American Economic Review 2022, 112(8): 2815–2849
https://doi.org/10.1257/aer.20201739

Dynamic Oligopoly and Price Stickiness†

By Olivier Wang and Iván Werning*

How does market concentration affect the potency of monetary policy? To address this question we embed a dynamic oligopolistic game into a general-equilibrium macroeconomic model. We provide a sufficient-statistic formula for the response to monetary shocks involving demand elasticities, concentration and markups. We discipline our model with evidence on pass-through and find that higher concentration amplifies nonneutrality and stickiness. We isolate strategic effects from oligopoly by comparing our model to one with naive firms. We derive an exact Phillips curve featuring novel higher-order terms, but show that a standard New Keynesian one recalibrated with higher stickiness provides an excellent approximation. (*JEL* D43, E12, E21, E31, E43, E51, E52)

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

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

*Wang: Stern School of Business, New York University (email: olivier.wang@nyu.edu); Werning: Department of Economics, Massachusetts Institute of Technology (email: iwerning@mit.edu). Emi Nakamura was the coeditor for this article. In memory of Julio Rotemberg, who was ahead of his time, many times. We thank excellent research assistance by Marc de la Barrera, Rebekah Dix, Juliette Fournier, Alex Carrasco, and Pedro Martinez Bruera, as well as comments by Fernando Alvarez, Anmol Bhandari, Ariel Burstein, Glenn Ellison, Fabio Ghironi, Francesco Lippi, Simon Mongey, and seminar and conference participants at MIT, UTDT lecture in honor of Julio Rotemberg, EIEF Rome, SED 2018 meetings, Minneapolis Federal Reserve Bank, NBER-SI 2020, Central Bank of Chile, Wisconsin, NYU, PSE, Wharton, Bank of Israel, Yale, and Harvard. All remaining errors are ours.

†Go to https://doi.org/10.1257/aer.20201739 to visit the article page for additional materials and author disclosure statements.

[1] For instance, Gutiérrez and Philippon (2017) document an increase in the mean Herfindahl-Hirshman index since the mid-nineties, and argue that it has weakened investment. Autor et al. (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.

every firm is a function of the prices of its competitors. We use this model to study the aggregate real effects of monetary shocks.

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

Given the importance of this slope, we next investigate its determinants. We provide a formula for it that inputs three sufficient statistics: (i) market concentration as captured by the effective number of firms within a sector, (ii) demand elasticities, and (iii) markups.[2] According to our formula, all other things the same, higher observed markups predict larger output responses to monetary shocks. At the heart of this result is the notion that, away from the monopolistic limit, markups reveal the strength of strategic complementarities. This is the case because greater strategic complementarities lead to higher markups in our dynamic oligopoly game.

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

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

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

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

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

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

higher concentration amplifies aggregate stickiness if the superelasticity (the elasticity of the elasticity) of demand is low, as under CES, but has the opposite effect if the superelasticity is high. Thus, it is essential to first understand the link between concentration and finer properties of demand functions.

We use evidence on the heterogeneity in the pass-through of idiosyncratic cost shocks across small and large firms from Amiti, Itskhoki, and Konings (2019) to calibrate how concentration affects the superelasticity of demand, and find that concentration amplifies stickiness substantially, even more than under CES: the half-life doubles when going from $n = \infty$ to $n = 3$. Translating, this corresponds to dividing the slope of the Phillips curve by four. Under this calibration the rise in the average Herfindahl-Hirschman index (HHI) observed in the United States since 1990 increases the half-life by 15 percent, or a one-third reduction in the slope of the Phillips curve.

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

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

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

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

Relative to the standard New Keynesian model, which is a simply first-order dynamical system, our Phillips curve includes higher-order terms. Thus, these terms can, in principle, generate endogenous inflation persistence and cost-push shocks. However, we find, again, that for a wide range of shocks the equilibrium with naïve firms provides an accurate approximation to the strategic model. Since the naïve

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

model is equivalent to a monopolistic setup, this implies that a standard (first-order) New Keynesian Phillips curve provides an excellent approximation to the actual (higher-order) Phillips curve.

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

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

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

Related Literature

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

The first paper to combine nonmonopolistic competition and nominal rigidities in general equilibrium is Mongey (2018). This paper uses a rich quantitative model with two firms, menu costs, and idiosyncratic shocks to show that duopoly can generate significant nonneutrality relative to the Golosov and Lucas (2007) benchmark. It also finds that duopoly is closer to monopolistic competition under Calvo price-setting than with menu costs. Our paper takes a complementary approach, more analytical but assuming Calvo pricing and abstracting from idiosyncratic shocks.[7] This

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

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

allows us to go beyond two firms and explore different questions, in particular by changing industry concentration, separating strategic complementarities from residual demand effects, and allowing for arbitrary shocks through the Phillips curve.[8] Modeling more than two firms also lets us incorporate recent evidence linking cost pass-through and market shares from Amiti, Itskhoki, and Konings (2019) to infer the relation between concentration and monetary nonneutrality. This evidence implies that even under Calvo pricing, oligopoly leads to significant amplification.

The literature on variable markups in international trade highlights the importance of market structure for cost (e.g., exchange rate) pass-through in static settings (e.g., Atkeson and Burstein 2008). We study a dynamic general equilibrium version of these models, as is needed to analyze monetary policy, and show how to map pass-through estimates to aggregate effects of monetary policy. Our results also share some of the mechanisms studied in partial equilibrium in the IO 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).

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

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

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

[8] Several papers, such as Benigno and Faia (2016); Corhay, Kung, and Schmid (2020)—with Rotemberg pricing— and Etro and Rossi (2015); 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.

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

I. A Macro Model with Oligopolies

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

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

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

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

Household Preferences: Utility is given by

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

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

$$C = \left(\int_S C_s^{1-\frac{1}{\omega}} ds\right)^{1/\left(1-\frac{1}{\omega}\right)},$$

$$C_s = H_s\Big(\{c_{i,s}\}_{i \in I_s}\Big),$$

with $C = \exp\int_S \log C_s ds$ when $\omega = 1$ where $\{C_s\}$ and $\{c_{i,s}\}$ are sectoral consumption across sectors $s \in S$ and good varieties across firms $i \in I_s$ within each sector. H_s is homogeneous of degree one and can be more general than CES (e.g., Kimball 1995).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where f is increasing and differentiable. If f is concave then this captures decreasing returns. We first assume no differences in productivity within sectors, so that the n_s firms are symmetric, and thus normalize to $z_{i,s} = 1$.[12] Section V extends the analysis to heterogeneous firms.

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

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

Profits for firm i are

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

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

$$E_0 \int_0^{\infty} e^{-\int_0^t R(s)ds} \Pi^{i,s}(p_{i,s}(t), p_{-i,s}(t); t) dt.$$

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

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

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

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

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

Equilibrium Definition: Given initial prices $\{p_{i,s}(0)\}$, an equilibrium is given by paths for the aggregate price $P(t)$, wage $W(t)$, interest rate $R(t)$, consumption $C(t)$, labor $\ell(t)$ and money supply $M(t)$, as well as demand functions for consumers $d^{i,s}$ and strategy functions for firms $g^{i,s}$ such that: (i) consumers optimize quantities taking as given the sequence of prices and interest rates; (ii) each firm's reset price strategy $g^{i,s}$ is optimal, given the path for $P(t)$, $C(t)$, its rivals' strategies $g^{j,s}$ and the demand functions d; (iii) consistency: the aggregate price level evolves in accordance with the reset strategy g employed by firms; (iv) markets clear: firms meet demand for goods, $y_{i,s}(t) = c_{i,s}(t)$, the supply of labor equals demand

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

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

A. *Stationary Markov Equilibrium*

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

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

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

where $g(p_{-i}) \in \operatorname{argmax}_{p_i} V^i(p_i, p_{-i})$ with necessary condition

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

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

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

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

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

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

Thus, B represents the reaction to average rival prices.

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

Limit $\lambda/\rho \rightarrow$ *0*: **Static Bertrand-Nash Equilibrium:** When prices are infinitely sticky or firms are infinitely impatient, so that $\lambda/\rho \rightarrow 0$ then (1) implies that

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

$V(p) \rightarrow \Pi(p)$ and, thus, firms play a static best-response. Intuitively, they take the current prices of other firms as fixed forever. The equilibrium then converges to a static Bertrand-Nash:

$$\lim_{\lambda/\rho \rightarrow 0} g(p_{-i}) = g^{Nash}(p_{-i}) = \operatorname*{argmax}_{p_i} \Pi^i(p_i, p_{-i}).$$

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

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

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

II. Monetary Shocks: Dynamics and Sufficient Statistics

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

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

A. *Exact Dynamics: Partial Equilibrium To General Equilibrium*

Firms must forecast the path that macroeconomic variables will take after the shock. Any given path for aggregates determines a Markov reset price strategy $g^{i,s}$.

[14] Figure J6 displays the locus of existence of these monotone equilibria in the (λ, η)-space (where η is the within-sector elasticity of substitution). While the curse of dimensionality prevents us from solving numerically for the full MPE with general n, we conjecture that the region of existence of these equilibria increases with the number of firms, since a higher n reduces the potential monopoly profit (the case of monopolistic competition $n \rightarrow \infty$ 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.

These strategies, in turn, determine the evolution of aggregates. It is possible to solve this fixed-point problem quite generally, as we do in Section VI, but we first focus on a simple case. In the spirit of Golosov and Lucas (2007), our assumptions on preferences lead to the following simplification.

PROPOSITION 1: *Equilibrium aggregates satisfy*

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

If in addition

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

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

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

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

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

The classic paper by Rotemberg and Saloner (1986) studied a partial equilibrium model of oligopoly, facing exogenous fluctuations in demand without price rigidities. They assumed a fixed real interest rate. Their analysis focused on non-Markov trigger strategies that sustain "collusive" prices in bad times, but lead to price wars during booms (creating an amplification mechanism for output). In their model, price wars occur because booms are periods with higher demand and, thus, high temporary profits to compete over, so the incentive to raise prices is greater. A similar effect is generally present in our model. However, Proposition 1 shows that when

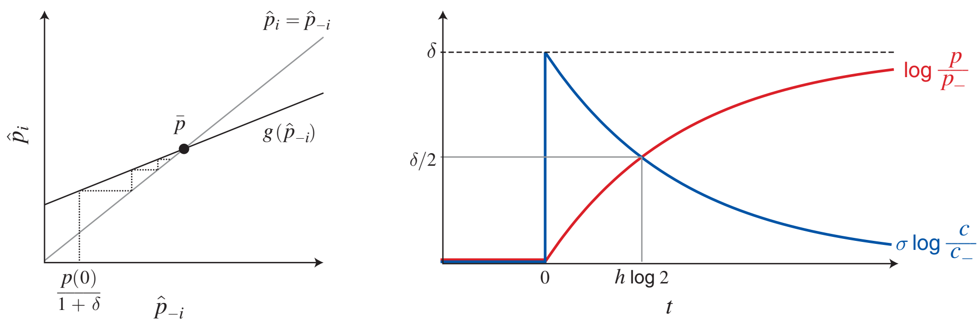

Figure 1. Dynamics Following a Monetary Shock δ.

Notes: Left: price dynamics within a sector. Right: aggregate dynamics. The half-life of the price level is denoted by $h\log 2$.

$\omega\sigma = 1$, these effects may not be present in general equilibrium since $P(t)$ moves in the opposite direction. An equivalent way to describe this is that Rotemberg and Saloner (1986) assumed a fixed real interest rate. But in our model, a boom lowers real interest rates, which makes firms care about the future more, exactly counterbalancing the increase in stakes from higher current profits.

B. *Approximate Dynamics: The Importance of B_s*

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

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

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

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

where $\zeta_s = \bar{P}_s \bar{C}_s / (\bar{P}\bar{C})$ is the steady state expenditure share of sector s.[15] *The present-value output effect of the shock discounted at any rate $r \geq 0$ is*

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

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

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

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

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

C. *On Half-lives and Phillips Curves*

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

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

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

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

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

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

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

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

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

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

Up to constants of proportionality and for ρ small

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

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

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

D. *Sufficient Statistics for B: Markups and Elasticities*

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

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

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

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

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

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

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

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

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

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

III. Market Concentration and other Comparative Statics

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

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

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

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

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

A. *Preliminaries: Method, Elasticity, and Superelasticity*

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

Our method selects Kimball preferences that generate an equilibrium that can be solved analytically locally around the steady state. Crucially, we can match any desired elasticities, superelasticities, and higher order elasticities of the demand function d^i, up to any desired order m. We employ $m = 2$ on the grounds that this is sufficient to flexibly match the first two derivatives, which are the only ones indirectly estimated in practice (using pass-through regressions, as we discuss later).[22] Equivalently, we are matching the elasticity ϵ and the "superelasticity" Σ of demand at the steady state:[23]

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

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

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

In online Appendix E, we derive an additional equation

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

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

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

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

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

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

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

we obtain the standard CES aggregator across firms. Letting $\Phi_s(x) = -\frac{\phi_s'(x)}{x\phi_s''(x)}$ it is standard to define

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

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

Then we have (dropping the s subscript):

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

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

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

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

B. *Market Concentration*

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

Exogenous Changes in Number of Firms: We first hold preferences, embedded in the Kimball aggregator ϕ_s, fixed when changing n. Therefore the parameters η and θ are fixed, but the elasticities ϵ and Σ will change with n according to (10)–(11). One interpretation is that the set of varieties demanded by consumers is unchanged,

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

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

but concentration increases due to mergers and acquisitions: firms expand and get to set prices for more varieties. We assume constant returns to scale in production, $f(\ell) = \ell$, to focus on complementarities coming from the demand side. The remaining parameters are described in Table 1.

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

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

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

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

A Calibration Based on Pass-Through: Previously, we fixed the "preference parameters" η, θ when changing the number of firms n. We now provide an alternative that does not hold preferences ϕ_s fixed as we change the number of firms. We have seen that the shape of demand is crucial to understand how market structure impacts the transmission of aggregate monetary shocks. As Atkeson and Burstein (2008) emphasized in a static setting, market structure also affects the pass-through of firm-level cost shocks, hereafter, simply, "pass-through." We now argue that calibrating the model to match the empirical relation between market

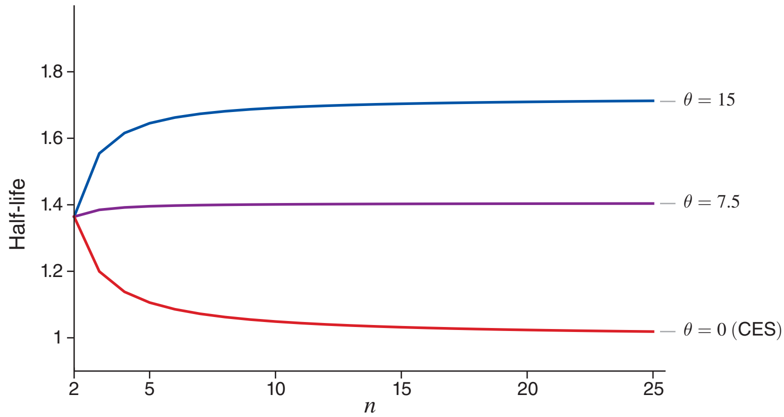

Figure 2. Half-Life as a Function of n for Different Values of θ.

share and pass-through implies that concentration significantly amplifies monetary nonneutrality, even more than under CES.[25]

Amiti, Itskhoki, and Konings (2019) estimate pass-through regressions,

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

separately for small and large firms, and find considerable heterogeneity in pass-through. Small firms behave as if 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 one-half of their own cost shocks, thus letting their markup decline to absorb the other half.[26] Amiti, Itskhoki, and Konings (2019) show that $\hat{\alpha}$ as a function of market share s is well approximated by

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

with $\eta = 10$ and $\omega = 1$ (equation (12) and calibration p. 2398 in their paper).[27] Empirical variation in s captures both differences in concentration across sectors

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

[26] Other papers such as Berman, Martin, and Mayer (2012); Chatterjee, Dix-Carneiro, and Vichyanond (2013); and Auer and Schoenle (2016) also find lower pass-through for larger firms.

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

and heterogeneity across firms within sectors. We will associate the share under our symmetric model to concentration: $s = 1/n$.[28]

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

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

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

PROPOSITION 5: *There exist unique scalars*

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

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

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

Remark 2.—The mapping $(\alpha, B) \mapsto (\hat{\alpha}, \hat{B})$ given by Proposition 5 cannot be inverted to obtain directly α, B as functions of empirical estimates $\hat{\alpha}, \hat{B}$. That is, for any $\hat{\alpha}, \hat{B}$ such that $\hat{\alpha} + \hat{B} = 1$ (a condition that $\hat{\alpha}$ and $\hat{B}$ in (15) must satisfy), the system (15) does not identify α, B.

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

[28] Pass-through as a function of market share is essentially the same, whether variation in market share comes from varying the number n of symmetric firms, or from within-sector heterogeneity among a fixed number of firms. This equivalence is exact in a static model, and holds approximately in our dynamic model as shown in Figure J1. In Section V we show that the equivalence also holds for the response to aggregate monetary shocks.

[29] The coefficients α, B, γ can be computed as before using envelope conditions applied to a generalization of the Bellman equation (49).

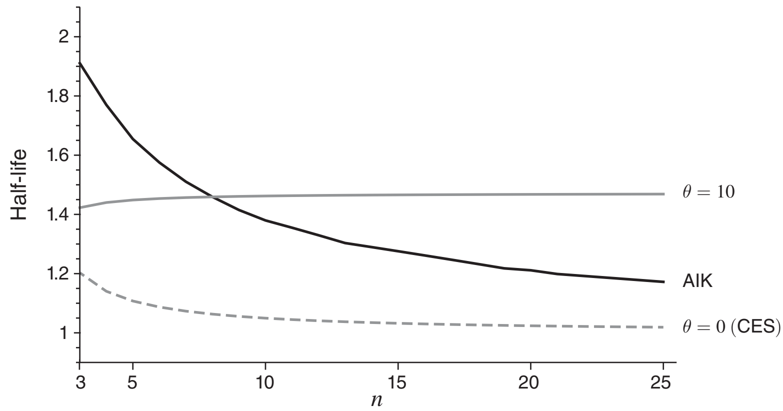

Figure 3. Half-Life as a Function of the Number of Firms n

Note: AIK: variable θ_n to match pass-through estimates from Amiti, Itskhoki, and Konings (2019).

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

Figure 3 shows our results. "AIK" is our calibration with a variable parameter θ_n as explained above. Concentration amplifies stickiness substantially, much more than under CES. When going from monopolistic competition to $n = 3$ firms, the half-life doubles; equivalently, the slope of the Phillips curve is divided by four. For comparison we include calibrations that hold θ fixed.

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

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

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

C. *Other Comparative Statics: the Determinants of Markups*

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

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

We argued that under dynamic oligopoly, markups are not fully determined by demand elasticities. Figure 4 illustrates this point. A higher superelasticity parameter θ increases the markup. Since this leaves ϵ unchanged, this experiment is the most transparent application of our sufficient statistic result: B, and thus, the half-life increase. Finally, markups increase with the frequency of price changes λ and decrease with discount rates ρ: more patient firms can sustain a higher markup, as in the literature on collusion.

IV. Inspecting the Mechanism: Strategic versus Naïve Firms

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

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

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

Therefore the naïve model provides a behavioral foundation for the notion of a "properly recalibrated" monopolistic economy.[32] We compute the half-life of the

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

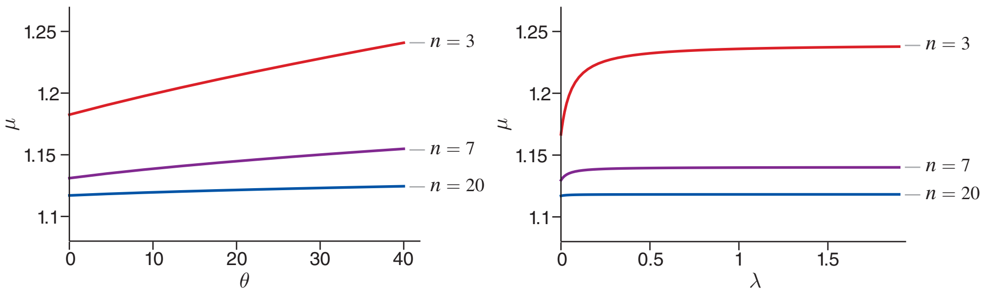

FIGURE 4. STEADY STATE MARKUP μ AS A FUNCTION OF θ (LEFT PANEL) AND λ (RIGHT PANEL).

price level h in the strategic model and $h^{Naïve}$ in the naïve model, and then define strategic effects as follows:

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

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

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

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

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

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

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

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

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

PROPOSITION 7: *Let* $\varphi(x,y) = 1 + 1/(2y) - \sqrt{\left[1 + 1/(2y)\right]^2 - (1 + 1/y)x}$. *Then,*

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

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

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

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

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

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

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

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

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

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

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

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

Comparative Statics in the Naïve Model: The effect of oligopoly on monetary policy transmission is transparent in the naïve model because changes in n, θ, η

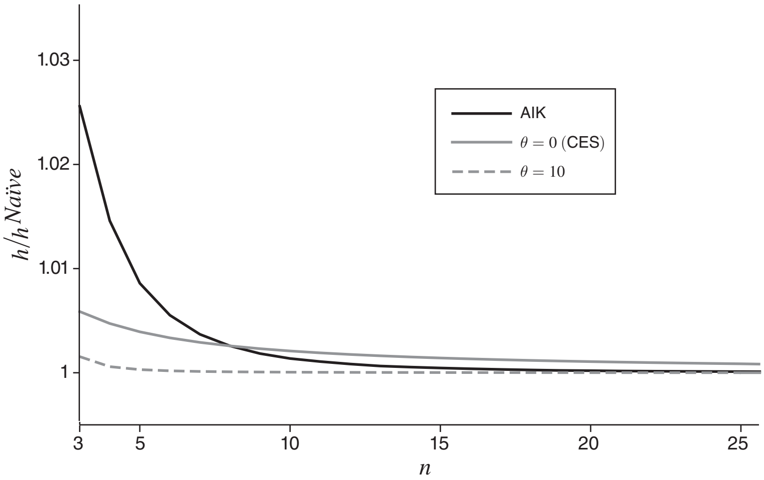

Figure 5. Strategic Effect $h/h^{Naïve}$ as a Function of n

Note: AIK: variable θ_n to match pass-through estimates from Amiti, Itskhoki, and Konings (2019).

affect $B^{Naïve}$ only through B^{Nash}. Going back to the findings from Section IIIC, the naïve model helps understand, for instance, when concentration amplifies or dampens aggregate price stickiness. Figure 2 shows that the effect of n in the strategic model depends on the value of θ; in the naïve model we can prove this property analytically and understand better the underlying intuition.

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

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

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

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

[34] Similarly, holding n fixed, B^{Nash} decreases with the elasticity of substitution η (and thus the observed markup) if and only if $\theta < \frac{n}{n-2} \times \frac{(\eta - 1)^2}{1 + (n-1)\eta^2}$, which explains why, in Figure G1, the half-life decreases with the markup μ under CES ($\theta = 0$) but not when θ is high enough.

V. 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 nonneutrality. 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-Hirschman index.

A. *Heterogeneous Price Stickiness across Sectors*

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

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

where $E[x_s] = \int_s \zeta_s x_s \, ds$ denotes the average of a variable x across sectors. In Section IIIB we saw that $\frac{1}{1 - B_s}$ is higher in more concentrated sectors. If these sectors are also characterized by a lower frequency λ_s, then the covariance term is positive, which increases aggregate nonneutrality. This channel differs from the role of heterogeneity in, e.g., Carvalho (2006) under Calvo pricing or Nakamura and Steinsson (2010) under menu costs. Oligopoly is a very natural reason to have heterogeneity in B_s, but note that even under monopolistic competition $n_s \to \infty$, heterogeneous B_s could arise from differences in Kimball demand or the degree of decreasing returns to scale across sectors.

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

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

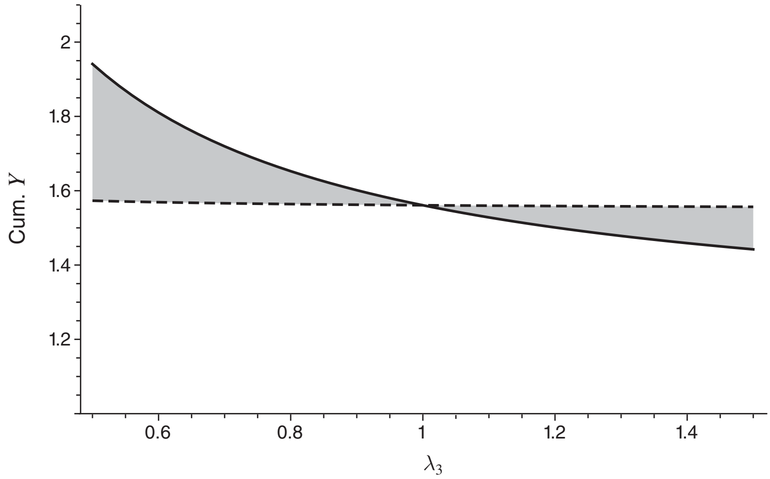

FIGURE 6. EFFECTS OF HETEROGENEOUS FREQUENCY ACROSS SECTORS

Notes: The black line shows the cumulative output effect and the gray area shows the covariance term in (18). Example with two sectors, one with $n = 3$ firms and one with $n = 20$ firms, with AIK calibration. The frequency in the sector with 3 firms is λ_3, and we set the frequency λ_{20} in the other sector to keep the average duration $\frac{1}{2}(1/\lambda_3 + 1/\lambda_{20})$ fixed at 1.

B. *Heterogeneous Firm Productivity within Sectors*

We now extend our baseline model to allow for permanent heterogeneity between firms *within* sectors, in terms of tastes and productivity. Focusing on one sector s, suppose that firms differ in their productivity z_i while consumers have different tastes captured by multiplicative demand shifters ξ_i.[36]

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

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

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

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

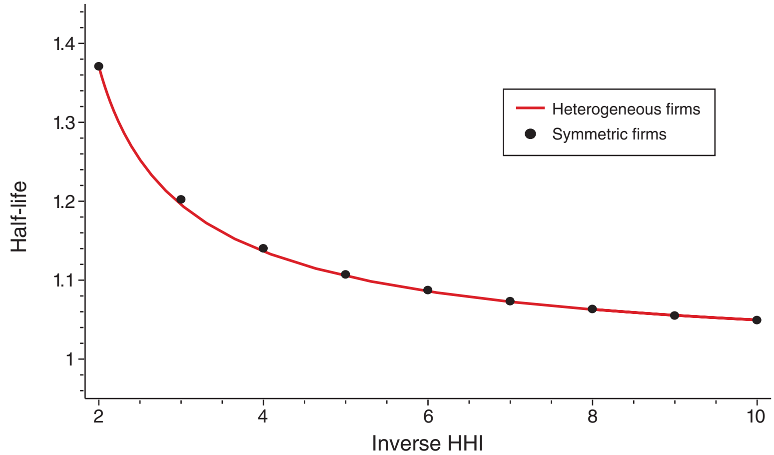

Figure 7. Heterogeneous versus Symmetric Firms

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

VI. The Oligopolistic Phillips Curve

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

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

A. *The Phillips Curve*

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

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

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

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

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

Under monopolistic competition, (19) simplifies drastically to

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

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

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

with $\kappa^{\text{Naïve}}(n) = \left[1 - B^{\text{Nash}}(n)\right]\lambda(\lambda + \rho)$.

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

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

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

[38] Figure J2 illustrates the coefficients $\gamma^{mc}(s), \gamma^c(s), \gamma^R(s)$ for different n.

[39] In general there are $q = 7$ eigenvalues, but q can be reduced to 3 under a simplifying condition (58) given in online Appendix I, which we assume to ease the exposition of (22); inflation dynamics are almost unchanged when we use $q = 7$, as we do in all the figures.

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

$$\kappa^{Naïve}(3) = 0.25,$$

$$\kappa^{Naïve}(\infty) = 1.05.$$

Going from $n = \infty$ to $n = 3$ reduces $\kappa^{Naïve}$ by a factor of *four*; in this sense the amplification from oligopoly appears very large. This result is consistent with Figure 3 in which the half-life h doubles when going from $n = \infty$ to $n = 3$, given the relation $\kappa \propto 1/h^2$ from Proposition 3.[40]

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

B. *Three-Equation Model*

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

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

and a monetary policy interest rate rule

$$R = \max\{0, \rho + \phi_\pi \pi + \epsilon^m\},$$

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

Date-0 Monetary Policy Shocks: We first consider unanticipated date-0 interest rate shocks that decay geometrically, $\epsilon^m(t) = \epsilon_0^m e^{-\xi t}$, while shutting other shocks, $\epsilon^r = 0$, so that the zero lower bound remains slack. The solution is detailed in online Appendix I. Under both monopolistic and oligopolistic competition, all the equilibrium variables $x \in \{c, \pi, mc, R - \rho\}$ are proportional to $e^{-\xi t}$, e.g., $x(t) = x(0)e^{-\xi t}$, hence differences across models are summarized by the

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

impact effects on consumption $c(0)$ and inflation $\pi(0)$.[41] Since $mc(0) = (\psi + \sigma) c(0)$, then

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

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

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

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

$$\epsilon^m(t) = \begin{cases} 0 & t < t_{shock}; \\ \epsilon_0^m e^{-\xi(t - t_{shock})} & t \geq t_{shock}. \end{cases}$$

This captures a news shock about monetary policy (the same results obtain with news about real shocks ϵ^r). Figure J3 shows the similarity of the impulse responses in the strategic and naïve models, for different shock dates $t_{\text{shock}} = 1, 2, 3$.

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

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

At $t = T$ the economy reverts to the steady state with $c(T) = \pi(T) = 0$ (for instance because the central bank lacks commitment). Figure J4 shows the impact effects $c(0)$ and $\pi(0)$ as a function of the length of the trap T, for two economies, $n = 3$ and $n = \infty$. Just like higher price flexibility λ leads to more deflation and deeper recession (e.g., Werning 2012), for given λ the stickiness due to oligopoly significantly dampens the severity of the trap by weakening the deflationary response. Figure J5 compares the impact effects $c(0)$ and $\pi(0)$ for the strategic and

[41] One can recover the permanent money shock case from Proposition 2 by setting $\xi = \lambda(1 - B)$ since then $\phi_\pi \pi + \epsilon^m = 0$ so $R(t)$ is unchanged.

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

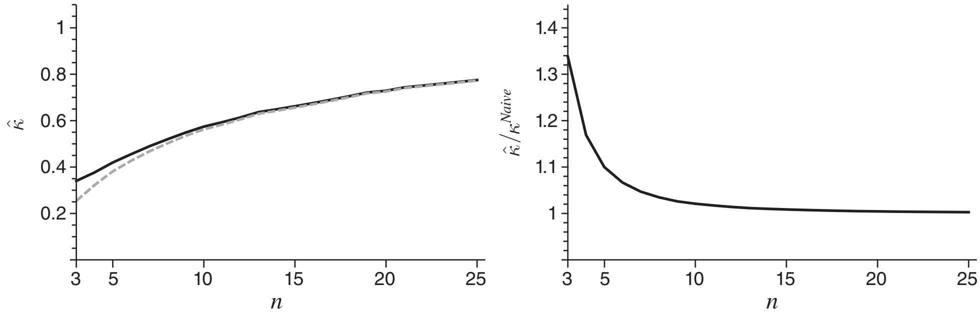

Figure 8. Effective Slope of the Phillips Curve $\hat{\kappa}$, Strategic versus Naïve Oligopoly

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

naïve models as a function of the length of the trap T: the responses are almost identical for short traps, but start diverging as T increases.

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

VII. Conclusion

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

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

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

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

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

Menu costs introduce several additional effects which we abstracted from here. Higher concentration may affect the frequency of price changes by reducing the profit losses from failing to adjust prices (e.g., Rotemberg and Saloner 1987). Moreover, for a given average frequency of price changes, concentration interacts with menu costs through two effects: the selection effect (i.e., which firms are more likely to adjust), and possibly by coordinating the price changes (i.e., increasing the correlation between different firms' price changes). In Mongey (2018), the first effect dominates and generates additional amplification from oligopoly relative to Calvo pricing. The second effect, on the other hand, works towards price flexibility. In extreme cases, this is what Nirei and Scheinkman (2021) call "repricing avalanches."

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

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

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

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

REFERENCES

Abreu, Dilip, David Pearce, and Ennio Stacchetti. 1990. "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring." *Econometrica* 58 (5): 1041–63.

Afrouzi, Hassan. 2020. "Strategic Inattention, Inflation Dynamics, and the Non-Neutrality of Money." Unpublished.
Alvarez, Fernando, and Francesco Lippi. 2014. "Price Setting with Menu Cost for Multiproduct Firms." *Econometrica* 82 (1): 89–135.
Alvarez, Fernando, Francesco Lippi, and Juan Passadore. 2016. "Are State- and Time-Dependent Models Really Different?" *NBER Macroeconomics Annual* 31 (1): 379–457.
Alvarez, Fernando, Hervé Le Bihan, and Francesco Lippi. 2016. "The Real Effects of Monetary Shocks in Sticky Price Models: A Sufficient Statistic Approach." *American Economic Review* 106 (10): 2817–51.
Amiti, Mary, Oleg Itskhoki, and Jozef Konings. 2019. "International Shocks, Variable Markups, and Domestic Prices." *Review of Economic Studies* 86 (6): 2356–2402.
Andrés, Javier, and Pablo Burriel. 2018. "Inflation and Optimal Monetary Policy in a Model with Firm Heterogeneity and Bertrand Competition." *European Economic Review* 103: 18–38.
Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review* 98 (5): 1998–2031.
Auer, Raphael, and Raphael Schoenle. 2016. "Market Structure and Exchange Rate Pass-Through." *Journal of International Economics* 98 (C): 60–77.
Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2017. "Concentrating on the Fall of the Labor Share." *American Economic Review* 107 (5): 180–85.
Barkai, Simcha. 2020. "Declining Labor and Capital Shares." *Journal of Finance* 75 (5): 2421–63.
Benigno, Pierpaolo, and Ester Faia. 2016. "Globalization, Pass-Through, and Inflation Dynamics." *International Journal of Central Banking* 12 (4): 263–306.
Berger, David, and Joseph Vavra. 2019. "Shocks versus Responsiveness: What Drives Time-Varying Dispersion?" *Journal of Political Economy* 127 (5): 2104–42.
Bergin, Paul R., and Robert C. Feenstra. 2000. "Staggered Price Setting, Translog Preferences, and Endogenous Persistence." *Journal of Monetary Economics* 45 (3): 657–80.
Berman, Nicolas, Philippe Martin, and Thierry Mayer. 2012. "How do Different Exporters React to Exchange Rate Changes?" *Quarterly Journal of Economics* 127 (1): 437–92.
Bilbiie, Florin O., Fabio Ghironi, and Marc J. Melitz. 2007. "Monetary Policy and Business Cycles with Endogenous Entry and Product Variety." NBER Working Paper 13199.
Burstein, Ariel, Vasco Carvalho, and Basile Grassi. 2020. "Bottom-Up Markup Fluctuations." Unpublished.
Carlton, Dennis W. 1986. "The Rigidity of Prices." *American Economic Review* 76 (4): 637–58.
Carvalho, Carlos. 2006. "Heterogeneity in Price Stickiness and the Real Effects of Monetary Shocks." *B.E. Journal of Macroeconomics* 2 (1): 1–56.
Chatterjee, Arpita, Rafael Dix-Carneiro, and Jade Vichyanond. 2013. "Multi-Product Firms and Exchange Rate Fluctuations." *American Economic Journal: Economic Policy* 5 (2): 77–110.
Corhay, Alexandre, Howard Kung, and Lukas Schmid. 2020. "Q: Risk, Rents, or Growth?" Unpublished.
DellaVigna, Stefano, and Matthew Gentzkow. 2019. "Uniform Pricing in U.S. Retail Chains." *Quarterly Journal of Economics* 134 (4): 2011–84.
Etro, Federico, and Lorenza Rossi. 2015. "New-Keynesian Phillips Curve with Bertrand Competition and Endogenous Entry." *Journal of Economic Dynamics and Control* 51: 318–40.
Gertler, Mark, and John Leahy. 2008. "A Phillips Curve with an Ss Foundation." *Journal of Political Economy* 116 (3): 533–72.
Goldberg, Pinelopi Koujianou, and Frank Verboven. 2001. "The Evolution of Price Dispersion in the European Car Market." *Review of Economic Studies* 68 (4): 811–48.
Golosov, Mikhail, and Robert E. Lucas. 2007. "Menu Costs and Phillips Curves." *Journal of Political Economy* 115 (2): 171–99.
Gopinath, Gita, and Oleg Itskhoki. 2010. "Frequency of Price Adjustment and Pass-Through." *Quarterly Journal of Economics* 125 (2): 675–727.
Gutiérrez, Germán, and Thomas Philippon. 2017. "Declining Competition and Investment in the U.S." NBER Working Paper 23583.
Kimball, Miles S. 1995. "The Quantitative Analytics of the Basic Neomonetarist Model." *Journal of Money, Credit and Banking* 27 (4): 1241–77.
Klenow, Peter J., and Jonathan L. Willis. 2016. "Real Rigidities and Nominal Price Changes." *Economica* 83 (331): 443–72.
Krusell, Per, Burhanettin Kuruscu, and Anthony A. Smith. 2002. "Equilibrium Welfare and Government Policy with Quasi-geometric Discounting." *Journal of Economic Theory* 105 (1): 42–72.
Levintal, Oren. 2018. "Taylor Projection: A New Solution Method for Dynamic General Equilibrium Models." *International Economic Review* 59 (3): 1345–73.

Maskin, Eric, and Jean Tirole. 1988. "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles." *Econometrica* 56 (3): 571–99.
Matsuyama, Kiminori, and Philip Uschchev. 2017. "Beyond CES: Three Alternative Classes of Flexible Homothetic Demand Systems." Unpublished.
Midrigan, Virgiliu. 2011. "Menu Costs, Multiproduct Firms, and Aggregate Fluctuations." *Econometrica* 79 (4): 1139–80.
Mongey, Simon. 2018. "Market Structure and Monetary Non-Neutrality." Unpublished.
Nakamura, Emi, and Dawit Zerom. 2010. "Accounting for Incomplete Pass-Through." *Review of Economic Studies* 77 (3): 1192–1230.
Nakamura, Emi, and Jón Steinsson. 2010. "Monetary Non-neutrality in a Multisector Menu Cost Model." *Quarterly Journal of Economics* 125 (3): 961–1013.
Nakamura, Emi, and Jón Steinsson. 2013. "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications." *Annual Review of Economics* 5 (1): 133–63.
Neiman, Brent. 2011. "A State-Dependent Model of Intermediate Goods Pricing." *Journal of International Economics* 85 (1): 1–13.
Nirei, Makoto, and José Scheinkman. 2021. "Repricing Avalanches." Unpublished.
Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter. 2020. "Diverging Trends in National and Local Concentration." *NBER Macroeconomics Annual* 35 (1): 115–50.
Rotemberg, Julio J., and Garth Saloner. 1986. "A Supergame-Theoretic Model of Price Wars during Booms." *American Economic Review* 76 (3): 390–407.
Rotemberg, Julio J., and Garth Saloner. 1987. "The Relative Rigidity of Monopoly Pricing." *American Economic Review* 77 (5): 917–26.
Rotemberg, Julio J., and Michael Woodford. 1992. "Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity." *Journal of Political Economy* 100 (6): 1153–1207.
Slade, Margaret E. 1998. "Optimal Pricing with Costly Adjustment: Evidence from Retail-Grocery Prices." *Review of Economic Studies* 65 (1): 87–107.
Weintraub, Gabriel Y., C. Lanier Benkard, and Benjamin Van Roy. 2008. "Markov Perfect Industry Dynamics with Many Firms." *Econometrica* 76 (6): 1375–1411.
Werning, Iván. 2012. "Managing a Liquidity Trap: Monetary and Fiscal Policy." Unpublished.