresponding to an increase in the average Herfindahl index $1/n$ from 0.05 to 0.1, reflecting the observed trends since 1990 in e.g. Gutiérrez and Philippon (2017), amplifies the real effects of monetary policy by around 15%.[20]

Remark 2. When studying the relation between market share and pass-through, we vary n and maintain the symmetry assumption, so that the number of firms is the only source of variation in market share. The results would be very similar with variation in market share stemming from within-sector heterogeneity instead. Indeed, under static Bertrand or Cournot competition, market share is a sufficient statistic for pass-through: a large firm with a given market share passes through its costs to its prices in the same way whether it faces many small competitors or a few large ones. The same insight applies quantitatively in the dynamic model: Figure A.5 shows that 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 heterogeneity among a fixed number of firms.

5.3 Other Comparative Statics: the Determinants of Markups

Our sufficient statistics formula (4) highlights the role of markups. We next describe how different parameters can affect markups and monetary policy transmission, holding concentration (i.e., n) fixed.

The main results in this section are as follows. As under monopolistic competition, a lower elasticity of substitution η increases the markup. But the effect on the half-life is ambiguous because a lower η also decreases the demand elasticity ϵ_i^i. Under dynamic oligopoly, a higher superelasticity parameter θ increases both the markup and the half-life, in contrast to both monopolistic competition and static oligopoly models. Finally, markups are also increasing in the frequency of price changes and decreasing in discount rates.

Changes in the Elasticity of Substitution η and the Superelasticity Parameter θ. With CES demand, the steady state markup and the demand elasticity are related

[20] Rossi-Hansberg et al. (2020), however, show a decline in *local* concentration. An interesting open question is then which level of geographic or economic aggregation (what we call "sectors" s) is most relevant for the competition that determines consumer price inflation. The answer depends in part on the prevalence of "uniform pricing" policies (DellaVigna and Gentzkow, 2019).

one-to-one, making it sufficient to know a single statistic, the markup, to infer the half-life of monetary shocks from (4). In Appendix C.1 we show that this is also true with any homothetic preferences when there are only $n = 2$ symmetric firms, whether the cross-sector aggregator has unit elasticity ω or not. In the case of Kimball demand, this means that θ is irrelevant when $n = 2$, as can be seen directly from (11), and in Figure 1, showing that 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 (4) also requires information on demand elasticities. It is thus important to allow for more than two firms in each sector to relax the strong restrictions imposed by the duopoly model.

Figure A.2 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. When $n = 2$, the value of the superelasticity parameter θ does not matter, and we have a negative relation between the markup and the half-life (which is also present in the duopoly model with menu costs of Mongey 2018). However, as soon as there are at least $n = 3$ firms, there is a crucial interaction between θ and η. 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 will provide an intuition behind this fact in section 7.

A crucial difference between our framework and a monopolistically competitive economy is that (when $n > 2$) the superelasticity parameter θ can generate variations in the steady state markup $\bar{\mu}$ for a given demand elasticity. Figure A.3 shows an example with $n = 3$. The left panel shows that as θ increases, the markup under dynamic oligopoly rises. Equilibrium markups depend on multiple factors, and varying θ is the most transparent way to apply formula (4), as in that case a higher markup unambiguously implies a larger half-life, as on the right panel. In a model with monopolistic competition and Kimball (1995) demand, θ would also increase non-neutrality through complementarities in pricing, but would have no effect on the markup around a zero inflation steady state, hence markups would be uninformative about the strength of monetary policy.

Changes in Discount Rates and Price Stickiness. In the oligopoly model, the discount rate ρ and the frequency of price changes λ can also affect steady state markups (and therefore the slopes β). 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.4 shows that markups increase with λ, and more so for low n. Together with the effect of the superelasticity parameter θ, this finding confirms that equilibrium markups are complex objects that depend on many features of the environment beyond residual demand elasticities, something that dynamic monopolistic competition models ($n = \infty, 0 < \lambda \leq \infty$) and static oligopolistic models ($n < \infty$, $\lambda = 0$) fail to capture. Yet recall that markups can be used as sufficient statistic in Proposition 1 exactly because we do not need to know exactly where they come from.

In the limit of infinitely sticky prices $\lambda \to 0$, firms play the one-shot best-response, and so the equilibrium coincides with the static Bertrand-Nash equilibrium, both in terms of steady state markup and reaction function. On the other hand, the limit of infinitely frequent price changes $\lambda \to \infty$ (or equivalently $\rho \to 0$) does not equal the frictionless (flexible price) model, in which firms would play the static Bertrand-Nash equilibrium at each instant.[21]

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.

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

6.1 Heterogeneous 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 correlations between frequency of price changes and market structure. 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. Models with menu costs such as those proposed in these papers provide a microfoundation for the effect of concentration on price flexibility. Although our Calvo framework does not endogenize these correlations, interesting insights still arise from taking these correlations as given, by letting λ_s in sector s vary with the number of firms n_s and deriving implications for the aggregate effects of monetary policy.

From (8), the cumulative output effect for a monetary shock of size δ is:

$$\frac{\delta}{\sigma} \times \left\{ \mathbf{E}\left[\frac{1}{\lambda_s}\right] \mathbf{E}\left[\frac{1}{1-(n_s-1)\,\beta_s}\right] + \mathbf{Cov}\left(\frac{1}{\lambda_s}, \frac{1}{1-(n_s-1)\,\beta_s}\right) \right\} \tag{12}$$

where $\mathbf{E}$ denotes the average across sectors. In section 5.2 we saw that $\frac{1}{1-(n_s-1)\beta_s}$ is higher in more concentrated sectors. If they are also characterized by a lower frequency λ_s, then the covariance term is positive, which increases aggregate non-neutrality relative to a case with homogeneous frequency of price changes across sectors. Figure A.6 shows, in an example with one concentrated sector ($n = 3$) and one competitive sector ($n = 20$), that the covariance term represented by the shaded area can be significant. If the two sectors have the same price duration of 12 months, then the cumulative output effect is already 73% higher than in a "standard" CES New Keynesian model with monopolistic competition. If, instead, the durations are 15 months in the concentrated sector and 9 months in the competitive sector, then the cumulative output effect is 87% higher than in the "standard" model. This channel is specific to the oligopoly model, and differs from the role of heterogeneity under monopolistic competition, e.g., in Carvalho (2006) under Calvo pricing or Nakamura and Steinsson (2010) under menu costs.[22]

[22]Even under monopolistic competition and CES demand, the cumulative output effect $\frac{\delta}{\sigma}\mathbf{E}\left[\frac{1}{\lambda_s}\right]$ is convex in the sectoral frequencies $\{\lambda_s\}$, hence non-neutrality is amplified relative to a homogeneous

6.2 Within-Sector Firm Heterogeneity

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 consumers have different tastes captured by multiplicative demand shifters ξ_i so that sectoral consumption C_s solves $\frac{1}{n}\sum_{i\in I_s}\phi\left(\frac{\xi_i c_i}{C_s}\right) = 1$ instead of (9). Firms also differ in their productivity z_i. As a result the nominal profit of firm i can be written as

$$\Pi^i(t) = \left(\tilde{p}_i(t) - \frac{W(t)}{\xi_i z_i}\right) d^i\left(\tilde{p}_i\left(t\right), \tilde{p}_{-i}\left(t\right)\right) \tag{13}$$

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.

We solve the heterogeneous firms in Appendix E using the same method as in the symmetric case, with the additional complexity that we now need to solve for an asymmetric steady state price vector and a *matrix* of strategy slopes $\beta^i_j = \frac{\partial g^i}{\partial p_j}$. In order to keep the solution tractable, we impose a simple form of heterogeneity: in each sector, there are two types of firm a and b, with n_a (resp. n_b) firms of type a (resp. b), with the convention $n_a \leq n_b$. Firms of a given type are identical; firms of different types can differ arbitrarily in (ξ, z). The two types of firms allow us to capture the case of small and large firms in a sector.[23]

For any relative productivity or demand shifters, we solve for the equilibrium of the heterogeneous firms model, and compute the resulting steady state market shares and inverse HHI for each type of firm, as well as the half-life of the aggregate price level in response to monetary shocks. By varying continuously the heterogeneity, we can generate any inverse HHI between n_a and $n_a + n_b$.

Results. The first finding is that heterogeneity amplifies non-neutrality. Figures A.7 and A.8 show the slopes β^i_j for $i, j \in \{a, b\}$ and the resulting half-life of the aggregate price level as functions of market shares, in an example with $n_a = 2$ and $n_b = 8$. The half-life is minimized when firms a and b are identical, which corre-

economy that matches the average frequency $\mathbf{E}[\lambda_s]$. We point out an additional effect stemming from the empirical positive correlation between concentration and price duration.

[23] In theory we could solve the model with any $m \leq n$ types of firms (in a sector with n firms). We would then need to solve for m prices and m^2 slopes. We assume $m = 2$ for computational simplicity.

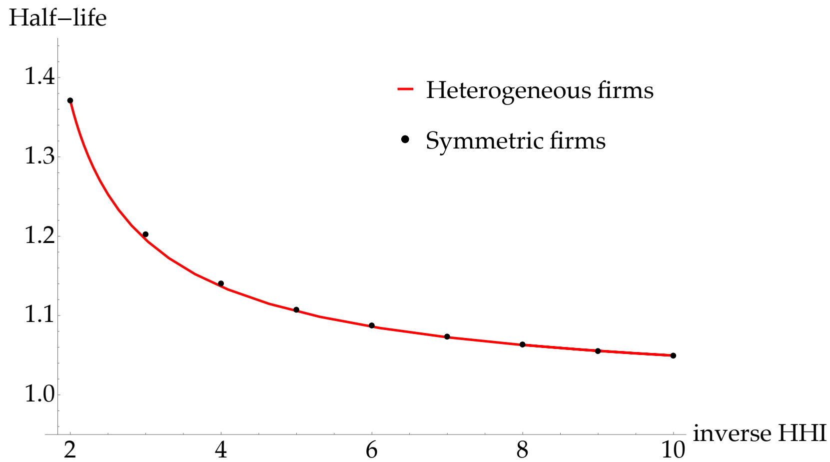

Figure 4: Heterogeneous vs. symmetric firms.

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

sponds to our baseline model with 10 symmetric firms. Intuitively, heterogeneity increases non-neutrality by reducing the effective number of firms; for instance, in the limit case where a firms become much larger than b firms (e.g., by being much more productive), the model becomes equivalent to one with only $n_a = 2$ firms and thus a large half-life. In fact, the analogy with an effective number of firms is more general, as we discuss next.

The second and main 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 4 shows the half-life as a function of the inverse HHI of type-a firms. In this example with $n_a = 2$ and $n_b = 8$, the inverse HHI of a firms can take any value between 2 and 10. Each black dot represents the half-life of a model with $n = 2, 3, \ldots, 10$ symmetric firms; under symmetry, the inverse HHI is just the number of firms. At the two extremes (inverse HHI equals to 2 or 10), the model with heterogeneous firms collapses exactly to a model with symmetric firms, so the corresponding black dots lie exactly on the red line. What the figure shows is that even between these two extremes, the black dots remain extremely close to the red line, that is, the half-life in a model with heterogeneous firms and inverse HHI equal to n is essentially the same as in a model with n symmetric firms. The same conclusion holds for other 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.

7 Inspecting the Mechanism: Strategic Behavior vs. Atomistic Feedback a la Kimball

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 what the literature with monopolistic competition calls strategic complementarities in pricing, that could arise from variable markups as in our setting, or other channels such as intermediate inputs or decreasing returns in production. The decomposition we propose is only meaningful under oligopoly, because under monopolistic competition, no single firm can affect the sectoral price index hence strategic effects are nil.

We disentangle the two effects through the lens of a "non-strategic" model. For each n, the associated non-strategic model is an economy with monopolistic competition ($n = \infty$) and modified Kimball preferences that match the residual demand elasticity ϵ_i^i and superelasticity ϵ_{ii}^i of the oligopolistic model with Kimball preferences ϕ and n firms.[24] The non-strategic model captures all the feedback (which in our context only arises from properties of the demand system), while suppressing strategic effects thanks to the monopolistic competition assumption.

We compute the half-life $\widetilde{hl}(n)$ of this non-strategic model, and then define strategic effects in the MPE as the increase in the half-life (relative to $1/\lambda$, the half-life in the standard New Keynesian model with monopolistic competition and CES demand) not explained by the non-strategic model:

$$\frac{hl\,(n)}{1/\lambda} = \underbrace{\frac{\widetilde{hl}\,(n)}{1/\lambda}}_{\text{feedback effect}} \times \underbrace{\frac{hl\,(n)}{\widetilde{hl}\,(n)}}_{\text{strategic effect}} .$$

[24]This non-strategic model also has a behavioral interpretation. Suppose that all firms are non-strategic 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.

As n goes to infinity, $hl/\widetilde{hl}$ goes to 1 and the strategic effect disappears; what is left is the standard feedback effect that can stem from a Kimball (1995) demand with positive superelasticity.

7.1 The Non-Strategic Model

The steady state price of the non-strategic model is the static Bertrand-Nash price p^{NE}, that solves $\Pi_i^i\left(p^{NE}\right) = 0$. We look for a symmetric equilibrium where, to first order, each resetting firm i sets $p_i^*(t) = \widetilde{\beta}\sum_{j\neq i} p_j(t)$. When it resets, given other firms' strategies $\widetilde{\beta}$, firm i chooses $p_i^*(t)$ to maximize

$$\mathbf{E}_t\left[\int_t^\infty e^{-(\lambda+\rho)(s-t)}\Pi^i(p_i^*(t), p_{-i}(t+s))ds\right].$$

The key difference with the MPE defined by the Bellman equation (1) is that here, firm i treats the evolution of rivals' prices as exogenous to its choice p_i^*. Define the feedback parameter as

$$\Gamma_n = \frac{(n-1)\Pi_{ij}}{-\Pi_{ii}}.$$

Γ_n is a measure of static feedback effects: it is the slope of the best response of a firm to a simultaneous price change by all its competitors in a static Bertrand-Nash equilibrium.[25] In Appendix F we prove the following:

Proposition 5. *The half-life of the aggregate price level in the non-strategic equilibrium is*

$$\widetilde{hl}(n) = \frac{1}{\lambda\left(1-\left(\frac{\rho+2\lambda}{2\lambda}\right)\left[1-\sqrt{1-\frac{4\lambda(\rho+\lambda)}{(\rho+2\lambda)^2}\Gamma_n}\right]\right)}. \tag{14}$$

and is thus equal to $1/\lambda$ when $\Gamma_n = 0$, and increasing in Γ_n.

The non-strategic price-setting strategy is forward-looking and differs from the static best-response that would maximize current profits. The static best-response would imply a half-life equal to $\frac{1}{\lambda(1-\Gamma_n)}$. The non-strategic half-life $\widetilde{hl}$ is strictly lower and depends on ρ.

[25] We can reexpress Γ_n around the Nash markup in terms of the demand elasticities $\epsilon_i^i = \frac{\partial \log d^i}{d \log p_i}$ and $\epsilon_{ii}^i = \frac{\partial^2 \log d^i}{\partial \log p_i^2}$ as $\Gamma_n = \frac{\epsilon_{ii}^i(n)/\epsilon_i^i(n)}{\epsilon_{ii}^i(n)/\epsilon_i^i(n)-\epsilon_i^i(n)-1}$. $\Gamma_\infty/\left(1-\Gamma_\infty\right)$ is also known as the *markup elasticity* (Gopinath and Itskhoki, 2010) or *responsiveness* (Berger and Vavra, 2019).

As n goes to infinity, Γ_n goes to 0 and $\widetilde{hl}$ to $1/\lambda$ in the standard CES case. Away from CES, Γ_n can converge to a positive limit. With a finite number of firms, even CES demand implies $\Gamma_n > 0$ and thus $\widetilde{hl} > 1/\lambda$.

Comparative Statics. The effect of oligopoly on monetary policy transmission is transparent in the non-strategic model, as it is entirely captured by Γ_n that we can compute in closed form. We can now see that the behavior of Γ_n plays a large part in section 5.3's findings. In the Klenow and Willis (2016) specification, Γ_n decreases with the elasticity of substitution η (and thus the observed markup) if and only if θ is low enough, which explains why, in Figure A.2, the half-life is decreasing in the markup $\bar{\mu}$ under CES ($\theta = 0$) but not when θ is high.[26] Similarly, we can use the non-strategic model to understand how concentration affects the half-life. As in Figure 1, this depends again on the value of θ: feedback Γ_n is decreasing in n (and thus increasing in concentration) if and only if θ is low enough.[27]

In principle, insights based on the non-strategic model could fail to be valid in the full MPE, due to sufficiently strong strategic effects that work in the opposite direction. But as we show next, strategic effects $hl/\widetilde{hl}$ are quantitatively modest.

7.2 Measuring Strategic Effects

While strategic effects are important determinants of steady state markups, as we saw in Figure A.3, we find that quantitatively, they do not explain much of the aggregate response to monetary shocks under oligopoly. Figure 5 displays the strategic effect, defined as $hl(n)/\widetilde{hl}(n)$, as n varies. We contrast our baseline calibration "AIK" with variable superelasticity (defined in section 5.2) with the CES case and a Kimball demand with fixed $\theta = 10$ "KW" (as in Klenow and Willis 2016). There is an interaction between strategic effects and feedback effects: strategic effects are considerably stronger in the "AIK" calibration, which also features stronger feedback effects. This interaction is intuitive: the only 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 5% higher than the non-strategic half-life. Consistent with their definition, strategic

[26] The exact condition is $\theta < \frac{n}{n-2} \times \frac{(\eta-1)^2}{1+(n-1)\eta^2}$.

[27] The exact condition is $\theta < \frac{(\eta-1)^2}{\eta+1}$.

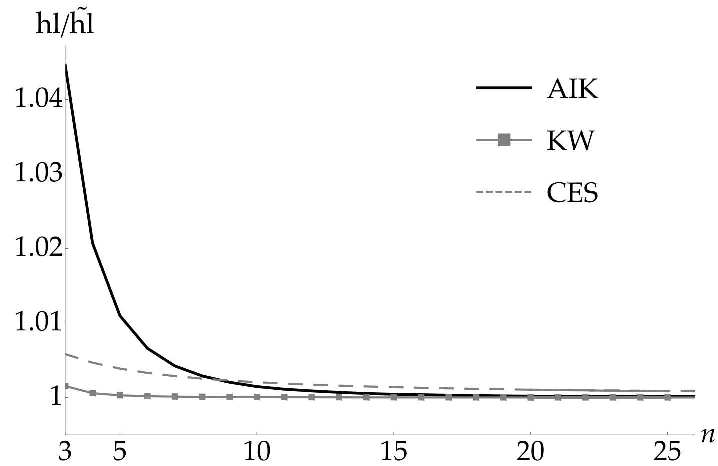

Figure 5: Strategic effect $hl(n)/\widetilde{hl}(n)$ as a function of n. AIK: variable superelasticity to match heterogeneity in pass-through from Amiti et al. (2019). KW: Fixed $\theta = 10$. CES: Fixed $\theta = 0$. In all cases, $\eta = 10$.

effects vanish as n grows and the economy approaches monopolistic competition: they fall below 1% when n exceeds 6.

Overall, our results suggest that oligopolistic competition significantly amplifies the real effects of monetary shocks, but primarily through "feedback effects", that is changes in residual demand elasticities as measured by Γ_n. While this implies that a simpler model of oligopolistic that abstracts away from strategic interactions goes a long way in explaining the economy's response to monetary shocks, this quantitative conclusion can only be reached after formulating and solving the fully strategic model. Moreover, in the next section we show that strategic effects can play a much more important role once we generalize the model to more complex monetary policy experiments.

8 The Three-Equation Oligopolistic New Keynesian Model

We focused so far on the dynamics following a permanent money supply shock, under the Golosov and Lucas (2007) assumptions (6). In this section we take a step closer to the New Keynesian model. We allow for general preferences as well as non-stationary dynamics and more realistic monetary policy shocks to the nominal interest rate. The derivations, in Appendix G, are much more involved, which is why we use a simpler framework as our baseline model. The main payoff is an oligopolistic Phillips curve that can be embedded in a standard DSGE model once combined with an Euler equation and an interest rate policy rule.

8.1 The Oligopolistic Phillips Curve

Denote $i(t)$ the nominal interest rate and $k(t)$ the log real marginal cost. We obtain the following Phillips curve, in integral form:

Proposition 6. *There exists a $q \times q$ matrix $\mathbf{A}$ with $q \leq 7$ that depends on the steady state demand elasticities, markup and slope β (described in Appendix G), with eigenvalues $\{\nu_j\}_{j=1}^{q}$, such that inflation follows*

$$\pi(t) = \int_0^\infty \gamma^k(s)\, k(t+s)\, ds + \int_0^\infty \gamma^c(s)\, c(t+s)\, ds + \int_0^\infty \gamma^i(s)\, (i(t+s) - \rho)\, ds \tag{15}$$

where for each variable $x \in \{k, c, i\}$, $\gamma^x(s)$ is a linear combination of $\{e^{-\nu_j s}\}_{j=1}^{q}$.

In general there $q = 7$ eigenvalues, but under a simplifying condition (A.26) given in Appendix G, which we assume in what follows, q can be reduced to 3.

Under monopolistic competition, even with Kimball preferences parametrized by Γ (as in section 7), there is a single eigenvalue $\nu_1 = \rho$ instead of three, and $\gamma^c = \gamma^i = 0$, hence the Phillips curve in integral form (15) simplifies to

$$\pi(t) = (1 - \Gamma)\, \lambda\, (\lambda + \rho) \int_0^\infty e^{-\rho s} k(t+s)\, ds. \tag{16}$$

The "slope of the Phillips curve" is equal to $(1 - \Gamma)\, \lambda\, (\lambda + \rho)$: higher feedback effects Γ flatten the Phillips curve, but in a way that is isomorphic to a higher degree of stickiness λ.

Under oligopolistic competition, inflation is also determined by a weighted average of future marginal costs, but oligopoly is not isomorphic to a higher stickiness parameter λ due to two important differences. First, there are multiple eigenvalues, so we cannot summarize the Phillips curve into a simple scalar "slope". Second, inflation depends on more than future marginal costs, as the second sum in (15) relates current inflation to future consumption and nominal interest rates. 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.

As with our earlier permanent money supply shocks, we can compare (15) to a "non-strategic" Phillips curve that corresponds to a monopolistic competitive economy with Kimball preferences that match the elasticity and superelasticity of the oligopolistic economy, characterized by (16) with $\Gamma = \Gamma_n$ given in section 7.1.

We can transform the integral form (15) into a scalar ordinary differential equation for inflation. For instance, for $n = 3$, under the AIK calibration and other parameters as in Table 1, we have

$$\dot{\pi} = 0.08\pi - 0.2k + 1.53\ddot{\pi} + \underbrace{0.37\dot{k} + 0.03\left(i - \rho\right)}_{=u}. \tag{17}$$

The corresponding non-strategic Phillips curve and the standard CES Phillips curve under the same parameters are respectively

$$\dot{\pi} = 0.05\pi - 0.17k, \tag{18}$$

$$\dot{\pi} = 0.05\pi - 1.05k. \tag{19}$$

Both (17) and (18) feature a smaller "slope" (i.e., the coefficient on k) than (19). Relative to (18), the oligopolistic Phillips curve (17) also features (i) more discounting, (ii) inflation persistence in the term $1.53\ddot{\pi}$, and (iii) a term u that resembles an endogenous "cost-push" shock. We study next how these differences can generate significant differences between the oligopoly model and Kimball monopolistic competition, that is, significant strategic effects.

8.2 Three-Equation Model

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

$$\dot{c} = \sigma^{-1}\left(i - \pi - r^n\right),$$

and a monetary policy interest rate rule

$$i = \kappa\rho + \left(1 - \kappa\right) r^n + \phi_\pi \pi + \epsilon^m,$$

where $r^n\left(t\right) = \rho + \epsilon^r\left(t\right)$ is the natural real interest rate and $\epsilon^m\left(t\right)$ is a monetary shock. $1 - \kappa$ measures how well the central bank is able to track the natural rate; κ can be thought of as monetary policy inertia. For simplicity, agents have perfect foresight over the shocks ϵ^r, ϵ^m.

Calibration. 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 $k = (\psi + \sigma) c$. We set standard values of $\sigma^{-1} = 1$ for the elasticity of intertemporal substitution (as in our monetary shock experiments), $\psi^{-1} = 0.5$ for the Frisch elasticity of labor supply, $\phi_\pi = 1.5$ for the Taylor rule coefficient on inflation, and $\kappa = 0.8$ for monetary policy inertia.

One-time Shocks. Consider first geometrically decaying unanticipated shocks

$$\epsilon^m(t) = \epsilon_0^m e^{-\xi t}, \quad \epsilon^r(t) = \epsilon_0^r e^{-\xi t}$$

with the same decay ξ (a particular case being only one type of shock). It is a standard result in the literature (Woodford, 2003) that under monopolistic competition, all the equilibrium variables are proportional to $e^{-\xi t}$. The same applies to the oligopolistic model, hence all the differences between economies are summarized by the impact effect, e.g. $c(t) = c(0) e^{-\xi t}$ and the cumulative output effect is $c(0)/\xi$. This contrasts with the case of permanent money supply shocks, for which impact effects were common to all economies and differences were summarized by the half-life.

Figure 6 displays the impact effect on consumption $c(0)$ for a 100 bps monetary shock $\epsilon_0^m = -0.01$ with $\xi = 1$. The message is consistent with what we found for permanent shocks to the money supply: concentration amplifies monetary non-neutrality by a significant amount. As Figure A.9 shows, a large part of the amplification can again be explained by feedback effects. Denoting $\tilde{c}(0)$ the initial consumption jump in the monopolistic Kimball economy calibrated to match the parameter Γ_n for each n, we find that $c(0)$ is actually lower than $\tilde{c}(0)$ (so that "strategic effects" are not amplifying) and can deviate from $\tilde{c}(0)$ by around 5% when $n = 3$.

More General Shocks. The one-time shocks are not without loss of generality. For instance, the common exponential decay leaves no room for the endogenous cost-push shocks to generate different inflation persistence across models.

Once we allow for a more general process for shocks, there are also meaningful differences between the oligopolistic economy and the non-strategic economy. Consider for instance paths for real and monetary shocks generated from an Ornstein-

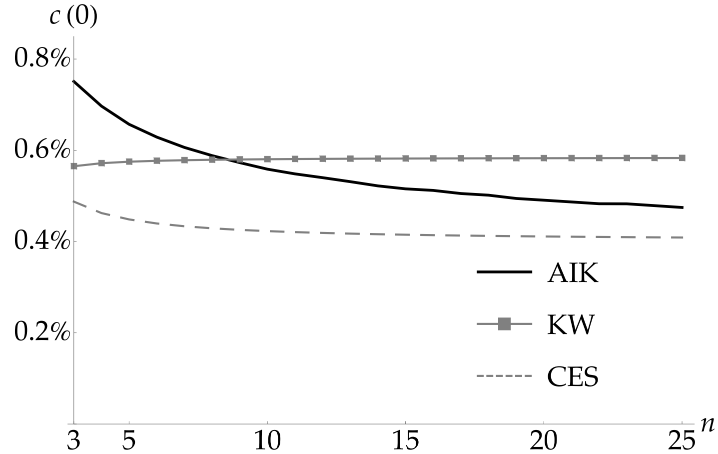

Figure 6: Impact effect of a $\epsilon_0^m = -1\%$ monetary shock on consumption $c(0)$ (log-deviation from steady state) as a function of number of firms n. AIK: variable superelasticity to match heterogeneity in pass-through from Amiti et al. (2019). KW: Fixed $\theta = 10$. CES: Fixed $\theta = 0$. In all cases, $\eta = 10$.

Uhlenbeck process (a continuous-time version of AR(1) processes) $d\epsilon = -a\epsilon + \sigma dZ$, where Z in a standard Brownian motion, and $a, \sigma^r > 0$ parametrize the speed of mean-reversion and variance of the shocks, respectively.[28] We set $a = 0.3$, $\sigma^r = 0.01$. Note that we are still assuming perfect foresight about the path, as in the case of exponentially decaying shocks.

Table A.1 shows the results for the two kinds of shocks under the "AIK" calibration. The standard deviations of inflation and consumption are smaller in the oligopolistic model than in the corresponding non-strategic model. The higher-order terms in the oligopolistic Phillips curve smooth out the path for inflation, which in turn makes the real rate and consumption less volatile. This example demonstrates that the strong equivalence between oligopoly and Kimball economies that we observe in the case of the literature's benchmark shocks (permanent money supply shocks and exponentially decaying interest rate shocks) does not necessarily transpose to more general processes.

[28]Technically we also multiply ϵ^r by a very slow exponential decay to ensure that the economy converges towards the deterministic steady state as $t \to \infty$.

9 Conclusion

In this paper, we studied how oligopolistic competition affects monetary policy transmission. We derived a closed-form formula for the response of aggregate output to monetary shocks as a function of three measurable sufficient statistics: demand elasticities, market concentration, and markups. Under our calibration, oligopolistic competition amplifies monetary non-neutrality, but, in the case of the standard shocks to money supply or interest rates studied in the literature, the response approximates a monopolistic competition model with Kimball demand that matches the residual demand elasticity and superelasticity of the oligopolistic model.

This does not imply, however, that oligopoly is isomorphic to monopolistic competition. First, a unique prediction of our model is the link between markups and subtle properties of demand functions such as superelasticities. Under monopolistic competition, superelasticities affect cost pass-through and thus monetary policy, but are irrelevant for markups. Under oligopolistic competition, higher superelasticities raise both markups and cost pass-through. Other factors, such as the frequency of price changes, also affect markups and pass-through: we discuss new implications for the role of sectoral heterogeneity in the transmission of monetary policy. Second, in the context of our three-equations oligopolistic New Keynesian model that allows for more general shocks and non-stationary dynamics, we find that the oligopolistic model can depart significantly from the recalibrated monopolistic model. In particular, the oligopolistic Phillips curve features a form of endogenous inflation persistence (or equivalently, endogenous cost-push shocks) that does not matter with standard shocks, but plays a role once we allow for richer dynamics.

Our calibration relies on estimates of exchange rate pass-through, as we believe they are the most relevant sources of information when studying strategic interactions. In the menu costs literature, it is more common to target moments of the distribution of price changes. The open economy literature on pass-through and the closed economy monetary literature have thus evolved mostly in parallel, with different conclusions regarding the strength of strategic complementarities in pricing. Our framework provides a natural way to reconcile these two strands: larger firms have more market power, only pass through a fraction of their idiosyncratic shocks, but drive most of the aggregate price stickiness. An interesting avenue for future empirical work would be to analyze how the distribution of price changes itself depends on firm size and market share.

References

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.

Atkin, David, Benjamin Faber, and Marco Gonzalez-Navarro, "Retail Globalization and Household Welfare: Evidence from Mexico," *Journal of Political Economy*, 2018, *126* (1), 1–73.

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.

Busso, Matias and Sebastian Galiani, "The Causal Effect of Competition on Prices and Quality: Evidence from a Field Experiment," *American Economic Journal: Applied Economics*, January 2019, *11* (1), 33–56.

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.

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.

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.

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.

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

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

Woodford, Michael, *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, 2003.

Appendix
For Online Publication

A Stationary Dynamics after a Permanent M shock

Proof of Proposition 2. 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

$$\begin{aligned}
\frac{\dot{C(t)}}{C(t)} &= \frac{1}{\sigma}\left(i(t)-\pi(t)-\rho\right) \\
N(t)^{\psi}C(t)^{\sigma} &= \frac{W(t)}{P(t)} \Rightarrow \psi\frac{\dot{N(t)}}{N(t)} = \frac{\dot{W(t)}}{W(t)} - i(t)+\rho \\
M(t)^{-\chi}P(t)^{\chi}C(t)^{\sigma} &= i(t)
\end{aligned}$$

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

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

To get constant wage $W(t) = W$ we need $i = \rho$ (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

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

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

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

Firm i maximizes the present discounted value of real profits using Arrow-Debreu SDF, that is

$$\begin{aligned} &\int e^{-\rho t} C(t)^{-\sigma} d^i\left(p_{i,s}, p_{-i,s}\right) C(t) P(t)^{\omega-1} \left[p_{i,s} - MC^i(t)\right] \\ =&\int e^{-\rho t} d^i\left(p_{i,s}, p_{-i,s}\right) C(t)^{1-\sigma} P(t)^{\omega-1} \left[p_{i,s} - MC^i(t)\right] \end{aligned}$$

so with $\sigma = 1$ and $\omega = 1$, firms can ignore the behavior of aggregate variables $P(t)$ and $C(t)$.

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

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

Therefore the demand shifter becomes

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

so if

$$\omega\sigma = 1$$

then firms can ignore the behavior of aggregate variables during the transition to the new steady state.

B Aggregation

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

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

where $z_i = p_i - \bar{p}$ (prices are in log). Denote first-order expansions of best responses by $p_i' = \alpha + \beta \left(\sum_{j \neq i} p_j \right)$ or equivalently $z_i' = \beta \left(\sum_{j \neq i} z_j \right)$. When firm i adjusts its price, the state of sector s changes to $v_s'(t) = M_i v_s(t)$ where M_i is the identity matrix except for row i which is equal to $(\beta, \dots, \beta, \underset{\substack{\uparrow \\ i}}{0}, \beta, \dots, \beta)$.

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

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

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

$$\dot{V}_t = n\lambda \left(\frac{\sum_i M_i}{n} - I_n \right) V_t$$

where

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

The aggregate price level is then, to first order, $\log P(t) = LV_t + \bar{p}$ where $L = \frac{1}{n}(1, \dots, 1)$. 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 $(1, \dots, 1)'$ is an eigenvector of μ_2, so if we start from symmetric initial conditions

$$V(0) = (p_0 - \bar{p}, \ldots, p_0 - \bar{p})$$

we have

$$V(t) = V(0)e^{\mu_2 t}$$

hence, to first order,

$$\log P(t) = \log \bar{P} + (\log P(0) - \log \bar{P})\, e^{-\lambda[1-(n-1)\beta]t}.$$

With heterogeneous sectors s the aggregation across sectors yields

$$\log P(t) = \log \bar{P} + (\log P(0) - \log \bar{P}) \int_s e^{-\lambda_s[1-(n_s-1)\beta_s]t} ds.$$

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

$$c^i + \sum_j p_j \frac{\partial c^j}{\partial p_i} = 0 \quad \text{(A.1)}$$

Then Slutsky symmetry and constant returns to scale imply

$$\epsilon^i_i + \sum_{j \neq i} \epsilon^i_j = -1 \quad \text{(A.2)}$$

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

$$\epsilon^i_j = -\frac{1 + \epsilon^i_i}{n-1} \quad \text{(A.3)}$$

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.2) with respect to $\log p_i$ to get

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

hence at a symmetric price,

$$\epsilon^i_{ii} + (n-1)\epsilon^i_{ij} = 0 \tag{A.4}$$

Differentiating once more the budget constraint with respect to p_i

$$2\frac{\partial c^i}{\partial p_i} + \sum_j \frac{\partial^2 c^j}{\partial p_i^2} = 0 \tag{A.5}$$

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.5)

$$\epsilon^i_{jj} = \epsilon^i_j\left(1 - \epsilon^i_j\right) - \frac{1}{n-1}\left[\epsilon^i_{ii} + \epsilon^i_i\left(1 + \epsilon^i_i\right)\right] \tag{A.6}$$

Finally, differentiating (A.1) 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,

$$2\epsilon^i_j + (n-2)\left[\epsilon^i_{jk} + \left(\epsilon^i_j\right)^2\right] + 2\left[\epsilon^i_{ij} + \epsilon^i_j \epsilon^i_i\right] = 0 \tag{A.7}$$

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

$$\Sigma_i = \frac{\partial \log(-\epsilon_i^i)}{\partial \log p_i} = \frac{\epsilon_{ii}^i}{\epsilon_i^i}$$

So in the end we have two degrees of freedom: $\{\epsilon_i^i, \epsilon_{ii}^i\}$ 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.4), (A.7) as before, but (A.3) and (A.6) become respectively

$$\epsilon_j^i = -\frac{\omega + \epsilon_i^i}{n-1}$$
$$\epsilon_{jj}^i = \epsilon_j^i \left(1 + \epsilon_j^i\right) - \frac{1}{n-1}\left[\epsilon_{ii}^i + \epsilon_i^i \left(\omega + \epsilon_i^i\right)\right].$$

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.7), the cross-superelasticity ϵ_{ij}^i is determined by elasticities, hence so is the own-superelasticity $\epsilon_{ii}^i = -(n-1)\epsilon_{ij}^i$.

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

$$\frac{1}{n}\sum_i \Psi\left(\frac{c_i}{C}\right) = 1 \tag{A.8}$$

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

$$p_i = \lambda \Psi'\left(\frac{c_i}{C}\right)\frac{1}{C} \tag{A.9}$$

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.9) as

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

Taking logs and differentating (A.9) 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.8) 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

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

$$R\left(\frac{c^k}{C}\right)\left[\epsilon_i^k - \frac{\partial \log C}{\partial \log p_i}\right] - R\left(\frac{c^i}{C}\right)\left[\epsilon_i^i - \frac{\partial \log C}{\partial \log p_i}\right] = 1 \tag{A.10}$$

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.6) 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.10) 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.7) we have

$$\epsilon_i^i=-\left[\left(\frac{n-1}{n}\right)\frac{1}{R(1)}+\frac{1}{n}\right] \tag{A.11}$$
$$\epsilon_j^i=\frac{\frac{1}{R(1)}-1}{n}$$
$$\epsilon_{ii}^i=-\frac{n-1}{n^2}\left[\frac{R(1)\left[1-R(1)\right]^2+(n-2)R'(1)}{R(1)^3}\right]$$
$$\epsilon_{ij}^i=\frac{R(1)\left[1-R(1)\right]^2+(n-2)R'(1)}{n^2R(1)^3}\quad (j\neq i)$$
$$\epsilon_{jj}^i=\frac{-(n-1)R(1)\left[1-R(1)\right]^2+(n-2)R'(1)}{n^2R(1)^3}\quad (j\neq i)$$
$$\epsilon_{jk}^i=\frac{R(1)\left[1-R(1)\right]^2-2R'(1)}{n^2R(1)^3}\quad (j\neq k, n\geq 3)$$

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

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

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

$$\epsilon_i^i = -\frac{\eta (n-1) + \omega}{n} \tag{A.12a}$$

$$\epsilon_j^i = \frac{\eta - \omega}{n} \tag{A.12b}$$

$$\epsilon_{ii}^i = -\frac{(n-1)}{n^2} \left[\eta^2 - (1+\omega)\eta + \omega + (n-2)\theta\eta\right] \tag{A.12c}$$

$$\epsilon_{ij}^i = \frac{\eta^2 - (1+\omega)\eta + \omega + (n-2)\theta\eta}{n^2} \tag{A.12d}$$

$$\epsilon_{jj}^i = \frac{(n-2)\theta\eta - (\eta-1)(n-1)(\eta-\omega)}{n^2} \tag{A.12e}$$

$$\epsilon_{jk}^i = \frac{\eta^2 - (1+\omega)\eta + \omega - 2\theta\eta}{n^2} \tag{A.12f}$$

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

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

with $S = 1/n$ denoting the market share. The approximation in the second line holds if S is small relative to $\eta/(1+\eta)$, 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 $\frac{\epsilon_{ii}^i}{\epsilon_i^i}$ 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 Perturbation of utility

Proof of Proposition 4. We start from the system that defines an MPE:

$$(\rho + n\lambda) V(p) = \Pi(p) + \lambda \sum_j V\left(g\left(p_{-j}\right), p_{-j}\right) \tag{A.13}$$

$$V_p\left(g\left(p_{-i}\right), p_{-i}\right) = 0 \tag{A.14}$$

Differentiating k times the Bellman equation (A.13) gives us for each $k \geq 1$ a linear system in the kth-derivatives $\mathbf{V}^{(k)} = (V_{11...11}, V_{11...12}, V_{11...22}, \dots)$ 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 $\mathbf{\Pi}^{(k)} = (\Pi_{11...11}, \dots)$ 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.14) gives us a sequence of equations that must be satisfied at a steady state

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

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

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

and so on. Section D.1 below shows that there are indeed enough degrees of freedom to make the equations $F^m, F^{m+1}, \dots$ 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 $\left(\epsilon_{(1)}, \dots, \epsilon_{(m)}, \tilde{\epsilon}_{(m+1)}, \dots, \tilde{\epsilon}_{(k)}\right)$ through the same computations as in Appendix C.

Given this construction, $\bar{p}, g', \dots, g^{(m-1)}$ are pinned down by $\left(\epsilon_{(1)}, \dots, \epsilon_{(m)}\right)$ as the solution to the system of equations F^k for $k = 1, \dots, m$.

D.1 Counting the degrees of freedom

The main potential impediment to the proof above 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.

Step 1: Computing $\#_n(m)$. 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*

$$\#_n(m) = \sum_{k=0}^{m} q_{n-1}(m-k) \tag{A.15}$$

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. □

Step 2: Computing the number of restrictions arising from demand integrability. Next, we want to use economic restrictions to reduce the number of degrees of freedom, ideally to 1, by having $\#_n(m) - 1$ independent equations. Our restrictions are

$$\Phi(p) = \sum_j p_j d^j(p) = 0 \quad \forall p \tag{A.16}$$

$$d^i_j(p) = d^j_i(p) \quad \forall p, \forall i, j \tag{A.17}$$

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.16) 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.15), 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 of partitions of m into n non-negative integers

$$q_n(m)$$

How many restrictions $\flat_n(m)$ do 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. The first terms are (see in next subsection)

$$\begin{aligned}
\flat_n(1) &= 0 \\
\flat_n(2) &= 1 \\
\flat_n(3) &= \begin{cases} 2 & \text{if } n \geq 3 \\ 1 & \text{if } n = 2 \end{cases} \\
\flat_n(4) &= \begin{cases} 5 & \text{if } n \geq 4 \\ 4 & \text{if } n = 3 \\ 3 & \text{if } n = 2 \end{cases}
\end{aligned}$$

Step 3: Comparing the two. 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 4:

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

$$q_n(m) + \flat_n(m) + 1 \leq \#_n(m) \tag{A.18}$$

Proof. We know by hand that (A.18) holds for $m = 1, 2$ so take $m \geq 3$. Then all the Slutsky conditions can be written as starting with

$$d_{12\ldots}^1 = \ldots$$

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.18) 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 $n \geq 3$. □

Note that so far we have considered general CRS demand functions. Restricting attention to the Kimball class makes the inequality (A.18) 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 4.

D.2 $m = 2$: Locally Linear Equilibrium

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

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

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

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

yields

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

with coefficients

$$A_{ii} = \beta\left((\beta+1)\lambda^3\left(\beta^2\left(-2n^2+9n-10\right)+\beta^3(n-2)+6\beta(n-2)-4\right)-\lambda^2\rho\left(\beta^3\left(n^2-5n+6\right)+\beta^2\left(2n^2-15n+22\right)+\beta(24-9n)+8\right)+\lambda\rho^2\left(\beta^2(n-2)+\beta(3n-8)-5\right)-\rho^3\right) \tag{A.19a}$$

$$A_{ij} = -\left(2(\beta+1)\lambda^3\left(-2\beta^3\left(n^2-3n+2\right)+\beta^4(n-1)+2\beta^2(n-1)-\beta(n-2)+1\right)+\lambda^2\rho\left(\beta^4\left(-2n^2+7n-5\right)-4\beta^3\left(n^2-4n+3\right)+3\beta^2 n-4\beta(n-3)+5\right)+\lambda\rho^2\left(\beta^2 n-2\beta(n-3)+4\right)+\rho^3\right) \tag{A.19b}$$

$$A_{jj} = \beta^2\lambda\left((\beta+1)\lambda^2\left(2\left(\beta^2+3\beta+2\right)+\beta(\beta+1)n^2-\left(3\beta^2+7\beta+2\right)n\right)+\lambda\rho\left(4\beta^2+10\beta+\beta(\beta+1)n^2-\left(5\beta^2+9\beta+3\right)n+6\right)+\rho^2(\beta-(\beta+1)n+2)\right) \tag{A.19c}$$

$$A_{jk} = -\beta\lambda(n-2)\left((\beta+1)\lambda^2\left(-\beta+\beta^3(n-1)+3\beta^2(n-1)+1\right)+\lambda\rho\left(2\beta^3(n-1)+\beta^2(3n-4)+2\right)+\rho^2\right) \tag{A.19d}$$

E Heterogeneous 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, their demand, or both. We know need to solve for six unknowns $\left\{\beta_a^a, \beta_b^a, \beta_a^b, \beta_b^b, p_a, p_b\right\}$ where β_j^i is 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

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

and, in the locally linear equilibrium:

$$\begin{aligned}
(\rho+\lambda)\, V_{ii}^{i,a} &= \Pi_{ii}^{i,a} + \lambda\,(n_a-1)\left[V_{j_aj_a}^{i,a}\left(\beta_a^a\right)^2 + 2V_{ij_a}^{i,a}\beta_a^a\right] + \lambda n_b\left[V_{j_bj_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\,(n_a-2)\left[V_{j_aj_a}^{i,a}\left(\beta_a^a\right)^2 + V_{j_ak_a}^{i,a}\beta_a^a + V_{ij_a}^{i,a}\beta_a^a\right] + \lambda n_b\left[V_{j_bj_b}^{i,a}\left(\beta_a^b\right)^2 + V_{j_ak_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\,(n_a-1)\left[V_{j_aj_a}^{i,a}\beta_a^a\beta_b^a + V_{j_ak_b}^{i,a}\beta_a^a + V_{ij_a}^{i,a}\beta_b^a\right] + \lambda\,(n_b-1)\left[V_{j_bj_b}^{i,a}\beta_a^b\beta_b^b + V_{j_bk_b}^{i,a}\beta_a^b + V_{ij_b}^{i,a}\beta_b^b\right] \\
(\rho+\lambda)\, V_{j_aj_a}^{i,a} &= \Pi_{j_aj_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\,(n_a-2)\left[V_{j_aj_a}^{i,a}\left(\beta_a^a\right)^2 + 2V_{j_ak_a}^{i,a}\beta_a^a\right] + \lambda n_b\left[V_{j_bj_b}^{i,a}\left(\beta_a^b\right)^2 + 2V_{j_ak_b}^{i,a}\beta_a^b\right] \\
(\rho+2\lambda)\, V_{j_ak_b}^{i,a} &= \Pi_{j_ak_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\,(n_a-2)\left[V_{j_aj_a}^{i,a}\beta_a^a\beta_b^a + V_{j_ak_b}^{i,a}\beta_a^a + V_{j_ak_a}^{i,a}\beta_b^a\right] \\
&\quad + \lambda\,(n_b-1)\left[V_{j_bj_b}^{i,a}\beta_a^b\beta_b^b + V_{j_bk_b}^{i,a}\beta_a^b + V_{j_ak_b}^{i,a}\beta_b^b\right] \\
(\rho+2\lambda)\, V_{j_ak_a}^{i,a} &= \Pi_{j_ak_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\,(n_a-3)\left[V_{j_aj_a}^{i,a}\left(\beta_a^a\right)^2 + 2V_{j_ak_a}^{i,a}\beta_a^a\right] + \lambda n_b\left[V_{j_bj_b}^{i,a}\left(\beta_a^b\right)^2 + 2V_{j_ak_b}^{i,a}\beta_a^b\right] \\
(\rho+\lambda)\, V_{j_bj_b}^{i,a} &= \Pi_{j_bj_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\,(n_a-1)\left[V_{j_aj_a}^{i,a}\left(\beta_b^a\right)^2 + 2V_{j_ak_b}^{i,a}\beta_b^a\right] + \lambda\,(n_b-1)\left[V_{j_bj_b}^{i,a}\left(\beta_b^b\right)^2 + 2V_{j_bk_b}^{i,a}\beta_b^b\right] \\
(\rho+2\lambda)\, V_{j_bk_b}^{i,a} &= \Pi_{j_bk_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\,(n_a-1)\left[V_{j_aj_a}^{i,a}\left(\beta_b^a\right)^2 + 2V_{j_ak_b}^{i,a}\beta_b^a\right] + \lambda\,(n_b-2)\left[V_{j_bj_b}^{i,a}\left(\beta_b^b\right)^2 + 2V_{j_bk_b}^{i,a}\beta_b^b\right]
\end{aligned}$$

We can use these 11 envelope conditions to solve linearly for $\left\{V_i^{i,a}, V_{j_a}^{i,a}, V_{j_b}^{i,a}, V_{ii}^{i,a}, \dots\right\}$, 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 Non-Strategic Model

The quadratic approximation of profit Π^i of firm i around the non-strategic steady state which is the static Nash p^{NE} writes (in log deviations)

$$\pi^i(p_i, Q_i, R_i) = BQ_i + CQ_i^2 + Dp_iQ_i + Ep_i^2 + FR_i$$

where

$$\begin{aligned} Q_i &= \sum_{j\neq i} p_j \\ R_i &= \sum_{j\neq i} p_j^2 \end{aligned}$$

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

$$D = \Pi_{ij}\left(p^{NE}(n)\right)$$

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

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

$$p_j^*(t) = \beta Q_j(t)$$

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 $p_i^*(t)$ such that

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

The FOC is

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

Therefore

$$(n-1)\beta = \Gamma_n \frac{1}{1 + \frac{\lambda}{\rho+\lambda}\left[1-(n-1)\beta\right]} \tag{A.20}$$

where the ratio $\Gamma_n = \frac{(n-1)\Pi_{ij}}{-\Pi_{ii}}$ 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 Γ_n to be strictly lower than 1 for a static symmetric Nash equilibrium to exist. (A.20) shows that the slope of the dynamic non-strategic best response at a stable steady state, if one exists, is always smaller than the slope of the static best response Γ_n. The stable root in $(0,1)$ is

$$(n-1)\beta = \left(\frac{\rho+2\lambda}{2\lambda}\right)\left[1-\sqrt{1-4\frac{\lambda(\rho+\lambda)}{(\rho+2\lambda)^2}\Gamma_n}\right]$$

G Oligopolistic Phillips Curve

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

$$(i_t+n\lambda)\,V^i(p,t) = V_t^i(p,t) + \Pi^i(p, MC_t, Z_t) + \lambda\sum_j V^i\left(g^j(p_{-j},t), p_{-j}, t\right) \tag{A.21}$$

$$V_i^i\left(g^i(p_{-i},t), p_{-i}, t\right) = 0 \tag{A.22}$$

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.[29]

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.21), 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 4 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

[29] In section 4, conditions (6) ensured a constant Z_t.

partial differential equations (PDEs)

$$0 = V^i_{it} + \Pi^i_i + \lambda\left(n-1\right)V^i_j\beta \tag{A.23a}$$

$$\left(i_t+\lambda\right)V^i_j = V^i_{jt} + \Pi^i_j + \lambda\left(n-2\right)V^i_j\beta \tag{A.23b}$$

$$\left(i_t+\lambda\right)V^i_{ii} = V^i_{iit} + \Pi^i_{ii} + \lambda\left(n-1\right)\left(V^i_{jj}\beta^2 + 2V^i_{ij}\beta\right) \tag{A.23c}$$

$$\left(i_t+2\lambda\right)V^i_{ij} = V^i_{ijt} + \Pi^i_{ij} + \lambda\left(n-2\right)\left(V^i_{jj}\beta^2 + V^i_{jk}\beta + \beta V^i_{ij}\right) \tag{A.23d}$$

$$\left(i_t+\lambda\right)V^i_{jj} = V^i_{jjt} + \Pi^i_{jj} + \lambda\left(n-2\right)\left(V^i_{jj}\beta^2 + 2\beta V^i_{jk}\right) + \lambda\left(V^i_{ii}\beta^2 + 2\beta V^i_{ij}\right) \tag{A.23e}$$

$$\left(i_t+2\lambda\right)V^i_{jk} = V^i_{jkt} + \Pi^i_{jk} + \lambda\left(n-3\right)\left(V^i_{jj}\beta^2 + 2\beta V^i_{jk}\right) + \lambda\left(V^i_{ii}\beta^2 + 2\beta V^i_{ij}\right) \tag{A.23f}$$

Denote the functions

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

and so on for all derivatives of the value function V^i. We can transform the system (A.23) into a system of ordinary differential equations in the functions $W^i_i\left(t\right), W^i_j\left(t\right)$, and so on. The partial derivatives with respect to time such as

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

in equations (A.23) 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):

$$0 = -\left[W^i_{ii} + (n-1)\,W^i_{ij}\right]\dot{\alpha} + \Pi^i_i + \lambda\,(n-1)\,W^i_j\beta \tag{A.25a}$$

$$(i_t+\lambda)\,W^i_j = \dot{W}^i_j - \left[W^i_{ij} + W^i_{jj} + (n-2)\,W^i_{jk}\right]\dot{\alpha} + \Pi^i_j + \lambda\,(n-2)\,W^i_j\beta \tag{A.25b}$$

$$0 = W^i_{ii}\beta + W^i_{ij} \tag{A.25c}$$

$$(i_t+\lambda)\,W^i_{ii} = \dot{W}^i_{ii} - \left[V^i_{iii} + (n-1)\,V^i_{iij}\right]\dot{\alpha} + \Pi^i_{ii} + \lambda\,(n-1)\left(W^i_{jj}\beta^2 + 2W^i_{ij}\beta\right) \tag{A.25d}$$

$$(i_t+2\lambda)\,W^i_{ij} = \dot{W}^i_{ij} - \left[V^i_{iij} + V^i_{ijj} + (n-2)\,V^i_{ijk}\right]\dot{\alpha} + \Pi^i_{ij} + \lambda\,(n-2)\left(W^i_{jj}\beta^2 + W^i_{jk}\beta + W^i_{ij}\beta\right) \tag{A.25e}$$

$$(i_t+\lambda)\,W^i_{jj} = \dot{W}^i_{jj} - \left[V^i_{ijj} + V^i_{jjj} + (n-2)\,V^i_{jjk}\right]\dot{\alpha} + \Pi^i_{jj} + \lambda\,(n-2)\left(W^i_{jj}\beta^2 + 2\beta W^i_{jk}\right) + \lambda\left(W^i_{ii}\beta^2 + 2\beta W^i_{ij}\right) \tag{A.25f}$$

$$(i_t+2\lambda)\,W^i_{jk} = \dot{W}^i_{jk} - \left[V^i_{ijk} + V^i_{jjk} + V^i_{jkk} + (n-3)\,V^i_{jkl}\right]\dot{\alpha} + \Pi^i_{jk} + \lambda\,(n-3)\left(W^i_{jj}\beta^2 + 2\beta W^i_{jk}\right) + \lambda\left(W^i_{ii}\beta^2 + 2\beta W^i_{ij}\right) \tag{A.25g}$$

Next, we linearize system (A.25) 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 marginal cost as

$$mc(t) = p(t) + k(t)$$

where $k(t)$ is the log-deviation of the real marginal cost. Profit derivatives such as $\Pi^i_i(t)$ in (A.25) are evaluated at the moving price $\alpha(t)$, hence become once linearized[30]

$$\pi^i_i(t) = \bar{\alpha}\left[\Pi^i_{ii} + (n-1)\,\Pi^i_{ij}\right]a(t) + \bar{M}C\Pi^i_{i,MC}\,(p(t)+k(t)) + \Pi^i_i\left(z_c c(t) + z_p p(t)\right)$$

$$\pi^i_j(t) = \bar{\alpha}\left[\Pi^i_{ij} + \Pi^i_{jj} + (n-2)\,\Pi^i_{jk}\right]a(t) + \bar{M}C\Pi^i_{j,MC}\,(p(t)+k(t)) + \Pi^i_j\left(z_c c(t) + z_p p(t)\right)$$

$$\pi^i_{ii}(t) = \bar{\alpha}\left[\Pi^i_{iii} + (n-1)\,\Pi^i_{iij}\right]a(t) + \bar{M}C\Pi^i_{ii,MC}\,(p(t)+k(t)) + \Pi^i_{ii}\left(z_c c(t) + z_p p(t)\right)$$

$$\pi^i_{ij}(t) = \bar{\alpha}\left[\Pi^i_{iij} + \Pi^i_{ijj} + (n-2)\,\Pi^i_{ijk}\right]a(t) + \bar{M}C\Pi^i_{ij,MC}\,(p(t)+k(t)) + \Pi^i_{ij}\left(z_c c(t) + z_p p(t)\right)$$

$$\pi^i_{jj}(t) = \bar{\alpha}\left[\Pi^i_{ijj} + \Pi^i_{jjj} + (n-2)\,\Pi^i_{jjk}\right]a(t) + \bar{M}C\Pi^i_{jj,MC}\,(p(t)+k(t)) + \Pi^i_{jj}\left(z_c c(t) + z_p p(t)\right)$$

$$\pi^i_{jk}(t) = \bar{\alpha}\left[\Pi^i_{ijk} + 2\Pi^i_{jjk} + (n-3)\,\Pi^i_{jkl}\right]a(t) + \bar{M}C\Pi^i_{jk,MC}\,(p(t)+k(t)) + \Pi^i_{jk}\left(z_c c(t) + z_p p(t)\right)$$

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

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

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

$$\left[V_{ii}^i + (n-1)\, V_{ij}^i\right] \dot{a}\,(t) = \frac{1}{\bar{\alpha}} \pi_i^i\,(t) + \lambda\,(n-1)\,\frac{V_j^i \beta}{\bar{\alpha}} \left[w_j^i\,(t) + b\,(t)\right]$$

$$(\rho+\lambda)\, w_j^i\,(t) + i_t - \rho = \dot{w}_j^i\,(t) - \bar{\alpha} \left[\frac{V_{ij}^i + V_{jj}^i + (n-2)\, V_{jk}^i}{V_j^i}\right] \dot{a}\,(t) + \frac{1}{V_j^i} \pi_j^i\,(t) + \lambda\,(n-2)\,\beta \left[w_j^i\,(t) + b\,(t)\right]$$

$$(\rho+\lambda)\, w_{ii}^i\,(t) + i_t - \rho = \dot{w}_{ii}^i\,(t) - \frac{\bar{\alpha}}{V_{ii}^i} \left[V_{iii}^i + (n-1)\, V_{iij}^i\right] \dot{a}\,(t) + \frac{1}{V_{ii}^i} \pi_{ii}^i\,(t)$$
$$+ \lambda\,(n-1) \left\{ \frac{V_{jj}^i \beta^2}{V_{ii}^i} \left[w_{jj}^i\,(t) + 2b\,(t)\right] + \frac{2 V_{ij}^i \beta}{V_{ii}^i} \left[w_{ij}^i\,(t) + b\,(t)\right] \right\}$$

$$(\rho+2\lambda)\, w_{ij}^i\,(t) + i_t - \rho = \dot{w}_{ij}^i\,(t) - \frac{\bar{\alpha}}{V_{ij}^i} \left[V_{iij}^i + V_{ijj}^i + (n-2)\, V_{ijk}^i\right] \dot{a}\,(t) + \frac{1}{V_{ij}^i} \pi_{ij}^i\,(t)$$
$$+ \lambda\,(n-2) \left\{ \frac{V_{jj}^i \beta^2}{V_{ij}^i} \left[w_{jj}^i\,(t) + 2b\,(t)\right] + \frac{V_{jk}^i \beta}{V_{ij}^i} \left[w_{jk}^i\,(t) + b\,(t)\right] + \beta \left[w_{ij}^i\,(t) + b\,(t)\right] \right\}$$

$$(\rho+\lambda)\, w_{jj}^i\,(t) + i_t - \rho = \dot{w}_{jj}^i - \frac{\bar{\alpha}}{V_{jj}^i} \left[V_{ijj}^i + V_{jjj}^i + (n-2)\, V_{jjk}^i\right] \dot{a}\,(t) + \frac{1}{V_{jj}^i} \pi_{jj}^i\,(t)$$
$$+ \lambda\,(n-2) \left\{ \frac{V_{jj}^i \beta^2}{V_{jj}^i} \left[w_{jj}^i\,(t) + 2b\,(t)\right] + \frac{2 V_{jk}^i \beta}{V_{jj}^i} \left[w_{jk}^i\,(t) + b\,(t)\right] \right\}$$
$$+ \lambda \left\{ \frac{V_{ii}^i \beta^2}{V_{jj}^i} \left[w_{ii}^i\,(t) + 2b\,(t)\right] + \frac{2 V_{ij}^i \beta}{V_{jj}^i} \left[w_{ij}^i\,(t) + b\,(t)\right] \right\}$$

$$(\rho+2\lambda)\, w_{jk}^i\,(t) + i_t - \rho = \dot{w}_{jk}^i - \frac{\bar{\alpha}}{V_{jk}^i} \left[V_{ijk}^i + V_{jjk}^i + V_{jkk}^i + (n-3)\, V_{jkl}^i\right] \dot{a}\,(t) + \frac{1}{V_{jk}^i} \pi_{jk}^i\,(t)$$
$$+ \lambda\,(n-3) \left\{ \frac{V_{jj}^i \beta^2}{V_{jk}^i} \left[w_{jj}^i\,(t) + 2b\,(t)\right] + \frac{2 V_{jk}^i \beta}{V_{jk}^i} \left[w_{jk}^i\,(t) + b\,(t)\right] \right\}$$
$$+ \lambda \left\{ \frac{V_{ii}^i \beta^2}{V_{jk}^i} \left[w_{ii}^i\,(t) + 2b\,(t)\right] + \frac{2 V_{ij}^i \beta}{V_{jk}^i} \left[w_{ij}^i\,(t) + b\,(t)\right] \right\}$$

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_j^i. From the optimality conditions we have

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

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

$$V_{iij}^i \left[1 - (n-1)\,\beta\right] + (n-1)\, V_{ijj}^i - \beta V_{iii} = 0 \tag{A.26}$$

where $V_{iij}, V_{ijj}, V_{iii}$ are solutions to the system (A.24). 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

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

After log-linearization we have

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

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

$$\dot{\mathbf{Y}}(t) = \mathbf{A}\mathbf{Y}(t) + \mathbf{Z}_k k(t) + \mathbf{Z}_c c(t) + \mathbf{Z}_i [i(t) - \rho]$$

where $\mathbf{A} \in \mathbb{R}^{3\times 3}$, $\mathbf{Z}_k, \mathbf{Z}_c, \mathbf{Z}_i \in \mathbb{R}^3$ 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

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

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 6 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 $\mathbf{B}(t) = \mathbf{Z}_k k(t) + \mathbf{Z}_c c(t) + \mathbf{Z}_r [r(t) - \rho]$. 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

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

Take any vector $\alpha^\pi = \left(\alpha_j^\pi\right)_{j=1}^{K}$ 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

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

and we can define $\alpha_0^\pi = -\sum_{k=1}^{K}\alpha_k^\pi\left[\mathbf{A}^k\right]_{11}$. This simplifies to

$$\begin{aligned} \dddot{\pi} = {} & \left(\mathbf{A}_{\pi\pi} + \mathbf{A}_{ww}\right)\ddot{\pi} \\ & + \left(\mathbf{A}_{\pi p} + \mathbf{A}_{\pi w}\mathbf{A}_{w\pi} - \mathbf{A}_{\pi\pi}\mathbf{A}_{ww}\right)\dot{\pi} \\ & + \left(\mathbf{A}_{\pi w}\mathbf{A}_{wp} - \mathbf{A}_{\pi p}\mathbf{A}_{ww}\right)\pi \\ & + \mathbf{A}_{\pi w}\dot{\mathbf{B}}_w + \ddot{\mathbf{B}}_\pi - \mathbf{A}_{ww}\dot{\mathbf{B}}_\pi \end{aligned} \tag{A.27}$$

G.1 One-time shocks

Given (15) we can guess and verify that $x = \psi_x e^{-\xi t}$ for all variables $x \in \{\pi, k, c, r-\rho, i-\rho\}$ and solve for the coefficients ψ_x using the system

$$\begin{aligned} \psi_\pi\left(\gamma_0^\pi - \gamma_1^\pi\xi + \gamma_2^\pi\xi^2 - \gamma_3^\pi\xi^3\right) &= \psi_k\left(\gamma_0^k - \gamma_1^k\xi + \gamma_2^k\xi^2\right) \\ &\quad + \psi_c\left(\gamma_0^c - \gamma_1^c\xi + \gamma_2^c\xi^2\right) \\ &\quad + \left(\psi_i - \psi_\pi\right)\left(\gamma_0^r - \gamma_1^r\xi + \gamma_2^r\xi^2\right) \\ -\xi\psi_c &= \sigma^{-1}\left(\psi_i - \psi_\pi - \epsilon_0^r\right) \\ \psi_i &= \phi_\pi\psi_\pi + \epsilon_0^m + (1-\kappa)\,\epsilon_0^r \end{aligned}$$

Thus

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

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

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

H Additional Tables and Figures

Table A.1: Standard deviations of inflation and consumption.

Number of firms n	Model	Std. dev. of π (%)		Std. dev. of c (%)	
		ϵ^r	ϵ^m	ϵ^r	ϵ^m
∞	Standard NK (CES)	2.2	2.7	0.8	1.0
∞	Klenow-Willis $\theta = 10$	2.0	2.4	1.0	1.3
3	MPE	1.3	1.6	0.8	1.0
	Non-strategic	1.8	2.2	1.4	1.8
10	MPE	2.2	2.8	1.1	1.4
	Non-strategic	2.6	3.2	1.3	1.7

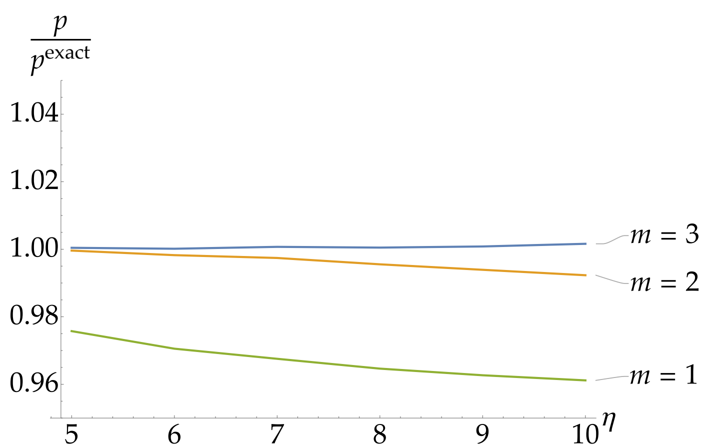

Figure A.1: Steady state markup p with $n = 2$ firms, under our solution method with $m = 1, 2, 3$, relative to exact solution p^{exact} (which corresponds to $m = \infty$).

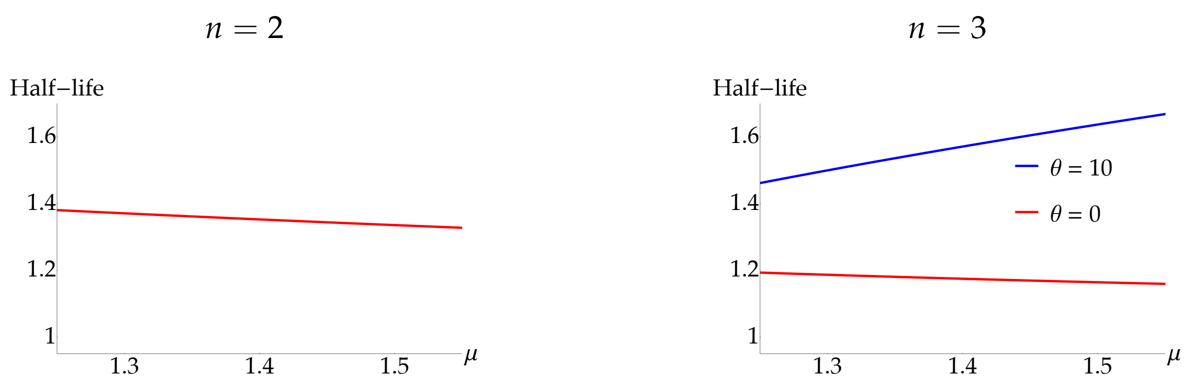

Figure A.2: Half-life as a function of resulting steady state markup when η varies.

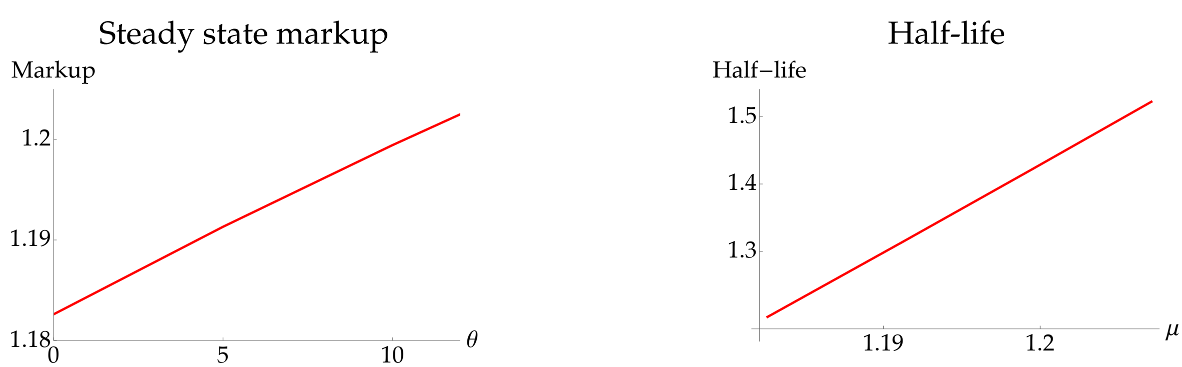

Figure A.3: Markup and half-life when θ varies in a model with $n = 3$ and $\eta = 10$.

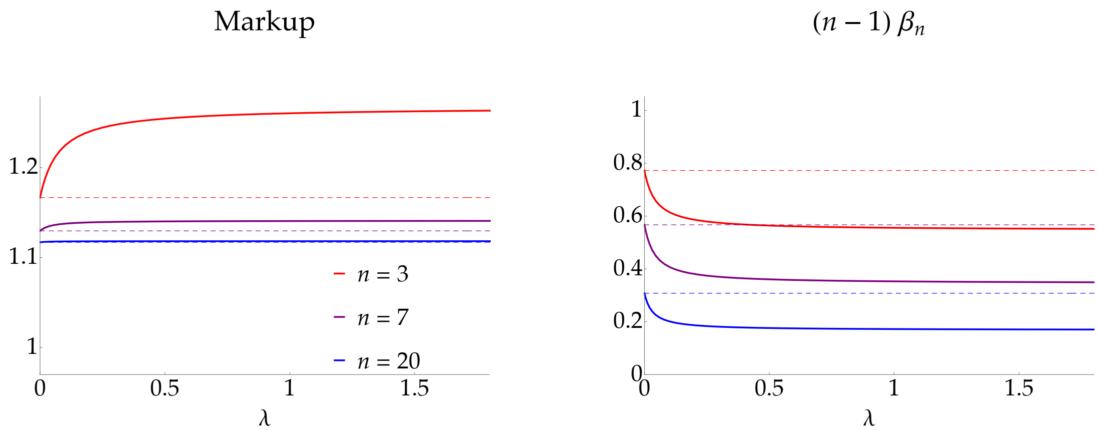

Figure A.4: Steady state markup and slope (weighted by number of rivals) as a function of frequency of price changes λ.

Note: The dashed horizontal lines show the markups and slopes of best-responses in the static Bertrand-Nash equilibrium, which corresponds to the dynamic model with $\lambda = 0$.

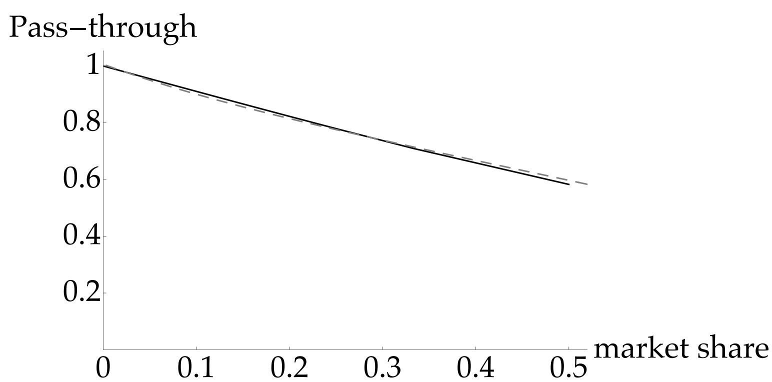

Figure A.5: Pass-through as a function of market share.

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 $n = 4$ of firms.

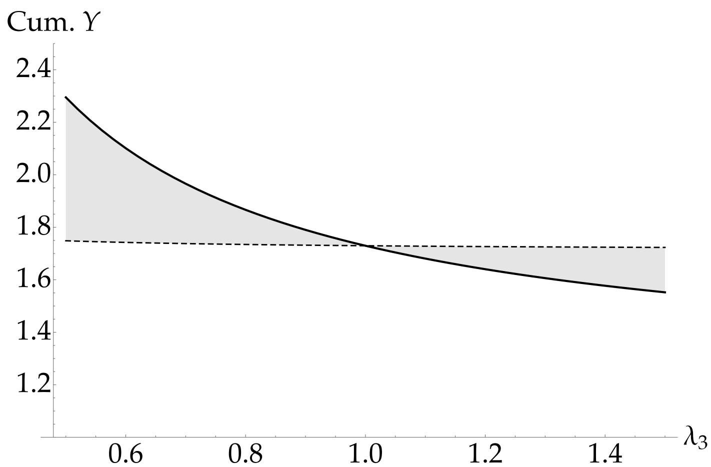

Figure A.6: Effects of heterogeneous frequency across sectors. The black line shows the total cumulative output effect (12) and the gray area shows the covariance term.

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.

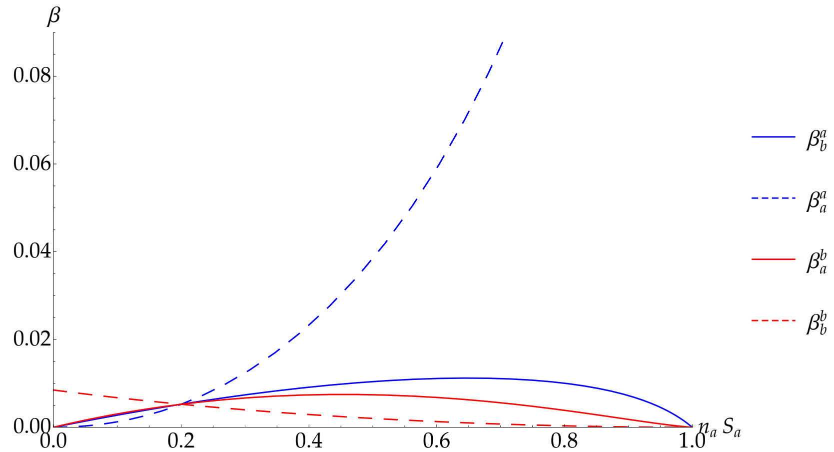

Figure A.7: Slopes $\beta_j^i = \frac{\partial g^i}{\partial p_j}$ for $i, j \in \{a, b\}$ around the steady state as a function of market share.

Note: The x-axis is the total market share of type-a firms, $n_a S_a$ (where S_a is the market share of one type-a firm). There are 10 heterogeneous firms with $n_a = 2, n_b = 8$. We vary market shares by changing the relative demand ξ_a/ξ_b. The y-scale is truncated at 0.1 for readability but β_a^a increases to 0.27 as $n_a S_a \to 1$.

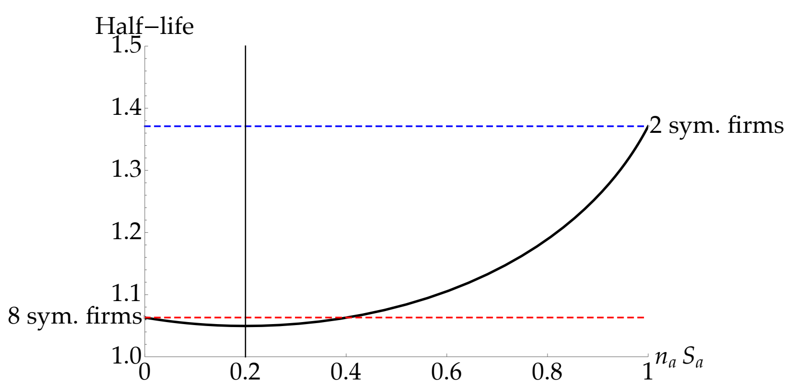

Figure A.8: Half-life of the aggregate price level as a function of market share.

Note: The x-axis is the total market share of type-a firms, $n_a S_a$ (where S_a is the market share of one type-a firm). There are 10 heterogeneous firms with $n_a = 2, n_b = 8$. We vary market shares by changing the relative demand ξ_a/ξ_b. The horizontal dashed lines show the half-lives with n_a symmetric firms (in blue) and n_b symmetric firms (in red), respectively. They correspond to the extreme cases of the heterogeneous firms model with one type of firm becoming dominant relative to the other. The vertical line denotes the case $\xi_a/\xi_b = 1$ which corresponds to a symmetric firms model with $n_a + n_b = 10$ firms.

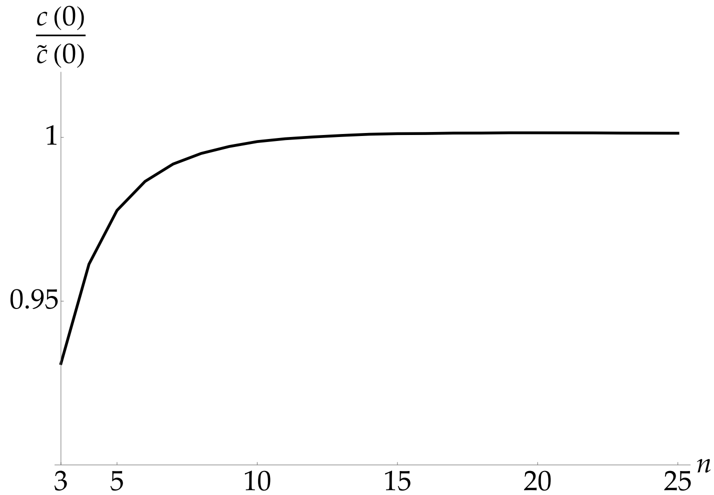

Figure A.9: Impact effect of a $\epsilon_0^m = -1\%$ monetary shock on consumption relative to non-strategic model $c(0)/\tilde{c}(0)$ as a function of number of firms n under AIK calibration.

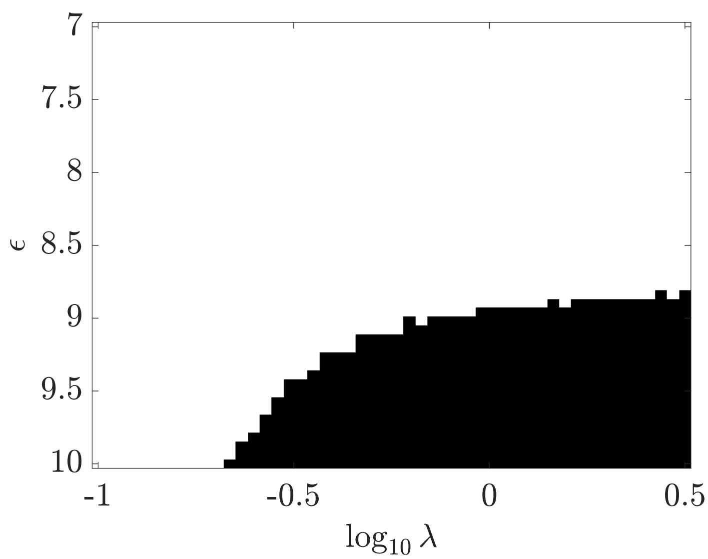

Figure A.10: In white: convergence of value function iteration algorithm towards a monotone MPE in (λ, ϵ) space, with $n = 2$ firms.