Dynamic Resource Misallocation and Amplification*

Marcus M. Opp   Christine A. Parlour   Johan Walden

December 30, 2012

Abstract

We develop a tractable dynamic general equilibrium model with a continuum of heterogeneous industries, each comprising a finite number of strategic price-setting firms. Firms in each industry collude on profit-maximizing markups, taking as given the behavior of all other industries. The strategic behavior of firms jointly determines the resource misallocation in each state, which in turn determines aggregate consumption in each state and the representative agent's marginal utility. Markups in any one particular industry can either be procyclical or countercyclical depending on the risk aversion of the representative agent and the correlation of sector-specific productivity with aggregate consumption. General equilibrium in the model is shown to exist under general conditions. Oligopolistic competition endogenously generates misallocation dynamics and may amplify aggregate technological shocks. This amplification channel is strongest when the dispersion of markups is countercyclical. Initial empirical tests support the importance of this novel channel for understanding aggregate fluctuations.

JEL classification: L16, E32, L13

Keywords: Oligopolistic Competition, Endogenous Aggregate Fluctuations, Markup Cycles, Allocative Efficiency, Industry Heterogeneity, Dynamic Games, DSGE Models.

*We thank Kyle Bagwell, Willie Fuchs, Nicolae Gârleanu, Zhiguo He, Anders Karlsson, Michael Katz, Debbie Lucas, Hanno Lustig, Miguel Palacios, Andrés Rodríguez-Clare, Julio Rotemberg, Stephanie Schmitt-Grohe, Alp Simsek, and Robert Staiger for very helpful comments and discussions. Research assistance by Michael Weber is gratefully acknowledged. We also appreciated the feedback from seminar participants at Aarhus University, Goethe University, MIT, University of California, Berkeley, University of Illinois, Urbana Champaign, University of Southern California, the Mitsui Finance Conference, and the 2012 Jackson Hole finance conference. We thank the Clausen Center for International Business and Policy at U.C. Berkeley Haas School of Business for financial support. All authors are at the Haas School of Business, UC Berkeley.

1 Introduction

How does industry-level market power influence aggregate fluctuations? Neo-Keynesian models in the spirit of Rotemberg and Woodford (1992) show how markups, if they are on average countercyclical, transmit aggregate demand shocks over time. In a static setting, more recent papers illustrate how the cross-sectional dispersion of markups across industries can generate a misallocation of resources and thereby welfare losses (see e.g., Bilbiie et al. (2008)).[1] An intriguing possibility is that if resource misallocation, in turn, affects industry markups, there could be a dynamic feedback between these two effects and aggregate fluctuations may be amplified or dampened by inter-industry differences in market power. To explore this possibility, we develop a general equilibrium model in which *oligopolistic* intra-industry competition generates markup dispersion across industries, misallocation, and time varying aggregate fluctuations in the economy.

Of course, for this dispersion channel to be relevant in practice, markups need to vary both over time and across industries. We provide first pass evidence that they do. We estimate a panel of price-cost margins (PCM) for 451 industries between 1959 and 2009 using the NBER manufacturing productivity database of Bartelsman and Gray.[2] Figure 1 plots the first and second moment of the cross-sectional PCM distribution from 1959 to 2009 corresponding to the relevant summary statistics in Neo-Keynesian models and the misallocation literature, respectively. Both first and second moments exhibit significant time series variation and, moreover, average market power and its dispersion have increased over time. This suggests that the importance of this channel may have increased. There is also considerable variation of markups across industries and, more interestingly, differences in how industry markups covary with the business cycle. Figure 2 plots the histogram of time-series correlation coefficients of industry markups with aggregate economic activity (GDP). Some industries exhibit strong countercyclical markup while others exhibit strong procyclical markups.[3]

Our paper is built on the seminal framework of Rotemberg and Woodford (1992). For our purposes, a limitation of their model – which assumes symmetric industries –

[1] Although this misallocation strand of literature is quickly expanding, the point that dispersion of markups may be more important than their actual levels from a welfare perspective dates back to Lerner (1934).

[2] While (average) price cost margins only correspond to precise markup estimates under special assumptions, e.g, if labor is the only factor input and production is constant returns to scale, they should be interpreted as a reasonable first pass proxy (see Nekarda and Ramey (2012) for more advanced methods). We discuss limitations of our approach and suggestions for future work in our empirical section.

[3] See also Bils et al. (2012), who provide evidence on variation of relative markups of durables and non-durables over the business cycle.

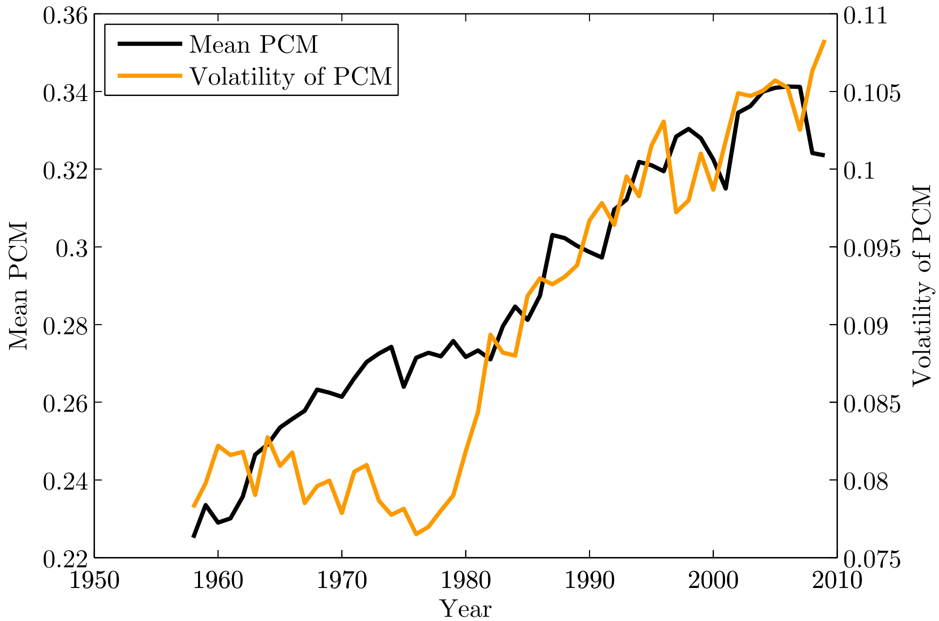

Figure 1. This graph plots the first two moments, i.e., mean and volatility, of the empirical price cost margin distribution from 1958 to 2009 based on 459 industries included in the NBER manufacturing productivity database of Bartelsmann and Gray. For the purpose of this graph, both moments are calculated assuming equal weights for each industry. See Section 6.2 for data description and precise definitions.

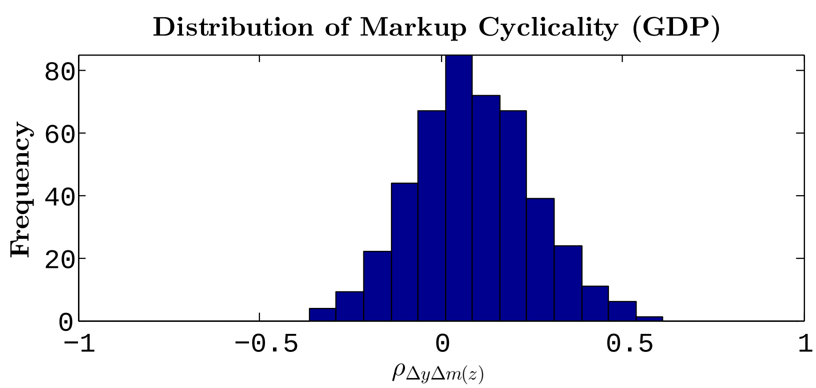

Figure 2. This graph plots a histogram of the distribution of markup cyclicality across industries. Specifically, the term $\rho_{\Delta y \Delta m(z)}$ refers to the time-series correlation coefficient of yearly log markup changes of a particular industry z with yearly log GDP changes. Since the average industry features $\rho_{\Delta y \Delta m(z)} > 0$, the evidence suggests slightly procyclical markups.

is that there is never any markup *dispersion* across industries. We therefore extend their framework to allow for cross sectional variation of industry concentration, and variations in productivity across industries and over time. This extension allows us to generate dynamic variation of markups levels and markup *dispersion*, microfounded by strategic behavior at the industry level. These dynamics translate into time-varying misallocations and represent an endogenous source of aggregate fluctuations which feed back into the firms' intra-industry optimization problems. A key effect is that, in general, misallocation amplifies technological shocks. The amplification channel is strongest if markup dispersion is countercyclical, i.e., if there is high dispersion in states of poor technological conditions (recessions).

To be more specific, we study a discrete time, infinite horizon general equilibrium economy with a continuum of industries (sectors), each of which is defined by a production technology. Within each industry, a finite number of identical strategic firms hire labor to produce a homogeneous good. The price of the good in each industry is determined by the outcome of an infinitely repeated pricing game. A representative agent consumes all goods, supplies all labor, and owns all the firms; thus all profits are valued by her preferences over consumption. We allow industries to differ cross-sectionally, both in their number of firms and their exposure to productivity shocks. These sources of heterogeneity, which are not present in Rotemberg and Woodford (1992), allow us to capture sector-specific strategic behavior, generate heterogenous markups, and analyze how idiosyncratic productivity shocks are transmitted to the aggregate economy. Firms in each industry maximize profits subject to intertemporal incentive compatibility constraints: In each period, each firm weighs the value of high short-term profits that can be obtained by aggressive pricing against the long-term profits that are obtained when all firms cooperate. In general equilibrium, the representative agent's consumption bundle depends on the output produced in each industry. This consumption affects the representative agent's valuation of each industry's profits and therefore feeds back into firms' ability to sustain collusion. Thus, while each industry takes the macro dynamics as given, industries jointly affect these macro dynamics. Our paper therefore provides a tight link between strategic industry behavior and aggregate outcomes.

Our theoretical contribution is three-fold. First, focusing on one industry, we characterize markups and derive conditions under which they are procyclical versus countercyclical. Countercyclical markups are often associated with oligopolistic competition, based on the risk-neutral setting of Rotemberg and Saloner (1986). In their framework, high product demand in good times increases firms' incentives to undercut competitors to reap immediate rewards; therefore equilibrium markups narrow in good times. Our paper shows that this intuition can be overturned if the representative agent's valuation of future profits are countercyclical. Indeed, in our framework these valuations can be endogenously countercyclical as determined by the preferences of a risk averse representative consumer. If valuations are sufficiently low in good times, then the present value of future cooperation is higher in booms, making procyclical markups possible. Within our model the intertemporal valuation effect can overturn the paradigm of countercyclical markups if the representative agent's intertemporal elasticity of substitution is low.[4] In general, the cyclicality of markups is therefore ambiguous.

While the cyclicality of the "average industry" is ambiguous, following the previous logic, we also show that one can decompose an industry's profit variations into a systematic (i.e., correlated with aggregates) and an idiosyncratic component and that the source of ambiguity lies in the systematic component. Markups are always countercyclical with respect to the idiosyncratic component i.e., controlling for the aggregate shock. This is natural, since there is no valuation or discount effect present for idiosyncratic shocks. A general implication of our analysis, consistent with the data, is that we should expect considerable variations in how industries markups vary with the business cycle.

Our second, and main, theoretical contribution is to analyze how the heterogeneous oligopolistic industry-level firm behavior may amplify technological shocks or even be the main source of aggregate volatility in the economy. These effects arise because of the feedback between industries in general equilibrium. Small changes in a few industries may become amplified if they affect other industries' ability to sustain collusive outcomes through the effects they have on the representative agents future valuation of consumption. In several examples we show that the amplification effects can indeed be drastic. We also highlight that shock amplification occurs whenever the endogenous cross-sectional dispersion of markups is higher in recessions than in good times, and that dampening of shocks is also theoretically possible in equilibrium, if markup dispersion is sufficiently procyclical.

Our third contribution is technical: We characterize the existence and qualitative behavior of equilibrium in our model. Given the complete generality of our set-up, allowing for full heterogeneity across industries and states, existence of equilibrium is by no means clear, a priori. Our main result in this part of the paper is Proposition 4, which shows the existence of equilibrium under minimal assumptions.

While the industry outcome given the behavior of all other industries is uniquely determined; multiple, qualitatively very different equilibria are consistent with industry-

[4]With constant relative risk aversion, this is isomorphic to requiring the coefficient of relative risk aversion to be high.

optimizing behavior. These arise because of the cross–sectional heterogeneity in our economy. If all industries were identical, equilibrium would be unique and efficient (see Proposition 5). The possibility of multiple equilibria in our model are reminiscent of the sunspot literature. However, multiple equilibria in our model are not due to self-fulfilling beliefs about demand; instead they arise as subgame perfect equilibrium outcomes from the strategic, rational, behavior of firms taking as given the (endogenous) valuation of future cash flows. This source of multiplicity driven by the representative agent's endogenous valuation is thus novel.

Although the main objective of our study is theoretical, we also provide a short discussion about the empirical implications of our model. We focus on the cyclicality of industry markups across the business cycle, and on aggregate shocks in the economy, and show some initial results for the latter. Especially, using the NBER manufacturing productivity database, we show that markup variations—defined in an appropriate way—is significantly (statistically and economically) positively related to aggregate shocks to consumption and GDP, with an R^2 of 17% and 11%, respectively. We believe that this discussion may serve as a basis for future empirical work.

1.1 Literature

We are certainly not the first researchers to address these issues and to explore micro foundations of macro shocks. Further, as our approach straddles multiple fields, it draws on various literatures including the industrial organization literature, the literature on misallocations and the literature on the propagation of macro shocks.

Our partial equilibrium results are most closely is related to the Industrial Organization literature on strategic competition over the business cycle (see, e.g., Chevalier and Scharfstein, 1995, Chevalier and Scharfstein, 1996, and Bagwell and Staiger, 1997). In particular, the partial equilibrium setup of Bagwell and Staiger (1997) highlights that procyclicality of markups may arise in a risk-neutral setting if expected future demand growth-rates are higher in boom times. In contrast to our paper, procyclicality is not driven via lower discount rates (as in our setup), but through even higher future (cash flow) growth rates in expansions. Further, dal Bo (2007) introduces i.i.d. interest rate fluctuations into the risk-neutral model of Rotemberg and Saloner (1986). Since the demand function is assumed to be constant and the interest rate process is exogenous, the paper cannot address pro- or countercyclicality of markups. Overall, our model takes this literature as a starting point, but extends the approach to allow for multiple industries and endogenous pricing of risk in an economy with risk averse agents.[5]

Since misallocations are the source of inefficiencies in our general equilibrium framework, our paper connects to a growing recent literature highlighting the welfare cost of variable markups. The formal expression for the welfare cost in our model is similar to the distortions that arise in sticky-price models in the spirit of Calvo (1983).[6] As Bilbiie et al. (2008) point out, the fundamental economics behind this can be traced back to early essays of Lerner (1934) and Samuelson (1949). Misallocations via variable markups have become particularly relevant for the literature on international trade since competition from abroad naturally affects industries in a heterogeneous way (see Epifani and Gancia (2011), Holmes et al. (2004), Edmond et al. (2012), and Dhingra and Morrow (2012)). From a modeling point, the literature on misallocations also highlights the special role of CES preferences under monopolistic condition in that market outcomes are efficient due to markups synchronization (see in particular Bilbiie et al. (2008) and Dhingra and Morrow (2012)). Instead, our paper shows that inefficiencies can arise even in settings with CES preferences (and inelastic labor supply) by allowing for oligopolistic competition with heterogeneous industries. This allows us to keep the tractability and standard aggregation results of CES, while being able to match relevant heterogeneity across industries.

Empirical studies suggest that losses from misallocation can be quantitatively large; at least in emerging market countries. Hsieh and Klenow (2009) estimate static losses ranging from $30\% - 50\%$ in China and $40\% - 60\%$ in India. In a dynamic setting, Peters (2012) considers the joint effect of misallocation, endogenous entry (see also Bilbiie et al. (2012)) and incentives to innovate (see also Kung and Schmid (2012)). Using a sample of manufacturing firms in Indonesia, he finds that a large proportion of the welfare gains from reducing barriers to entry results from the effect on the equilibrium growth rate rather than the reduction in (static) misallocation.

Since our paper combines real technology shocks with the just described endogenous misallocations, our paper also relates to an extensive literature on business cycles (e.g., Kydland and Prescott, 1982; Long and Plosser, 1983; Gabaix, 2011; Acemoglu et al., 2011). In contrast to the real business cycle literature, however, significant aggregate fluctuations may arise even when aggregate "technological" shocks are small. A recent

[5] The asset pricing implications of our framework are analyzed in a companion paper. Opp et al. (2012) show that industry characteristics (product demand, industry concentration and markups) should be informative about a firm's expected returns and volatility in the stock market.

[6] In contrast to sticky-price models, however, prices in our model are fully flexible and are determined *endogenously* as the outcome of a strategic game in each sector.

strand of literature has aimed at explaining how technological shocks at the individual firm or industry level do not diversify out, but may affect aggregate productivity. Gabaix (2011) notes that if the distribution of firm size is heavy-tailed, firm-specific shocks may indeed affect aggregate productivity. Acemoglu et al. (2011), suggest that inter-sectoral input-output linkages between industries may lead to "cascades effects" where a shock in one industry spreads through the economy and thereby becomes an aggregate shock. In our setup, such "cascade effects" may arise through the channel of the pricing kernel even if there is no direct input-output linkage between sectors. The mechanism in our model is also quite different, more along the lines suggested in Jovanovic (1987), who shows that idiosyncratic shocks may not cancel out in strategic games with a large number of players. We develop examples, in which aggregate productivity is close to constant across states, but because it varies at the sectoral level, the strategic behavior of firms leads to aggregate shocks in equilibrium. We believe that this provides an important mechanism for understanding the sources of aggregate fluctuations in the economy.

Our results highlight how strategic interaction between firms can generate endogenous fluctuations. These results are related to Gali (1994) and Schmitt-Grohe (1997) who, building on Woodford (1986) and Woodford (1991), study stationary sunspot equilibria in models with markups and investments. Both papers focus on the symmetric case with monopolistic competition, in which case the multiplicity of equilibria arises because of self-fulfilling expectations about future growth rates.[7] In contrast, our model features a unique equilibrium under symmetric behavior, i.e., homogeneous industries. Our key contribution is to allow for multiple, heterogeneous sectors in which multiple equilibria and welfare distortions arise from dispersion of markups in the cross-section.

The rest of the paper is organized as follows. In Section 2 we present the economic framework of the model. The equilibrium analysis of each industry and their joint effect on aggregate outcomes is presented in Section 3. Section 4 shows the existence of general equilibrium under general conditions, and Section 5 analyzes how oligopolistic competition can amplify, and even cause aggregate fluctuations. All proofs are delegated to the Appendix.

[7]In Jaimovich (2007), sunspot equilibria and countercyclical markups arise via entry and exit decisions (also see Jaimovich and Floetotto (2008)).

2 Model Framework

2.1 Physical Environment

Consider an infinite horizon, discrete time, discrete state economy in which time is indexed by $t \in \mathbb{Z}_+$ and the time t state of the world is denoted by $s_t \in \{1, 2, \ldots S\}$.[8] Each period there is a transition between states which is governed by a Markov process with time invariant transition probabilities:

$$\mathbb{P}(s_{t+1} = j | s_t = i) = \Phi_{i,j}. \tag{1}$$

Here, $\Phi_{i,j}$ refers to the element on the ith row and jth column of the matrix $\Phi \in \mathbb{R}_+^{S \times S}$. We assume that Φ is irreducible and aperiodic, so that the process has a unique long-term stationary distribution.

2.1.1 Production

There is a continuum of industries, indexed by $z \in [0, 1]$, each consists of $N(z) \geq 1$ identical strategic firms that produce and sell a unique non-storable consumption good. The nature of the strategic environment is discussed in Section 2.2. The production technology for each good z at time t is linear in labor with stochastic productivity $A(z, t) = A_{s_t}(z)(1 + g)^t$. Here, with some abuse of notation, $A_{s_t}(z)$ represents a state-dependent and sector-specific productivity component whereas $g \geq 0$ represents a common long-term productivity growth rate across all sectors. For ease of exposition, we set $g = 0$ in the main text and refer the reader to Appendix B, which shows the minor modifications necessary for the general case $g > 0$. Also, for tractability we assume that $A : S \times [0, 1] \to \mathbb{R}_{++}$ is a function that satisfies standard integrability conditions so that aggregation across industries is possible. Labor is supplied inelastically by a representative agent, who in each period allocates her one unit of human capital across industries, earning a competitive wage, $w(t)$, in return.[9]

[8] Here, $\mathbb{Z}_+ = \{0\} \cup \mathbb{N} = \{0, 1, \ldots\}$ is the set of non-negative integers. Also, we follow the standard convention that $\mathbb{R}_+$ is the set of nonnegative real numbers, whereas $\mathbb{R}_{++}$ is the set of strictly positive real numbers.

[9] We deliberately shut off the channel of endogenous labor supply to sharpen our findings of factor misallocation across heterogeneous sectors. Thus, our production factor in fixed supply could also be interpreted as "land" that has to be allocated to different sorts of crops (industries). We deliberately excluded physical capital accumulation from our model, to avoid the issue of disentangling effects of dynamic investment decisions from the effects of markups.

2.1.2 Preferences / Demand

The representative agent possesses iso-elastic preferences over aggregate consumption with risk aversion parameter γ, i.e., EIS is $\frac{1}{\gamma}$, and subjective discount factor δ, i.e.,

$$U = \mathbb{E}\left[\sum_{t=0}^{\infty} \delta^t \frac{C(t)^{1-\gamma}}{1-\gamma}\right], \tag{2}$$

where $C(t)$ represents the Dixit-Stiglitz *CES* consumption aggregator of goods (see Dixit and Stiglitz, 1977).[10] Thus,

$$C(t) = \left(\int_0^1 c(z,t)^{\frac{\theta-1}{\theta}} dz\right)^{\frac{\theta}{\theta-1}}. \tag{3}$$

The parameter $\theta > 1$ is the (constant) elasticity of substitution across goods. We note in passing that preferences with a more general state dependent utility specification are also covered by our specification.[11] The *CES* specification leads to standard *period-by-period* demand functions as a function of prices $p(z,t)$ and total income $y(t)$:[12]

$$c(z,t) = \frac{y(t)}{p(z,t)^{\theta} P(t)^{1-\theta}}, \tag{4}$$

where $P(t) \equiv \left(\int_0^1 p(z,t)^{1-\theta} dz\right)^{\frac{1}{1-\theta}}$ can be interpreted as the ideal price index. Total income, $y(t)$, is derived from wages, and distribution of firm profits, $\pi(z,t)$, across all sectors z:

$$y(t) = w(t) + \int_0^1 \pi(z,t) dz, \tag{5}$$

$$\pi(z,t) = \left[p(z,t) - \frac{w(t)}{A(z,t)}\right] c(z,t). \tag{6}$$

[10] See van Binsbergen (2007) or Ravn et al. (2006) for using *CES* preferences in a dynamic context.

[11] Consider the more general $\tilde{C}(t) = \left(\int_0^1 v_{s_t}(z) c(z,t)^{\frac{\theta-1}{\theta}} dz\right)^{\frac{\theta}{\theta-1}}$ as in Opp (2010). The state dependent "taste" function $v_s(z)$ can then easily be reduced to the case where $v_s(z) \equiv 1$, by transforming the productivity, $A_s(z) \mapsto v_s(z)^{(\theta-1)/\theta} A_s(z)$. Such a transformation can be interpreted as a numeraire change, where the amount of a unit of goods is redefined in each state. A state dependent taste function could, for example, represent an agent's higher utility of an umbrella in a rainy state than in a sunny state of the world.

[12] The demand functions $c(z,t)$ yield maximal $C(t)$ given an arbitrary price vector $p(z,t)$ and income $y(t)$. They are obtained via simple first-order conditions.

Going forward, we will normalize the nominal price index $P(t)$ to 1. (This is without loss of generality, since the wage rate $w(t)$ is a free variable.) Hence, income, wages, and profits are measured in units of aggregate consumption; in particular $y(t) = C(t)$.

2.2 Strategic Environment

Within each industry z, $N(z)$ identical firms play a dynamic Bertrand pricing game with perfect public information as in Rotemberg and Saloner (1986) taking as given the behavior of all other industries. The timing of the stage game in each period, t, is as follows. First, the state, s_t is revealed. Then all firms $i \in \{1, 2, \ldots N(z)\}$ in industry z simultaneously announce their gross markup, $M^{(i)}(z,t)$. For tractability, we express each firm's strategy in terms of gross markups instead of prices, satisfying $p^{(i)}(z,t) = M^{(i)}(z,t)\frac{w(t)}{A(z,t)}$. Consumers demand the product from the producer with the lowest markup. If all firms announce the same M, total demand in sector z is evenly shared between all $N(z)$ firms. The firms then go out and hire workers to meet demand.

Each industry z coordinates on the *symmetric, subgame* perfect equilibrium outcome that maximizes the present value of industry profits. Due to symmetry, the equilibrium gross markup function of each firm i satisfies: $M^{(i)}(z,t) = M(z,t)$, with the associated industry price

$$p(z,t) = M(z,t)\frac{w(t)}{A(z,t)}. \tag{7}$$

While the equilibrium outcome of this game is in general non-trivial (see Section 3.3), the two polar cases of a monopoly, i.e., $N(z) = 1$, and perfect competition provide useful bounds.

If the industry is served by a monopolist, he maximizes industry profits (equation 6) subject to consumer demand (equation 4) which leads to an optimal markup of:

$$M^m(z,t) = M^m = \frac{\theta}{\theta - 1}. \tag{8}$$

If, on the other hand, $N(z)$ is infinite, then we expect prices to be set competitively. In this case, the markup is 1. If the number of firms is finite but greater than one, we expect equilibrium markups to be somewhere in between the competitive and monopolistic prices, i.e., $M \in \left[1, \frac{\theta}{\theta-1}\right]$.

3 Partial Equilibrium Analysis

Before proceeding with our formal equilibrium analysis, it is convenient to transform our growing economy into a time-invariant economy in which outcomes only depend on time t through the state at time s_t. The resulting implications and other normalizations are presented in Section 3.1.

Our partial equilibrium analysis consists of two parts. First, for an arbitrary exogenous distribution of markups across industries, we characterize aggregate consumption, and show that it together with a measure of aggregate markups determines the efficiency losses in the economy (Section 3.2). Second, given the aggregate consumption and aggregate markup dynamics, we solve for the partial equilibrium outcome of one sector z in the economy, i.e., the optimal state-contingent markups (Section 3.3).

3.1 Preliminaries

We focus on equilibria which are time invariant in that equilibrium outcomes are the same at t_1 and t_2 if the states are the same, i.e., if $s_{t_1} = s_{t_2}$. Hence, we introduce the following notation for equilibrium markups (and similarly for other variables):

$$M(z,t) = M_{s_t}(z). \tag{9}$$

The focus on time invariant equilibria is natural in the stationary environment, since we prove that optimizing firm behavior in one particular industry is endogenously time invariant provided that all other industries exhibit time-invariant behavior. Moreover, it is ensured that (at least) one time-invariant equilibrium exists (see Proposition 4). We want to emphasize that this formulation does not impose any restriction on *off-equilibrium path* behavior.

For ease of exposition, we decompose productivity shocks $A_s(z)$ into the functions $\alpha_s(z)$ and $\bar{A}_s$ where $\alpha : S \times [0,1]$ and the vector $\bar{A} \in \mathbb{R}_+^S$. Specifically,

$$\alpha_s(z) \equiv \frac{A_s(z)^{\theta-1}}{\int_0^1 A_s(z)^{\theta-1} dz} = \left(\frac{A_s(z)}{\bar{A}_s}\right)^{\theta-1}, \qquad \text{where} \tag{10}$$

$$\bar{A}_s \equiv \left[\int_0^1 A_s(z)^{\theta-1} dz\right]^{\frac{1}{\theta-1}}. \tag{11}$$

Here, $\bar{A}$ represents the average productivity shock to the economy and $\alpha_s(z)$ captures the industry productivity shock relative to the economy. In other words, changes in $\alpha(z)$ across states are *idiosyncratic* shocks to individual industries, whereas changes in $\bar{A}$ are *systematic* shocks. We can also view $\alpha(z)$ as an S-vector, $\alpha(z) \in \mathbb{R}^S$.

As a result of the normalization, the average relative industry state is equal to one, i.e., $\int_0^1 \alpha_s(z)dz = 1$. Now instead of specifying A, we can equivalently specify the function of idiosyncratic shocks, α, and the vector of systematic shocks, $\bar{A} \in \mathbb{R}^S_{++}$. Given the previous argument, the exogenous variables in the economy can then be represented by the tuple $\mathcal{E} = (\alpha, \bar{A}, N, \Phi, \theta, \gamma, \delta)$.

3.2 Aggregate Consumption

Aggregate consumption is an important endogenous variable. As outlined above, we will first treat the outcome of the strategic game for each industry and each state as exogenously given, as summarized by the gross markup functions for each industry, $M_s(z)$. Together with the exogenous functions, $\alpha_s(z)$ and $\bar{A}_s$, the real outcome in the economy or the consumer's consumption bundle is completely determined, state-by-state. We will use aggregate consumption in two ways. First, as a measure of welfare and second as a determinant of the pricing kernel which governs the valuation of risky cash flows.

3.2.1 Misallocations and Aggregate Markups

This section illustrates the intuition of Lerner (1934) within the concrete setup of our model, i.e., the state-by-state misallocations caused by markup heterogeneity. For ease of exposition, we introduce two statistics of the cross-sectional markup distributions for the macro-economy in each state s:

$$\bar{M}_s = G_{1-\theta}\left(M_s\right), \tag{12}$$

$$\eta_s = \left(\frac{G_{-\theta}\left(M_s\right)}{G_{1-\theta}\left(M_s\right)}\right)^{\theta} \leq 1. \tag{13}$$

where $G_p\left(M_s\right) = \left(\int \alpha_s(z) M_s\left(z\right)^p dz\right)^{\frac{1}{p}}$ refers to the p-th order cross-sectional power mean of $M_s\left(z\right)$.[13] These statistics capture distinct elements of the cross-sectional markup

[13]Notice that by construction $\int_0^1 \alpha_s(z)dz = 1$, so we interpret α as a weighting measure where each industry obtains a weight according to its relative productivity.

distribution, and are jointly sufficient in describing the aggregate economy. The variable $\bar{M}_s$ captures the notion of aggregate market power, i.e., an appropriate average markup across industries. The variable η_s captures the (inverse of) heterogeneity of markups across industries. By Jensen's inequality, η_s is bounded above by one (obtained when all industries charge the same markup) and is decreasing in the heterogeneity of markups.[14] Thus, η_s can be interpreted as a measure of allocative production efficiency.

Lemma 1. *Given the functions M_s, α_s and $\bar{A}_s$, aggregate consumption, C_s, real income y_s, in state s are given by:*

$$C_s = y_s = \bar{A}_s \eta_s. \tag{14}$$

The fraction of real *income that is derived from labor income is given by:*

$$\omega_s = \frac{1}{\eta_s \bar{M}_s}. \tag{15}$$

The outcome in state s is Pareto efficient if $M_s(z) \equiv k_s$ for all z, so that $\eta_s = 1$.

From equation 14, aggregate consumption only depends on the exogenous aggregate shock $\bar{A}_s$ and allocative efficiency η_s implied by the markup distribution. As long as markups do not vary across industries in each state (i.e., $M_s(z) \equiv k_s$ for all z and s), the allocation of labor to industries is efficient so that aggregate consumption, i.e., potential output, is given by the aggregate shock $\bar{A}_s$. In all such economies, *relative* goods prices match the perfectly competitive and hence efficient outcome. Allocative efficient economies can only differ in terms of the decomposition of income, i.e., the fraction of income derived from labor ω_s and from firm profits, which are redistributed to the representative agent. An important benchmark case is the monopolistic economy, in which $k_s = \frac{\theta}{\theta-1}$ and $\omega = \frac{\theta-1}{\theta}$. The greater the cross-sectional dispersion of markups, the greater the misallocations, so that η_s falls.

3.2.2 Valuation

To value claims, we assume that a complete market of Arrow-Debreu securities is traded in zero net supply, in addition to the stocks of the firms. The unique one-period stochastic discount factor ("pricing kernel" or valuation operator) of the time-invariant economy, SDF, can be decomposed the subjective discount factor, δ, the (exogenous) productivity

[14] This follows from the fact that $G_p(\tilde{x}) > G_q(\tilde{x})$ for any non-degenerate random variable $\tilde{x}$ as long as $p > q$.

component and the (endogenous) misallocation factor.

$$SDF_{t+1} = \delta \left(\frac{C_{s_{t+1}}}{C_{s_t}} \right)^{-\gamma} = \delta \left(\frac{\bar{A}_{s_{t+1}}}{\bar{A}_{s_t}} \right)^{-\gamma} \left(\frac{\eta_{s_{t+1}}}{\eta_{s_t}} \right)^{-\gamma}. \tag{16}$$

Since time-t profits of a firm depend only on the state, s, the information about the firm's future profits can be summarized in an S-vector, π, where π_s is the profit in state s. As a result, the present value of expected future firm profits in each state s can be conveniently summarized in an S-vector:

$$V = \Theta \pi, \tag{17}$$

where the valuation operator Θ is defined as:

$$\Theta = \Lambda_m^{-1} (I - \delta \Phi)^{-1} \Lambda_m - I, \tag{18}$$

Here, Λ_m is a diagonal matrix, with the marginal utility in state s, $m_s = C_s^{-\gamma}$, as its sth diagonal element and I is the $S \times S$ identity matrix.[15] The valuation operator Θ has strictly positive elements. This simply represents the fact that higher profits in some state s strictly increases the present value of future profits, $V_{s'}$, in all states $s' = 1, \ldots, S$.[16]

3.3 Industry equilibrium

Understanding strategic price setting behavior in one industry z is the first step towards endogenizing the entire markups function M. We therefore characterize, as a function of industry and aggregate characteristics, when firms in a specific industry behave competitively, when monopolistic markups can be sustained, and when the outcome is neither of these extremes. Since each industry is small compared with the aggregate economy, firms in industry z take the dynamics of all other industries as exogenously given, i.e., they take M as exogenously given for all $z' \neq z$. In particular, the $S \times 2$ matrix consisting of the vectors C and $\bar{M}$ are jointly sufficient in describing the economic environment.

It is helpful to write real firm profits in sector z as a function of the choice variable

[15] The expression for Θ follows from solving the expression $V \equiv \delta \Lambda_m^{-1} \Phi \Lambda_m (\pi + V)$ for V.

[16] Recall that Φ is irreducible, so each state will be reached with positive probability, regardless of the initial state.

$M_s(z)$ and the exogenous variables $C, \bar{M}$ and $\alpha(z)$.

$$\pi_s(z) = C_s \alpha_s(z) \bar{M}_s^{\theta-1} \frac{M_s(z) - 1}{M_s(z)^\theta}.^{17} \tag{19}$$

While C_s and $\bar{M}_s$ are macro variables and hence affect all industries in a systematic fashion, the idiosyncratic productivity shock $\alpha_s(z)$ affects by definition only industry z. Note that industry z profits depend positively on the aggregate market power $\bar{M}_s$ since goods are substitutable (with $\theta > 1$).

Following Abreu (1988), we are interested in industry equilibria that generate the highest industry profits sustainable by credible threats. We restrict attention to *symmetric, pure strategy subgame perfect* equilibria of the infinitely repeated stage game described in Section 2.2. Firms condition their action at time t on the entire history of past actions of industry z and states up to time t. The relevant history of each industry z, h_t is defined as the entire sequence of markups, states, and aggregate variables:

$$h_t = \left\{ \left\{ M^{(i)}(z, \tau) \right\}_{i=1}^{N(z)}, s_\tau, \bar{M}_{s_\tau}, C_{s_\tau} \right\}_{\tau=0}^{t}, \tag{20}$$

with h_0 representing the empty history. Thus, a time-t, industry-z strategy for firm i is a mapping from $h_{t-1} \times S$ to a chosen markup, $M^i(z, \tau)$, $f_t^i : h_{t-1} \times S \to R_{++}$, (i.e., $f_t^i \in R_{++}^{h_{t-1} \times S}$). Here, the second parameter, $s \in S$, represents time t information about the state, which is available for the firm. A strategy for firm i is a sequence of time τ strategies, $\{f_\tau^i\}_{\tau=0}^\infty$.

The entire set of subgame perfect equilibria can be enforced with the threat of the worst possible subgame perfect equilibrium. In this case, the most severe punishment is given by the perfectly competitive outcome, i.e., zero profits forever after a deviation. Therefore, any subgame perfect equilibrium must satisfy the following incentive constraints for each state s,

$$\frac{\pi_s(z) + V_s(z)}{N(z)} \geq \pi_s(z). \tag{21}$$

That is, the share of discounted present value of profits under collusion, $\frac{\pi_s + V_s}{N}$, must be greater or equal to the best-possible one period deviation of capturing the entire industry demand π_s and zero profits thereafter. Industry profit maximization subject to this incentive constraint represents the only friction in our economy.[18]

[17] This equation follows from Lemma 1.

[18] We are implicitly assuming that firms can coordinate within an industry to achieve this best outcome

In the maximum profit equilibria, in each state, s, firms in an industry choose the vector of state contingent markups to maximize the value function, $V_s(z)$, given the value maximizing behavior in each of the other states of the world, $V_{-s}(z)$, and subject to incentive compatibility (equation 21),

$$V_s(z) = \arg\max_{M_s} : V_s(z) \,|V_{-s}(z), \tag{22}$$

for all s. Here, M_s maps to V_s via (19,17).

Within our model's setting, finding the solution to the optimization problem (22) is actually quite straightforward. First, we note that Equation 19 provides a bijection, $\pi_s \leftrightarrow M_s$, where $1 \leq M_s \leq \frac{\theta}{\theta-1}$, $0 \leq \pi_s \leq \pi_s^m \equiv \zeta C_s \alpha_s(z) \bar{M}_s^{\theta-1}$, and $\zeta \equiv \frac{(\theta-1)^{(\theta-1)}}{\theta^\theta}$ is a constant. Thus, the dynamic equilibrium can be viewed as a linear programming problem in which firms choose profits instead of prices, and replace M_s in (22) with π_s. The specific form of this corresponding linear programming problem makes it clear that the solution is the same for each state, and the optimization therefore collapses to a static, state independent, linear programming problem. Put differently, choosing an incentive compatible profit vector which maximizes firm value in state 1, i.e., $V_1(z) = \iota_1^T \Theta \pi(z)$ also maximizes firm value in all other states.[19]

Proposition 1. *Given aggregate consumption C and the average markup $\bar{M}$, the industry equilibrium outcome $\pi(z)$ (or equivalently $M(z)$) is uniquely determined by the solution to the following linear program.*

$$\pi(z) = \arg\max_{\hat{\pi}(z)} \ \iota_1^T \Theta \hat{\pi}(z), \quad s.t., \tag{23}$$

$$\hat{\pi}(z) \leq \pi^m(z), \tag{24}$$

$$0 \leq (\Theta - (N(z) - 1) I) \hat{\pi}(z), \tag{25}$$

Equilibrium profits in state s are either given by monopoly profits, $\pi_s^m(z)$, or the IC constraint in state s binds, i.e., $\pi_s(z) = \frac{\iota_s^T \Theta \pi(z)}{N(z)-1}$.

Going forward, it will be important to understand when the incentive constraint binds, and so markups deviate from the maximal. This is because, as we have observed, Pareto

with this equilibrium selection mechanism. This trivially rules out any outcomes where markups are higher than $\frac{\theta}{\theta-1}$, and outcomes where markups are lower than necessary. We do *not*, however, assume that firms can coordinate across industries, since in a large economy there are many industries and global coordination therefore typically is not possible.

[19] Here, ι_1 is the first column of the identity matrix I.

inefficiencies arise if markups differ across industries. To measure the "tightness" of the monopolistic incentive constraint, we introduce the "tightness" vector, $\Gamma(z)$, with element s denoting the s-state ratio of the present value of industry profits under monopoly markups to monopoly profits:

$$\Gamma_s(z) = \frac{\pi_s^m(z) + V_s^m(z)}{\pi_s^m(z)} = 1 + \frac{V_s^m(z)}{\pi_s^m(z)}. \tag{26}$$

If $\Gamma_{s_1} > \Gamma_{s_2}$ the incentive to deviate in state s_1 is smaller than in state s_2, i.e., the present value of collusion is high relative to current period profits.

Lemma 2. *The tightness vector satisfies:*

$$\Gamma(z) = (\Lambda_{\kappa(z)}^{-1}(I - \delta\Phi)^{-1}\Lambda_{\kappa(z)})\mathbf{1}, \tag{27}$$

where $\Lambda_{\kappa(z)} = diag(\kappa(z))$*, and the vector* $\kappa(z)$ *has elements:*

$$\kappa_s(z) = \pi_s^m(z)\, m_s = \zeta \bar{M}_s^{\theta-1} C_s^{1-\gamma} \alpha_s(z). \tag{28}$$

The variable κ_s captures an important determinant of the incentive to cheat in a certain state, Γ_s. It consists of the state component of the industry profit, $\pi_s^m(z)$, weighted by marginal utility in state s, $m_s = C_s^{-\gamma}$. We also define the minimum, $\underline{\kappa}(z) = \min_s \kappa_s(z)$. Substituting in the definition of $\pi_s^m(z)$ reveals that $\kappa_s(z)$ is increasing in the idiosyncratic industry productivity $\alpha_s(z)$, whereas the net dependence on aggregate consumption C_s depends on the EIS, i.e., $\frac{1}{\gamma}$. If $\gamma > 1$ ($\gamma < 1$), the variable $\kappa_s(z)$ is decreasing (increasing) in aggregate consumption. This is because aggregate consumption is not only a driver of profits (as in the risk-neutral case), but also influences the marginal utility (discount rate) due to risk aversion. If the discount rate channel dominates, procyclical markups may occur as the example following Proposition 2 reveals.

Using the definition of the tightness vector, we are now able to derive closed-form expressions for the threshold number of firms that leads to perfect competition and the monopoly outcome, respectively. Intuitively, for a small number of firms $N(z) \leq N^m(z)$, the monopoly outcome is sustainable in all states, while too many firms in one industry, $N(z) > N^c$, generates the competitive outcome in all states. In between, markups may vary across states. This intuition is formalized in the following proposition.

Proposition 2. *Given aggregate consumption* C *and the average markup* $\bar{M}$*, equilibrium* *profits in state s, $\pi_s(z)$, satisfy:*

$$\begin{aligned}
&\pi_s(z) = \pi_s^m(z) && \textit{for} \quad N(z) \leq N^m(z), \\
&\pi_s(z) \in (\underline{\kappa}(z)C_s^\gamma, \pi_s^m(z)] && \textit{for} \quad N(z) \in (N^m(z), N^c), \\
&\pi_s(z) = \underline{\kappa}(z)C_s^\gamma && \textit{for} \quad N(z) = N^c, \\
&\pi_s(z) = 0 && \textit{for} \quad N(z) > N^c.
\end{aligned}$$

where the respective threshold values satisfy $N^m(z) \stackrel{\text{def}}{=} \min_s(\Gamma_s(z))$ and $N^c \stackrel{\text{def}}{=} \frac{1}{1-\delta}$.

The different regions are best shown in a stylized example which highlights the intuition for our results. Assume that aggregate consumption satisfies $C = (1, 2, 4)^T$, that aggregate markups are competitive in all states, $\bar{M} = (1, 1, 1)^T$, and that $\alpha(z) = \left(\frac{1}{2}, \frac{3}{4}, 1\right)^T$. For simplicity, assume that the economy is i.i.d. with all states being equally likely. Finally, assume preference parameters of $\delta = 8/9$, $\gamma = 2$, and $\theta = 2$. Is is easy to show that the tightness vector in this example satisfies:

$$\Gamma(z) = (7, 9, 13)^T. \tag{29}$$

Thus, monopoly markups are sustainable for $N(z) \leq N_m(z) = 7$.[20] Given δ, the number of firms necessary to induce the competitive outcome is $N_c = 9$. Figure 3 plots state-contingent profits in the left panel and the corresponding state-contingent markups as a function of the number of firms, confirming the four cases in Proposition 2.

It is useful to explain the intuition for why the example exhibits procyclical markups, i.e., $M_1(z) \leq M_2(z) \leq M_3(z)$. In general, state-contingent markups arise because the incentive to cheat is higher in some states of the world than others. This incentive depends on the comparison between current-period profits and the present value of future profits (see equation 21). Since current period profits are higher in good states of the world, the incentive to cheat is higher in better states unless this effect is overwhelmed by the present value of future profits, V_s. Due to our i.i.d. specification in this example, expected future profits are equal across states. Nonetheless, the *present value* of future (monopoly) profits is higher in good states of the world, i.e., $V_1^m(z) < V_2^m(z) < V_3^m(z)$. This valuation effect is driven by countercyclical discount rates that arise from a high marginal utility of consumption in worse states of the world. Since the $EIS < 1$, this valuation effect is sufficiently strong, so that the incentive to reap the short-term profits

[20] Recall that aggregate markups are competitive even if a zero measure of industries are non-competitive. Thus, there is no inconsistency in having one non-competitive industry in an economy that in aggregate is competitive.

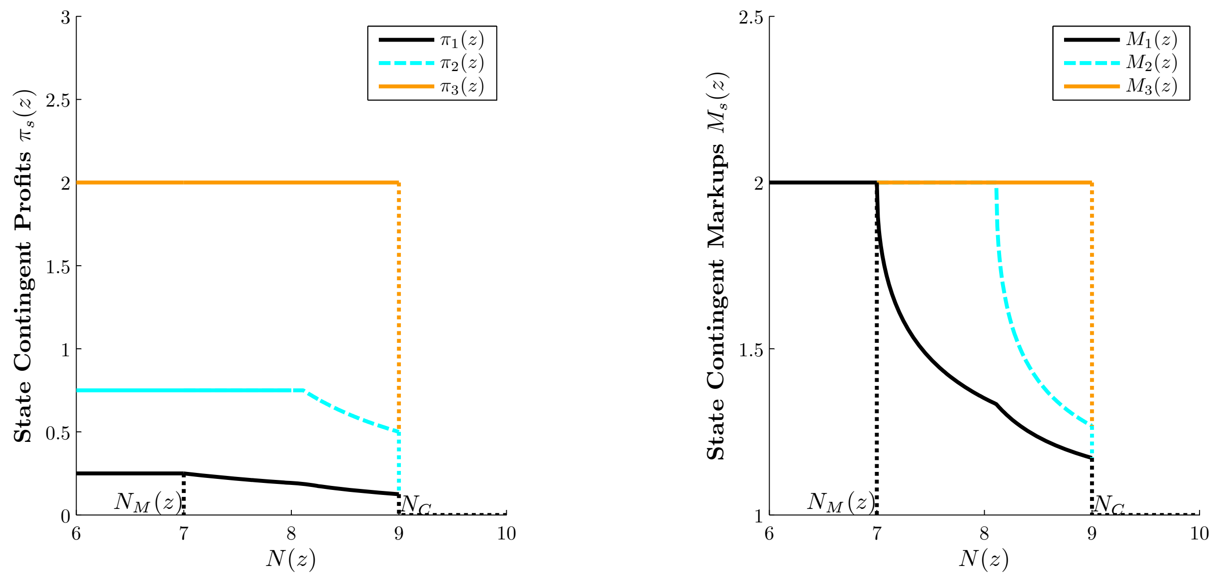

Figure 3. This graph plots the state contingent profits and markups of one particular industry given aggregate consumption of $C = (1, 2, 4)^T$, aggregate markups of $\bar{M} = (1, 1, 1)^T$, and the relative industry state of $\alpha(z) = (2, 3, 4)^T$. If there are fewer than 7 firms in the industry, monopoly markups are sustainable in all states. Increasing the number of firms further causes the incentive constraint in state 1 to bind first, then in state 2 and finally, at $N_C = 9$, all markups collapse discontinuously to the competitive outcome, i.e., 1.

is highest in bad states of the world, leading to procyclical markups.[21]

Also consider a slight perturbation of the example by considering the knife-edge of log utility, i.e., $\gamma = EIS = 1$. In this case, $\kappa_s(z)$ is only a function of the idiosyncratic productivity shock $\alpha_s(z)$. With respect to the idiosyncratic component, the conventional result of Rotemberg and Saloner (1986) applies: markups are countercyclical. For an industry with shocks $\alpha(z) = \left(\frac{1}{2}, \frac{3}{4}, 1\right)^T$ countercyclicality with respect to the idiosyncratic component also translates into "overall countercyclical" markups. In contrast, an industry with shocks $\alpha(z) = \left(1, \frac{3}{4}, \frac{1}{2}\right)^T$ will exhibit "overall procyclical markups". Thus, our setup allows for heterogeneous cyclicality of industries consistent with the stylized facts presented in Figure 2.

To summarize, while the exact conditions for pro-/ countercyclicality are certainly special to our setup, the fundamental asset pricing implications for industrial organization hold more generally. If discount rates for risky assets are countercyclical, then the conventional wisdom of overall countercyclical markups following Rotemberg and Saloner

[21] This valuation effect in this example is so strong that it also outweighs the procyclical idiosyncratic component of profits $\alpha(z)$.

(1986) may be overturned. Moreover, we want to emphasize that this set of results is not driven by industry heterogeneity and should even hold in a setup with homogeneous firms (such as in Rotemberg and Woodford, 1992).

Even with the generality of our setup, we are able to put a lot of structure on the industry equilibrium outcome. Intuitively, the threshold number of firms that allows the monopoly outcome is directly linked to $\Gamma(z)$. It is determined by the state in which the incentive to deviate is the highest, i.e., the state in which $\Gamma(z)$ attains its minimum. The maximum number of firms beyond which collusion completely breaks down is simply given by $N^c \overset{\text{def}}{=} \frac{1}{1-\delta}$, i.e., it only depends on the growth adjusted discount rate. Quite surprisingly, the threshold value is *independent* of industry characteristics as captured by $\alpha(z)$ and aggregate properties such as aggregate consumption C or the average markup $\bar{M}$.[22] Moreover, we are able to derive an analytical formula for the profits of any industry z with N^c firms.

What remains is to characterize the solution for industries with $N^m(z) < N(z) < N^c$ firms. For a special case, this region is empty, i.e., $N^m(z) = N^c$.

Lemma 3. *If $\kappa_s(z) = k$ for all s and some arbitrary constant k, the threshold value for the monopoly outcome is given by N^c, i.e., $N^m(z) = N^c$.*

In such industries markups are never state dependent (regardless of the number of firms in the industry), since they are neither state dependent in the monopolistic case, nor in the competitive case. One benchmark specification delivers this scenario: If the representative agent has log utility, i.e., $\gamma = 1$, and all industries are homogeneous, i.e., if $A_s(z) \equiv \bar{A}_s$, for all z and s, and $N(z) \equiv N$, then $\kappa_s(z) = k$ for all s and z.[23] As a result, all industries in the economy either behave like a monopolist or are perfectly competitive. Except for this knife-edge case, the region between $N^m(z)$ and N^c is non-empty, and represents the economically most interesting region, since it gives rise to state-contingent markups.

Proposition 3. *For an industry in which $N^m(z) < N(z) < N^c$,*

[22] At N^c, the incentive constraint is characterized by the indifference condition of a risk-neutral firm that compares the shared perpetuity value under collusion, $\frac{\pi^*(z)}{1-\delta}\frac{1}{N^c}$, and the best possible one-period deviation, $\pi^*(z)$.

[23] By homogeneity, we obtain $M_s(z) = \bar{M}_s$ and $\alpha_s(z) = 1$. Since $\gamma = 1$, κ_s does not depend on C. In turn, this will also imply that $\bar{M}_s$ does not depend on C. Thus, $\bar{M}_s = M^m$ (for $N \leq N^c$), so that $\kappa_s = \frac{1}{\theta}$, or, $\bar{M}_s = 1$ (for $N > N^c$) so that $\kappa_s = \zeta$. Proposition 5 implies that this is the unique equilibrium outcome.

1. *There will be at least one state in which monopolistic profits are obtained,* $\pi_s(z) = \pi_s^m(z)$ *for some* s.

2. *Equilibrium profits,* $\pi_s(z)$, *are decreasing in* $N(z)$ *for each* s, *as are markups.*

3. *Equilibrium profits,* $\pi_s(z)$, *are increasing in* $\alpha_{s'}(z)$, *for each* s, s', *as are markups.*

4. *Equilibrium profits and markups depend continuously on all parameters (*N, C, $\bar{M}$, Φ, α, *and* $\bar{A}$*).*

It is straightforward to verify properties 1, 2, and 4 in Figure 3. Thus, given that the aggregate variables of the economy C and $\bar{M}$ are known, the choice of state-contingent markups in a specific industry z is exactly characterized.

4 General Equilibrium

We show the existence of general equilibrium in which firms in each industry choose optimal markups given the (optimal) markups chosen by firms in all other industries. Recall that the economy's environment is characterized by the tuple $\mathcal{E}$, i.e., by the real variables $\alpha : S \times [0,1] \to \mathbb{R}_+$, $N : [0,1] \to \mathbb{N}$, $g \geq 0$, $\bar{A} \in \mathbb{R}^S_{++}$, the irreducible aperiodic stochastic matrix, $\Phi \in \mathbb{R}^{S\times S}_{++}$, and the preference parameters, γ, θ, and $\hat{\delta}$. We note that a given equilibrium is completely characterized by the markup function, $M : S \times [0,1] \to \left[1, \frac{\theta}{\theta-1}\right]$, together with $\mathcal{E}$, since all other real and financial variables can be calculated from M using (7) and (12-19). This motivates the following

Definition 1. *General Equilibrium in economy* $\mathcal{E}$ *is given by a markup function* $M : S \times [0,1] \to \left[1, \frac{\theta}{\theta-1}\right]$ *for which,*

1. $\bar{M}$ *and* C *are defined by Equations 12 and 14,*

2. *For all* z, $M(z)$ *is the solution to the maximization problem given by Equations 23-25, where* $\pi^m(z)$ *in the optimization problem is given by Equation* **??**.

We note that the existence and uniqueness of the second part of the definition is guaranteed by Proposition 1, industry by industry, i.e., given $\bar{M}$ and C there is a unique optimal markup function. It is a priori unclear, however, whether there exists a general equilibrium, i.e., whether both parts can be solved simultaneously. In other words, both the mappings, $M \mapsto (\bar{M}, C)$ (part 1) and $(\bar{M}, C) \mapsto M'$ (part 2) are well defined, but it is unclear whether M can be chosen such that the second step maps to the same markup function that was used in the first step, i.e., such that $M' = M$.

It turns out that we are able to prove the existence of equilibrium under very general conditions. Specifically, we assume that the functions N and α are Lebesgue measurable functions, and impose the following technical condition:

Condition 1. *For all s, for almost all z, $c_0 \leq \alpha_s(z) \leq c_1$ for constants, $0 < c_0 \leq c_1 < \infty$.*

We now have the following general result:

Proposition 4. *General equilibrium exists in any economy that satisfies Condition 1.*

Thus, only the technical conditions of integrability and boundedness of productivity functions across industries is needed to ensure the existence of equilibrium. The generality of this existence result is a priori quite surprising. In static general equilibrium models with imperfect competition, additional conditions in the form of quasi-concavity of firms' profit functions, and uniqueness of market clearing price functions given a productive allocation, are typically needed to show the existence of general equilibrium (see Gabszewicz and Vial, 1972; Marschak and Selten, 1974; and Benassy, 1978). These conditions are indeed satisfied in our model, as seen in Section 2.1. Instead, the major challenge is the dynamic setting, where the move from a static to a dynamic Bertrand game between firms drastically enlarges the strategy space. Since all firms are intertwined through the effects their actions have on the pricing kernel, showing the existence under general conditions seems out of reach. Previous literature (e.g., Rotemberg and Woodford, 1992; Gali, 1994; and Schmitt-Grohe, 1997) has avoided the issue by assuming complete symmetry, in which case the state space collapses. Of course, the focus on symmetric economies also restricts the type of effects that may arise, e.g., in terms of efficiency losses.

The reason why existence is still provable in our setting is the special structure of the model. The key property is that the game played between firms is simple enough so that we can completely characterize their behavior under general parameter values and show that this behavior has some needed properties. Specifically, the structure of firms' constrained optimization problems in equations 23 - 25 allows us to show uniqueness and uniform continuity of industry outcomes with respect to all parameters. This follows from two properties of the optimization problem. First, the objective function is linear.

Second, the IC constraints have a specific form such that (i) for any number of firms less than the competitive threshold, $N < N^c$, the domain of optimization is uniformly bounded, closed, convex with nonempty interior, (ii) for industries with $N = N^c$ the domain is a closed bounded line, and (iii) for industries with $N > N^c$ the domain contains a single point, the origin. These properties imply well behaved (unique and uniformly continuous) outcomes industry-by-industry, which in turn implies that the mapping $M \mapsto (\bar{M}, C) \mapsto M'$ is continuous (in the function space L^1).

Technically, the proof of Proposition 4 depends Schauder's fixed point theorem.[24] Specifically, it is shown in the proof of Proposition 4 that the space of markup functions is compact and convex, which via Schauder's theorem then guarantees the existence of a fixed point, i.e., an equilibrium. Details are given in the proof.

We note that Proposition 4 makes no claim about equilibrium uniqueness — a subject that will be explored further in the next section.

5 Endogenous Misallocation Dynamics

In this section we analyze the properties of general equilibrium with a sequence of examples. These qualitative examples are meant to deliver the main economic intuition for our results, without any attempt towards a real world calibration.[25] In particular, Section 5.1 shows how strategic competition can endogenously amplify technological shocks. In Section 5.2 we discuss conditions for equilibrium uniqueness and reveal how industry heterogeneity might produce multiple equilibria. Finally, we present various comparative statics in Section 5.3.

5.1 Shock Amplification

In general equilibrium, the decisions of firms in one part of the economy affect aggregate consumption and hence the pricing kernel and thereby the decisions of all other firms in the economy. Thus, in equilibrium, technological shocks are transmitted through the oligopolistic interaction between firms. To illustrate the mechanism of our model, consider the simple economy described in Table 1, with three distinct types of industries, I_1, I_2 and I_3, and $S = 2$ states. Here, all industries with $z \in I_j$ belongs to industry type

[24] We use this because we have a continuum of industries.

[25] We discuss empirical implications in Section 6.

Type, j	I_j	N	A_1	A_2	α_1	α_2
1	$z \in [0, 0.02)$	19	0.25	1	0.8728	1
2	$z \in [0.02, 0.81)$	19	1	1	1.0026	1
3	$z \in [0.81, 1]$	1	1	1	1.0026	1
$\bar{A}$					$\bar{A}_1 = 0.974$	$\bar{A}_2 = 1$

$$\Phi = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$

$$\gamma = 6, \qquad \theta = 1.1, \qquad \delta = 0.95.$$

Table 1. Economy with three industries and two states.

j. With a slight abuse of terminology, we will call the I_j sets "industries," although each set represents many identical industries. Thus, there is one very small industry (I_1), one large industry (I_2) and one medium-sized industry (I_3). The first two industries have many firms, $N = 19$, but they will still not be perfectly competitive, since $N^c = \frac{1}{1-\delta} = 20$. The third industry is monopolistic, so that it will be charge the markup $\frac{\theta}{\theta-1}$ regardless of the behavior in the first two industries.

Columns 4 and 5 in Table 1 describe the absolute productivity shocks, A, in the two states. We see that only the very small first industry experiences any variation in productivity across the two states. The aggregate variation in productivity will therefore be small. In columns 6 and 7, we show the decomposition of the absolute productivity shocks into relative and aggregate components, α and $\bar{A}$ (see equations 10 and 11). The effect on aggregate productivity of the first industry's shock is about 2.5%, since aggregate productivity is 0.974 in the low-productivity state and 1 in the high-productivity state. This would also be the aggregate consumption in the two states in an efficient outcome. Note that the shock to industry 1 also affects the relative productivity in industries 2 and 3, since α is normalized to sum to one across industries, state by state.

Before analyzing the equilibrium in this economy, it is instructive as a reference case to study the economy which is identical to that in Table 1, except for that $A_1 = 1$ in industry 1. This is thus an economy with no productivity shocks, neither idiosyncratic nor aggregate, and it follows that $\bar{A}_1 = \bar{A}_2 = 1$ and $\alpha_s(z) \equiv 1$ in this reference economy. One easily verifies that the monopolistic outcome, in which markups $M \equiv \frac{\theta}{\theta-1} = 11$ are chosen by all firms in all states, is feasible in this case (this also follows as a consequence from Lemma 3, since $N \leq N^c$ in all industries), leading to the efficient outcome where $C_1 = \bar{A}_1 = 1$, $C_2 = \bar{A}_2 = 1$.

The situation is different for the economy given in Table 1. The fully monopolistic outcome is no longer feasible, because it does not satisfy the IC constraints for firms in industry 2. Instead, an equilibrium is given by the following markups:

Markups	$s = 1$	$s = 2$
$M(I_1)$	1.580	11
$M(I_2)$	1.465	11
$M(I_3)$	11	11

(30)

leading to aggregate consumption

$$C_1 = 0.795, \qquad C_2 = 1.$$

Thus, the small productivity shock ($\approx 2.5\%$) leads to a significant decrease in equilibrium output ($\approx 20\%$) in state 1.

The intuition for why amplification occurs in this example is exactly in line with our main theme in this paper, that technological shocks that are small in aggregate — in that they only affect a few industries — change the strategic behavior of firms in other industries through the effect they have on the pricing kernel.

This mechanism is explained in Figure 4, focusing on the behaviors of industries 1 and 2.[26] In the upper part of the figure, the reference economy with identical industries is shown, in which case monopolistic profits are feasible for both industries. In the lower part of the figure, the economy in Table 1 is shown. Line A shows the relevant IC constraint in state 1, given the pricing kernel in the monopolistic outcome. Monopolistic profits are indeed feasible in industry 1 (lower left figure), but infeasible in industry 2 (lower right figure). Thus, the lower productivity in industry 1, through its effect on the pricing kernel, affects the outcome in sector 2, which moves the IC constraint in state 1 to line B. This in turn changes the pricing kernel even further, making monopolistic profits in industry 1 infeasible and further changing the outcome in industry 2, moving to lines C in the two industries, and generating further feedback effects. The ultimate effect of this mechanism is that the equilibrium moves to line D in the two figures, substantially

[26]Industry 3 is always monopolistic. The reason that it is still important for the example is that substantial efficiency losses only occur when there is high variability in markups across sectors. If industry 3 was not present then the economy would always be close to efficient, since markups would be the same for the vast majority of industries in each state — almost identical to the markups charged in industry 2. In contrast, when industry 3 is present and industry 2 charges low markups, efficiency will be low.

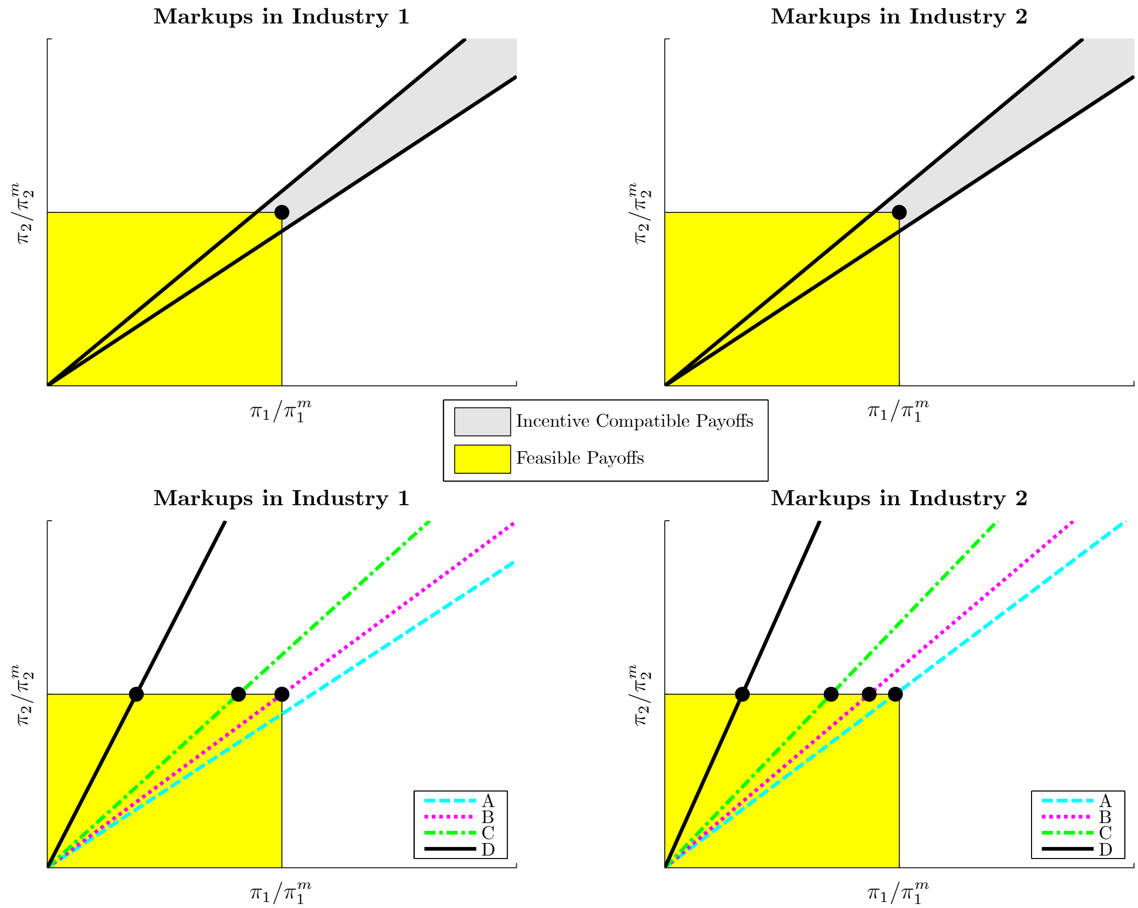

Figure 4. In each of the 4 panels, we plot incentive compatible and feasible profits (scaled by monopoly profits) in both states of the world. Feasibility refers to the upper bound imposed by monopoly profits in each state, i.e., $\pi_s/\pi_s^m = 1$. Incentive compatibility in both states is governed by two lines. The upper line refers to the IC constraint in state 2. The lower one refers to the IC constraint in state 1. The upper 2 panels refer to the benchmark economy with identical industries. The outcome in industry 1 (2) is plotted on the left (right). In both industries and states monopolistic profits are sustainable. Below: We only plot the relevant IC constraint in state 1. Monopolistic profits violate IC constraint in state 1 for industry 2 (line A), in turn changing the IC constraints in state 1 for industry 1 (line B). The resulting equilibrium (line D) is substantially different.

different from monopolistic equilibrium in the reference economy.

5.2 Uniqueness

Our general existence result (see Proposition 4) makes no claims with regards to uniqueness. We will show in this section that there *may* be multiple equilibria whenever a nonzero measure of firms fails the condition of perfect competition. Uniqueness of equilibria can, however, be proved for the important benchmark case of homogeneous industries.

Proposition 5. *If industries in the economy $\mathcal{E}$ are homogeneous, i.e., if $A_s(z) \equiv \bar{A}_s$, for all z and s, and $N(z) \equiv N$ for all z, then the equilibrium is unique.*

Of course, it is also immediate the outcome with homogeneous industries is Pareto efficient, i.e., $C_s = \bar{A}_s$. While uniqueness of aggregate consumption follows directly from Proposition 1 as $M_s(z) = \bar{M}_s$, Proposition 5 also implies uniqueness of state-contingent markups $\bar{M}_s$.

As a result of this proposition, multiplicity of equilibria must be driven by industry heterogeneity and the implied feedback from aggregate consumption to the pricing kernel. Indeed, it can be verified that the heterogeneous economy $\mathcal{E}$ parameterized in Table 1 exhibits (exactly) one more equilibrium supported by the equilibrium markups:

Markups	$s=1$	$s=2$
$M(I_1)$	11	2.104
$M(I_2)$	11	2.605
$M(I_3)$	11	11

(31)

and leading to aggregate consumption

$$C_1 = 0.974, \qquad C_2 = 0.884.$$

Again, aggregate fluctuations are endogenously determined. However, despite the same technology specification, the second equilibrium is very different from the first one. First, although the state with low productivity is the first, aggregate output is the lowest in the second state in this second equilibrium. There is thus a second way to ensure that firms do not deviate from equilibrium strategies, namely to decrease the attractiveness of state 2. We note that the first equilibrium leads to higher output than the second equilibrium in state 2 (1 versus 0.884), whereas the second equilibrium dominates in state 1 (0.974 versus 0.795). It is indeed easily verified that the second equilibrium Pareto dominates the first in expected utility terms, regardless of the current state.

It turns out that there are multiple equilibria even in the reference economy with no productivity shocks, i.e., in which $A_1 = 1$ also for the first industry. In such an economy, sector heterogeneity is purely driven by the differing number of firms in industries 1 and

2 compared to industry 3.

One can verify that

Markups	$s=1$	$s=2$
$M(I_1)$	11	1.634
$M(I_2)$	11	1.634
$M(I_3)$	11	11

(32)

with aggregate consumption.

$$C_1 = 1, \qquad C_2 = 0.830,$$

is also an equilibrium. Moreover, a third equilibrium (symmetrically) exists in which markups and consumption are low in state 1. Since productivity (idiosyncratic and aggregate) is constant across states in this case, aggregate fluctuations in this equilibrium are completely endogenous, and the COST is infinite. Thus, truly endogenous business cycles can arise as a result of strategic competition in our model if sectors differ purely in the number of firms. If all sectors were completely homogeneous, the equilibrium markup in each state of the world would be the same and the equilibrium would be unique (see Proposition 5).

We note that equilibrium multiplicity in models with markups, in the form of stationary sunspot equilibria, have also been generated in Gali (1994) and Schmitt-Grohe (1997). The analysis in Gali (1994), especially, has similarities to ours in that he assumes linear production technologies and also covers the case with inelastic labor supply. However, his mechanism is different from ours. Since he focuses on the symmetric case with monopolistic competition, there is no role for heterogeneity in markups across firms, and the corresponding inefficiencies such heterogeneity creates. Instead the multiplicity of equilibria arises because of self-fulfilling expectations about future growth rates. In our setup, industry homogeneity implies uniqueness as it prohibits efficiency losses due to cross-sectional variation of markups and hence shuts down the feedback channel through the pricing kernel.

5.3 Comparative Statics

The equilibrium outcome may be very sensitive to small changes in some parameter values, whereas it is remarkably stable in other aspects. The results together suggest that cross economy (e.g., cross country) comparisons need to be carefully designed to capture meaningful relationships when studying the determinants of an economy's dynamics.

We show that the equilibrium outcome may be extremely sensitive to small differences in long-term growth rates, g, and specifically that small differences can have large welfare effects by taking the economy from a Pareto efficient, perfectly competitive, outcome to one in which some industries are competitive and others are not. We study a modified version of our workhorse example given in Table 2. There are now 20 firms in each

Type, j	I_j	N	A_1	A_2	α_1	α_2
1	$[0, 0.2)$	20	1	1	0.972	0.972
2	$[0.2, 0.6)$	20	1	2	0.972	1.041
3	$[0.6, 1]$	20	2	1	1.041	0.972
$\bar{A}$					$\bar{A}_1 = 1.33$	$\bar{A}_2 = 1.33$

$$\Phi = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$

$$\gamma = 6, \qquad \theta = 1.1, \qquad \delta = 0.95,$$

Table 2. Modified economy with three industries and two states.

industry, that the asymmetry in industry sizes is not as large as in the previous example, and that there are productivity variations across states also in the large industries. It is straightforward to verify that there is an equilibrium with aggregate consumption

$$C_1 = C_2 = 1.19,$$

and markups

Markups	$s = 1$	$s = 2$
$M(I_1)$	11	11
$M(I_2)$	11	4.44
$M(I_3)$	4.44	11

(33)

and that the efficiency therefore is $\eta_s = \frac{C_s}{\bar{A}_s} \equiv \frac{1.19}{1.33} = 0.89$, about 11% below the Pareto efficient outcome in both states.

Now, in an identical economy as the one in Table 2, except for that $\delta = 0.949$ instead of 0.95, only the competitive outcome is an equilibrium, leading to $M \equiv 1$ and efficient consumption

$$C_1 = C_2 = 1.33.$$

This follows immediately since $N > N^c = \frac{1}{\delta}$ in all industries. Thus, the discontinuity of markups close to N^c, analyzed in Section 3.3, leads to extreme sensitivity of the equilibrium outcome when there is a substantial number of industries in which the number of firms close to N^c. Surprisingly, the mass of firms that are perfectly competitive in our economy, $\lambda(\{z : M_s(z) = 1, \forall s\})$, only depends on the exogenous subjective discount factor δ and the exogenous distribution of firms.

Suppose now that we relax the exogeneity assumption with regards to the number of firms and allowed for free entry with zero entry costs. Then, the discontinuity of firms' value function in the number of other competing firms also implies that free entry into each industry would not necessarily drive the economy to the efficient or competitive outcome. For a potential entrant, knowing that, on entering, industry profits would drop to zero means that he does not have a strict incentive to enter. For example, the outcome with $N = 20$ firms in each industry and aggregate consumption in Table 2 is an equilibrium in the economy with zero costs of entry. By this logic, even though the number of firms in each industry is exogenously given in our economy, the feedback effects between market power and industry equilibrium may be robust to alternate specifications.

6 Empirics

6.1 Testable Predictions

Our theory has testable predictions for both the individual industry behavior as well as their joint effect on aggregate economic activity in general equilibrium. The partial equilibrium results suggest that the analysis of markup cyclicality should disentangle the systematic component of industry demand/ profits and the idiosyncratic component. While the systematic "average" industry might exhibit procyclical (via the channel of discount rates) or countercyclical markups, the predictions with regards to idiosyncratic shocks, $\alpha(z)$ are unambiguous: If an industry is relatively procyclical, i.e., $\alpha_s(z)$ increasing with $C_s(z)$, then this industry will exhibit *relatively* countercyclical markups *compared* to the average industry. One can test these predictions by relating the cross-sectional distribution of markup cyclicality (see Figure 2) to idiosyncratic demand / productivity proxies and the number of firms in a sector.

Our main general equilibrium result relates variation in aggregate economic activity to variations in allocative efficiency and technological shocks, i.e., $C_t = A_t\eta_t$. Since empirical studies are mostly concerned about growth, it is useful to express this identity as:

$$\Delta c = \Delta a + \Delta e \tag{34}$$

where $c_t = \log(C_t)$, $a = \log(\bar{A})$ and $e = \log(\eta)$ and Δ refers to first differences. From this expression, it is immediate that amplification of technological shocks, i.e., greater consumption volatility than suggested by technological condition ($\sigma_{\Delta c} > \sigma_{\Delta a}$), occurs if and only if

$$\rho_{\Delta a \Delta e} > -\frac{1}{2}\frac{\sigma_{\Delta e}}{\sigma_{\Delta a}} \tag{35}$$

where $\rho_{\Delta a \Delta e}$ measures the coefficient of correlation between Δa and Δe. As a result, two factors can give rise to amplification: a high variation in efficiency relative to the variation in productivity ($\frac{\sigma_{\Delta e}}{\sigma_{\Delta a}}$) or a high positive correlation between efficiency and productivity ($\rho_{\Delta a \Delta e}$), i.e., countercyclical dispersion of markups. Both these factors are quite intuitive.

The just described partial and general equilibrium predictions are both in principle testable. While a rigorous examination is beyond the scope of this paper, we provide a first-pass empirical inspection of the decomposition that is at the heart of this paper (see equation 34). Variation in consumption growth should be explained by variations in allocative efficiency and technological growth. Since consumption and output are equivalent in our theory, we will also use GDP growth Δy as a measure of aggregate economic activity.

6.2 Data

To compute the time series of misallocations, we require a panel data set with markups for a large number, ideally all, industries in an economy. The requirement of a large cross-section of industries makes it impossible to use state-of-the-art estimation techniques for markups, that work well for one particular industry. Instead, we make use of the standard NBER manufacturing productivity database by Bartelsman and Gray containing valuable information on 459 industries between 1959 and 2009. We exclude 8 discontinued industries leaving us with 451 industries.[27] We use (average) price cost margins (see Aghion et al. (2005)) as a proxy for markups. Thus, the empirical markup

[27] Our results are virtually equivalent when we include those industries until their year of discontinuation.

$\hat{M}_t(z)$ estimate for industry z at time t is calculated as follows:

$$\hat{M}_t(z) = 1 + PCM(z) = 1 + \frac{\text{Value added}(z) \quad - \quad \text{Payroll}(z)}{\text{Value of Shipment} \quad (z)} \tag{36}$$

While this proxy is subject to shortcomings, such as not differentiating between marginal and average costs, it represents a reasonable proxy for a large scale study as ours.[28] This procedure allows us to generate a large panel data set of markups as required by our theory.

Next, we need to compute the empirical "weight" of each industry, i.e. a proxy for $\alpha_t(z)$:

$$\hat{\alpha}_t(z) = \frac{\text{Value of Shipment} \quad (z)\, M_s(z)}{\sum_{z'=1}^{451} \text{Value of Shipment} \quad (z')\, M_s(z')}.^{29} \tag{37}$$

We can now compute the proxies for the relevant aggregate markup variables, $\hat{M}_t$ and $\hat{\eta}_t$ (see equations 12 and 13) given the discrete analog of the non-linear power mean:

$$G_p(M_s) = \left(\sum_{z=1}^{451} \alpha_t(z)\, M^p(z) \right)^{\frac{1}{p}} \tag{38}$$

The estimates $\hat{\eta}_t$ and $\hat{M}_t$ are a function of the (free) parameter θ where higher θ translates into greater misallocations fixing the empirical input $\hat{M}_t(z)$. Intuitively, misallocation created by a *given empirical* dispersion of markups is larger if goods are substitutable, since customers switch to different products.[30] While the level of θ should depend on the level of disaggregation, we use a benchmark value of $\theta = 4$.[31]

Using sector-specific four factor TFP from the manufacturing database, we can now also determine a proxy of the aggregate productivity shock weighted by each industry's size.

$$\hat{A}_t = \sum_{z=1}^{451} \alpha_t(z)\, A_t(z) \tag{39}$$

[28] The proxy is consistent with our theory as the production function is constant returns to scale in labor (see De Loecker (2011)).

[29] This proxy directly follows from our theory based on rearranging equation 19 for $\alpha_s(z) = \frac{\pi_s(z)}{C_s M_s^{\theta-1}} \frac{M_s(z)^\theta}{M_s(z)-1}$, using $\pi(z) =$ Value added(z) - Payroll(z), and setting $\theta = 1$.

[30] For this result to hold, it is important to understand that this comparative static fixes the distribution of markups (as given by the data). In contrast, if markups are produced by a model, then smaller substitutability generates higher monopoly markups and allows for greater dispersion.

[31] Our results are robust to using different parameters for θ.

Finally, we obtain yearly NIPA non-durable consumption data and GDP estimates from the BEA.

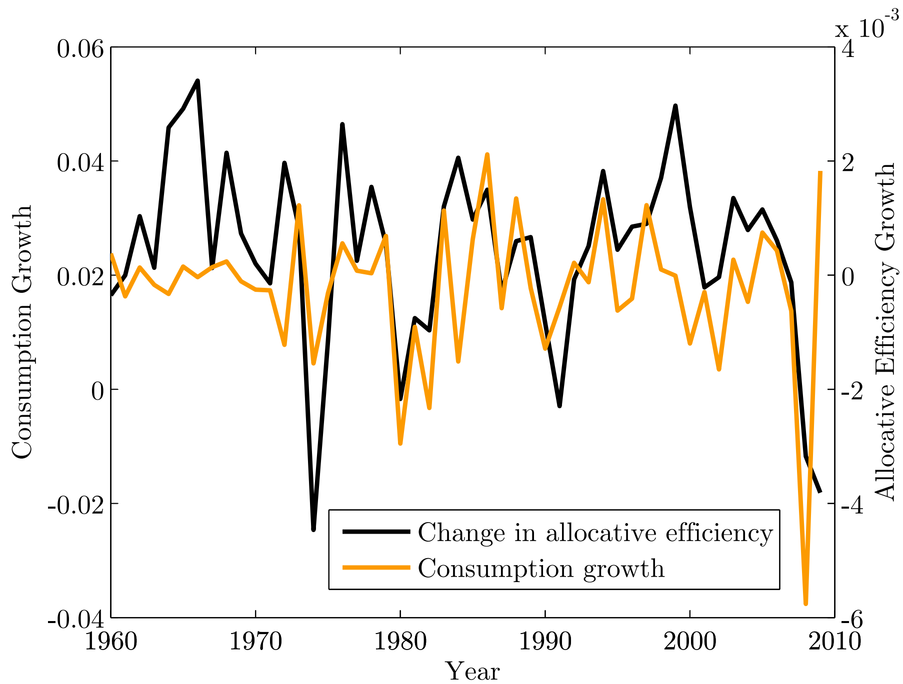

Figure 5. This figure plots the time-series of yearly (NIPA non-durable) consumption growth Δc and changes in allocative efficiency Δe implied by equation 13 and 38. The data covers the 50-year period between 1960 and 2009.

In our empirical analysis, we aim to relate these measures of aggregate economic activity, consumption growth Δc and GDP growth Δy, to efficiency growth Δe and technological growth Δa. Figure 5 plots the time series of Δc and Δe and reveals a strong positive correlation (with the exception of the outlier in the final year, i.e., the $2008 - 2009$ financial crisis). The correlation coefficient is 0.41 including the outlier and 0.54 without the final year. The graph looks similarly if we measure economic activity with GDP, as also becomes evident from Table 3 in which we also include our TFP growth proxy as an explanatory variable for both measures of economic activities, Δc and Δy. TFP growth is positively correlated with both measures (see columns 2 and 5). The increase in the R^2 in the multivariate regression of economic activity relative to both univariate regressions (see columns 3 and 6) reveals that allocative efficiency and the technological growth component capture different sources of variation.

We want to conclude this section with various shortcomings that future empirical work could address. As mentioned above, we did not consider adjustments for marginal

	Δc	Δc	Δc	Δy	Δy	Δy
Constant	0.025***	0.022***	0.023***	0.032***	0.028***	0.029***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Δe	4***		4.641***	4.045***		4.888***
	(1.270)		(1.169)	(1.771)		(1.654)
Δa		0.099***	0.123***		0.135***	0.16***
		(0.041)	(0.037)		(0.055)	(0.052)
R^2	0.17	0.11	0.33	0.1	0.11	0.25
N	50	50	50	50	50	50

Table 3. Misallocation, TFP, and Aggregate Economic Activity: OLS time-series regressions from 1960 – 2009 (i) with annual consumption growth, Δc, as the dependent variable in the first three columns, and (ii) annual GDP growth, Δy, as the dependent variable in the last three columns. The variable Δe refers to the log change in efficiency based on the cross-sectional dispersion of markups across 451 industries in the NBER manufacturing database (see equation 38). The variable Δa refers to the log change in the weighted average TFP of the 451 industries (see equation 39). Standard errors are in parentheses. *** denotes statistical significance at the 1% level.

costs or capital costs. To the extent that these adjustments are sector specific, our proxies for markups and dispersion thereof are mismeasured. However, the induced measurement error should bias our results against finding statistical significance. Also, while our data captured a large cross-section of industries, our data set only included manufacturing firms, an important, but not certainly not fully representative part of the economy. We hypothesize that markup dispersion within a group, i.e., the manufacturing sector, should be smaller than across all sectors. Thus, our measure of allocative efficiency would be downward biased. As long as this bias is close to a multiplicative constant, the statistical significance of our growth regressions would be unaffected. Finally, while our model assumes no input-output relation between producers, empirical work should take into account the entire markup chain to determine efficiency losses.

7 Concluding Remarks

We have developed general equilibrium in a dynamic economy with a continuum of different industries, each of which comprises a finite number of firms. The framework is tractable, and the strategic interaction between firms in each industry is straightforward to characterize. We establish the existence of general equilibrium and establish dynamic properties of the economy including equilibrium markups, firm profits and aggregate consumption.

The central premise of our model is that firms, maximizing shareholder value, are not always price takers but can be price setters. High prices in an industry can be sustained if firms value the future flow of profits over any immediate increases in market share garnered by undercutting. Of course, the rate at which future profits are discounted depends both on the representative agent's preferences and on the behavior of the aggregate economy. Specifically, the misallocation of resources that arises from the equilibrium cross-sectional dispersion of markups affects aggregate consumption and therefore the representative agent's valuation of future profits. This feedback effect between industry equilibrium and the macro economy is the central intuition in our paper.

The strategic interaction yields various general equilibrium effects that can be interpreted in light of the macro-economy. Even in an economy with no aggregate uncertainty, if the relative productivity of various industries changes, so does their ability to sustain collusive outcomes. These changes can affect both the level and the volatility of aggregate consumption; in short our model exhibits endogenous volatility.

An interesting extension of our model would be to consider asset pricing implications. The sub game perfect industry equilibria that we characterize naturally pins down the future value of each firms' cash flows. This of course, is the unlevered equity value of the firm. With an appropriate calibration, one could generate the relationship between returns, industry characteristics and the macro economy. We plan to explore these relationships in future work.

In conclusion, it is worthwhile highlighting how industry heterogeneity drives our results. In an economy with homogeneous industries as in Rotemberg and Woodford (1992), the markup in each industry is the same and therefore the equilibrium allocation of labor is efficient. This also precludes amplification of idiosyncratic industry shocks. Thus, our extension of their framework to allow for more realism, i.e., industry heterogeneity, generates a rich set of novel predictions.

A Proofs

Proof of Proposition 1

As explained in Section 3.1 we focus on time-invariant economies, so that all variables are solely expressed as state-dependent. Using the expression for prices, $p_s(z)$, (see equation 7) and the definitions of $\alpha_s(z)$, $\bar{A}_s$ and $\bar{M}_s$ (see equations 10, 11, and 12), we can solve for nominal prices and the nominal wage rate via normalizing the price index $P_s = \left(\int_0^1 p_s(z)^{1-\theta} dz\right)^{\frac{1}{1-\theta}}$ to one.

$$w_s = \frac{\bar{A}_s}{\bar{M}_s}, \tag{40}$$

$$p_s(z) = \frac{M_s(z)}{\bar{M}_s} \alpha_s(z)^{\frac{1}{1-\theta}}. \tag{41}$$

Finally, plugging the demand function of each sector, $c_s(z)$ (see equation 4) into the profit function of each sector $\pi_s(z)$ (see equation 6) yields an expression for y_s via the aggregate budget constraint (see equation 5)

$$y_s = \bar{A}_s \eta_s, \tag{42}$$

where we have used the expression for nominal wages and prices (see equations 40 and 41) and the definition of η_s (see equation 13). Since the price index is normalized to one, real consumption $C_s = \frac{y_s}{P_s}$ is given by y_s. The fraction of income derived by labor income, $\omega_s = \frac{w_s}{y_s}$, is readily obtained via equations 40 and 42. Real profits following immediately from 4, 6, 40, 41, and 42.

Proof of Proposition 1

The lemma is a special case of the following general lemma (by choosing $b = \Theta^T \iota_j$).

Lemma 4. *Consider a strictly positive vector $\pi^m \in \mathbb{R}^S_{++}$, a strictly positive matrix $\Theta \in \mathbb{R}^{S\times S}_{++}$, and a scalar $n \in \mathbb{R}_{++}$. Then there is a unique $\xi \in \mathbb{R}^S_+$ so that for all strictly positive $b \in \mathbb{R}^S_{++}$,*

$$\xi = \arg\max_x \; b^T x, \;\; s.t.,$$
$$x \leq \pi^m,$$
$$0 \leq (\Theta - nI)\, x.$$

For each s, the solution has either the first or the second constraint binding, i.e., for each s, $\xi_s = \pi^m_s$ or $n\xi_s = \Theta\xi_s$.

Proof: Let $x < y$ denote that $x \leq y$ and $x \neq y$. Also, define $z = x \vee y \in \mathbb{R}^S$, where $z_s = \max(x_s, y_s)$ for all s. Clearly, $x \leq x \vee y$, where the inequality is strict if there is an s such that $y_s > x_s$. Finally, define the set $K = \{x : 0 \leq x, \, x \leq \pi^*, nx \leq \Theta x\}$. Note that K is compact.

Now, there is a unique maximal element of K, that is, there is a unique $\xi \in K$, such that for all $x \in K$ such that $x \neq \xi$, $\xi > x$. This follows by contradiction, because assume that there are two distinct maximal elements, y and x, then clearly $z = x \vee y$ is strictly larger that both x and y. Now, it is straightforward to show that $z \in K$. The only condition that is not immediate is that $\Theta z \geq Nz$. However, this follows from $\Theta(x \vee y) \geq \Theta x \vee \Theta y \geq nx \vee ny = n(x \vee y) = nz$.

Now, since b is strictly positive, it is clear that ξ is indeed the unique solution to the optimization problem regardless of b. That one of the constraint is binding for each s also follows directly, because assume to the contrary that neither constraint is binding in some state s. Then ξ_s can be increased without violating either constraint in state s and, moreover, the constraints in all the other states will actually be relaxed, so such an increase is feasible. Further, since $b_s > 0$, it will also increase the objective function, contradicting the assumption that ξ is optimal.

Proof of Lemma 2

By definition: $V = \Lambda_\pi(\Gamma - \mathbf{1})$, so from (**??**), $\Lambda_\pi(\Gamma - \mathbf{1}) = \delta\Lambda_m^{-1}\Phi\Lambda_m(\pi + \Lambda_\pi(\Gamma - \mathbf{1}))$, leading to $\Gamma - \mathbf{1} = \delta\Lambda_\pi^{-1}\Lambda_m^{-1}\Phi\Lambda_m\Lambda_\pi\Gamma$. Now, observing (from (**??**)) that $\Lambda_\kappa = \Lambda_\pi\Lambda_m$, the result follows immediately.

Proof of Proposition 2

Let $n = N - 1$ and $K^*(n) \overset{\text{def}}{=} \{x : 0 \le x, nx \le \Theta x\}$. Now, $nx \le (\Lambda_m^{-1}(I - \delta\Phi)^{-1}\Lambda_m - I)$ is equivalent to $Ny \le (I - \delta\Phi)^{-1}y$, where $y = \Lambda_m x \in \mathbb{R}_+^S$. We first show that $K^*(n) = \{0\}$ when $N > \frac{1}{1-\delta}$, which immediately implies that the only solution to the optimization problem in Lemma **??** is indeed the competitive outcome. Define the matrix norm $\|A\| = \sup_{x\in\mathbb{R}^S\setminus\{0\}} \frac{\|Ax\|}{\|x\|}$, where the l^1 vector norm $\|y\| = \sum_s |y_s|$ is used. Since Φ is a stochastic matrix, $\|\Phi^i\| = 1$ for all i and using standard norm inequalities it therefore follows immediately that

$$\|(I - \delta\Phi)^{-1}\| = \left\|\sum_0^\infty \delta^i\Phi^i\right\| \le \sum_0^\infty \delta^i\|\Phi^i\| = \frac{1}{1-\delta},$$

and thus $\|(I - \delta\Phi)^{-1}y\| \le \frac{1}{1-\delta}\|y\|$. Now, $Ny \le (I - \delta\Phi)^{-1}y$ implies that $N\|y\| \le \|(I - \delta\Phi)^{-1}y\|$, and therefore it must be the case that $N \le \frac{1}{1-\delta}$, for the inequality to be satisfied for a non-zero y. Now, consider the case when $N = \frac{1}{1-\delta}$. Since $y = 1$ is an eigenvector to Φ with unit eigenvalue, it is also an eigenvector to $(I - \delta\Phi)^{-1}$ with corresponding eigenvector $\frac{1}{1-\delta}$, leading to $x = \Lambda_m^{-1}1 = m^{-1}$. It is easy to show that this is the unique (up to multiplication) nonzero solution. Given the properties of Φ, the Perron-Frobenius theorem implies that this is indeed the *only* eigenvector with unit eigenvalue, and therefore also the only eigenvector to $(I-\delta\Phi)^{-1}$ with eigenvalue $\frac{1}{1-\delta}$. Now, take an arbitrary $y \in \mathbb{R}_+^S\setminus\{0\}$ as a candidate vector to satisfy the inequality, i.e., such that $z = (I - \delta\Phi)^{-1}y$ satisfies $z_i \ge Ny_i = \frac{1}{1-\delta}y_i$ for all i. Then, since $\|(I - \delta\Phi)^{-1}\| = \frac{1}{1-\delta}$, it follows that $\sum_i z_i \le \frac{1}{1-\delta}\sum_i y_i$. The two inequalities can only be satisfied jointly if $z_i = \frac{1}{1-\delta}y_i$ for all i, and thus y is the already identified eigenvector. Thus, $K^*\left(\frac{1}{1-\delta}\right) = \{\iota m^{-1}, \iota \ge 0\}$. It follows immediately from the definition of the λ vector that the maximal ι that satisfies $\iota m_s^{-1} \le \pi_s^* = q_s C_s \alpha_s$ for all s is $\min_s \lambda_s$, leading to the given form of the profit vector.

Proof of Lemma 3

If $\kappa_s = k$, the diagonal matrix Λ_κ becomes $\Lambda_\kappa = kI$ so that we obtain for Γ (see (27)):

$$\Gamma = (I - \delta\Phi)^{-1}\mathbf{1} = \frac{1}{1-\delta}\mathbf{1} = N^c\mathbf{1}. \tag{43}$$

This is because the eigenvalue of $(I - \delta\Phi)^{-1}$ associated with the eigenvector of $\mathbf{1}$ is given by $\frac{1}{1-\delta}$ (see Proof of Proposition 2). So, $N^m = \min_s(\Gamma_s) = N^c$.

Proof of Proposition 3

(1,2) follow from the definition of K in the proof of Lemma **??**. It immediately follows that the set K is decreasing in N and increasing in each of α_s, which in turn immediately implies (1,2).

(3) follows from (1), and the fact that $\pi_s > 0$ for all s when the number of firms is N^c.

(4) follows from (1) and that $\pi_s = m_s^{-1}\pi_s^m m_s$ for the s that minimizes μ_s (see Proposition 3).

(5) follows from the fact that the objective function in Lemma **??** is a continuous function of all parameters and that (as long as N is strictly below N^c) the set K is compact, and depends continuously on all parameters, in the sense that if K and K' are defined for two sets of parameter values, then $D(K, K')$ approaches zero when the parameter values that define K' approach those that define K. Here, $D(K, K') = \sup_{x\in K'} \inf_{y\in K} |x - y|$.

Proof of Proposition 4

Before showing existence, we discuss some invariance results which will be helpful in the proof. We first note that the following result follows immediately from Proposition 2:

Lemma 5. *In any general equilibrium, any two industries with the same N and α have the same markups, M, and profits, π.*

Also, we observe that it is only the distributional properties of N and α that are important for the aggregate characteristics of an equilibrium. This should come as no surprise given that the aggregate variables important for industry equilibrium only depend on the distributions. To be specific, we define the (cumulative) distribution function $F : \mathbb{N} \times [c, C]^S \to [0, 1]$, where $F(n, s_1, \ldots, s_S) = \lambda(\{(z : N(z) \leq n \wedge \alpha_1(z) \leq s_1 \wedge \cdots \wedge \alpha_S(z) \leq s_S\})$, and λ denotes Lebesgue measure. Thus, $F(n, \alpha_1, \ldots, \alpha_S)$ denotes the fraction of industries with number of firms less than or equal to n, and productivities $\alpha_s(z) \leq \alpha_s$ for all s. We say that two economies, $\mathcal{E}_1$ and $\mathcal{E}_2$, are equivalent in distribution if they have the same distribution functions, and agree on the other parameters: g, $\bar{A}$, Φ, γ, θ and $\hat{\delta}$. Also, two outcomes—in two different economies—are said to be equivalent if any two industries, z and z' in the first and second economy, respectively, for which $N^1(z) = N^2(z')$ and $\alpha_s^1(z) = \alpha_s^2(z')$ for all s, have the same industry markups in each state of the world, $M_s^1(z) = M_s^2(z')$ for all s.

We then have

Lemma 6. *Given two economies that are equivalent in distribution. Then for each equilibrium in one of the economies there is an equivalent equilibrium in the other.*

We now prove the proposition with a fixed point argument, and therefore define a fixed point relationship for the markup function, M, which ensures that it defines an equilibrium. We define $R \stackrel{\text{def}}{=} \bar{N} \times [c, C]^S$, where $\bar{N} = \{1, 2, \lfloor N_c \rfloor + 1\}$, with elements $x = (n, \alpha_1, \ldots, \alpha_S) \in R$. We will then work with functions $M^0 : R \to [0, 1]^S$, and given such a function, the transformation to the standard markup function is given by $M_s(z) = M_s^0(\min(N(z), \lfloor N_c \rfloor + 1), \alpha_1(z), \ldots, \alpha_S(z))$. The reason why we work with the canonical domain, R, rather than $S \times [0, 1]$, is that compactness properties needed for a fixed point argument are easier obtained in this domain. Given a function, $M^0 : R \to \left[1, \frac{\theta}{\theta-1}\right]^S$, we define

$$p_s^0 = G_{-\theta}(M_s) = \left(\int \alpha_s(z) M_s(z)^{-\theta} dz\right)^{\frac{1}{-\theta}} = \left(\int_{x\in R} x_{s+1} M^0(x)^{-\theta} dF(x)\right)^{\frac{1}{-\theta}}, \tag{44}$$

$$p_s^1 = G_{1-\theta}(M_s) = \left(\int \alpha_s(z) M_s(z)^{1-\theta} dz\right)^{\frac{1}{1-\theta}} = \left(\int_{x\in R} x_{s+1} M^0(x)^{1-\theta} dF(x)\right)^{\frac{1}{1-\theta}}. \tag{45}$$

It follows immediately that the mapping from M^0 to p_0 and p_1 is continuous (in L^1 topology) and since $\int \alpha(z)dz = 1$, that p_s^0 and p_s^1 lie in $[1, \theta/(\theta-1)]$. From (14), it follows that

$$C_s = \bar{A}_s \left(\frac{p_1}{p_0}\right)^\theta, \tag{46}$$

and from (**??**) that

$$\pi_s^m = \frac{1}{p_1^{1-\theta}} \frac{(\theta-1)^{\theta-1}}{\theta^\theta} \alpha_s C_s = \frac{1}{p_1^{1-\theta}} \frac{(\theta-1)^{\theta-1}}{\theta^\theta} x_{s+1} C_s. \tag{47}$$

Now, for each z, given $\pi^m \in \mathbb{R}_+^S$, the program in Lemma **??** provides a continuous mapping from π^m to

$$\pi_s \in \prod_1^S [0, \pi_s^m]. \tag{48}$$

We use (19) to define the operator $\mathcal{F}$, which operates on functions, and which is given by:

$$M_s^1(x) = (\mathcal{F}(M^0)(x))_s = 1 + \frac{p_1(s)^{1-\theta}}{C_s x_{s+1}} (M_s^0(x))^\theta \pi_s.$$

Since each operation in (44-48) is continuous, it follows that $\mathcal{F}$ is a continuous operator (in $L^1(\mathbb{R}^{1+S})$-norm). Further, it also follows that if $M_s^0(x) \in \left[1, \frac{\theta}{\theta-1}\right]$, then since $0 \leq \pi \leq \pi^m$, $1 \leq M_s^1(x) \leq 1 + \frac{(\theta-1)^{\theta-1}}{\theta^\theta}(M_s^0)^\theta \leq \frac{\theta}{\theta-1}$. Define, Z as the set of all functions, $M : R \to [1, \theta/(\theta-1)]^S$, such that M is nonincreasing in its first argument and nondecreasing in all other arguments. Then, from what we have just shown, together with Proposition 3, it follows that $\mathcal{F}$ is a continuous operator that maps Z into itself. We also have

Lemma 7. *Z is convex and compact.*

We prove that the set, W, of nondecreasing functions $f : [0,1] \to [0,1]$, is convex and compact. The generalization to functions with arbitrary rectangular domains and ranges, $f : \prod_1^N [a_i, b_i] \to \prod_1^M [c_i, d_i]$, is straightforward, as is the generalization to functions that are nonincreasing in some coordinates and nondecreasing on others (as is Z). Convexity is immediate. For compactness, we show that every sequence of functions $f^n \in W$, $n = 1, 2, \ldots$, has a subsequence that converges to an element in W. First, note that W is closed, since a converging (Cauchy) sequence of nondecreasing functions necessarily converges to a nondecreasing function. To show compactness, define the corresponding sequence of vectors $g^n \in [0,1]^{2^j}$, for some $j \geq 1$, by $g_k^n = f_n(2^{-j}k)$, $k = 0, 1, \ldots 2^j - 1$. Now, since $[0,1]^{2^j}$ is compact it follows that there is a subsequence of $\{f^n\}$, $\{f^{n_m}\}$ that converges at each point $2^{-j}k$, to some $g^* \in [0,1]^{2^j}$. Define the function $h^j : [0,1] \to [0,1]$ by $h^j(x) = g_k^*$, for $2^{-j}k \leq x < 2^{-j}(k+1)$, which is obviously also in W. Next, take the sequence $\{f^{n_m}\}$, and use the same argument to find a subsequence that converges in each point $2^{-(j+1)}k$, $k = 0, \ldots, 2^{j+1} - 1$, and the corresponding function $h^{j+1}(x)$. By repeating this step, we obtain a sequence of functions in W, $h^j, h^{j+1}, \ldots$, such that for $m > j$,

$$\int_0^1 |h^m(x) - h^j(x)| dx \leq \sum_k (g_{k+1}^j - g_k^j) 2^{-j} \leq 2^{-j}.$$

Thus, $h^j, h^{j+1}, \ldots$ forms a Cauchy-sequence, which consequently converges to some function $h^* \in W$. Take a subsequence of the original sequence of functions, $\{f^{n_j}\}$, such that $\int |f^{n_j} - h^j| dx \leq 2^{-j}$. Then, for $m > j$, since

$$\begin{aligned}\int_0^1 |f^{n_m}(x) - f^{n_j}(x)|dx &= \int_0^1 |f^{n_m}(x) + h^m(x) - h^m(x) + h^j(x) - h^j(x) - f^{n_j}(x)|dx \\ &\leq \int_0^1 |f^{n_m}(x) - h^m(x)|dx + \int_0^1 |f^{n_j}(x) - h^j(x)|dx \\ &+ \int_0^1 |h^m(x) - h^j(x)|dx \\ &\leq 3 \times 2^{-j},\end{aligned}$$

$\{f^{n_j}\}$ is also a Cauchy sequence and converges to $h^* \in W$. Thus, W is compact and the lemma is proved. Given Lemma 7 and the continuity of $\mathcal{F}$, a direct application of Schauder's fixed point theorem

implies that there is a $M^* \in Z$, such that $\mathcal{F}(M^*) = M^*$. Now, given such a M^*, and its associated π^m defined by (47), and given the functions, $N(z)$ and $\alpha_s(z)$, $0 \leq z \leq 1$, Lemma **??** can be used to construct $M_s(z)$. Since M and M^* have the same distributional properties, and C, p_0 and p_1, only depend on distributional properties, it immediately follows that M constitutes an equilibrium. We are done.

Proof of Proposition 5

First note that an equivalent formulation of Lemma **??** is the following: Define the sets $V_s = \{x \in \mathbb{R}_+^S : x_s \leq \pi_s^m\}$, $Q_s = \{x \in \mathbb{R}_+^S : 0 \leq ((\Theta - nI)\, x)_s\}$, and $R = (\cap_{s=1}^S V_s) \cap (\cap_{s=1}^S Q_s)$. Then there is a unique element, $r \in R$, such that for all s, $r_s = \max_{q \in R} q_s$. That is, there is a unique element that jointly maximizes all coordinates of elements in R. Moreover, for each s, such that $r_s < \pi_s^m$ it must be that $r_s = \frac{1}{n}(\Theta x)_s$.

For coordinates such that $r_s < \pi_s^m$, if any number of the π_s^m is replaced by $\hat{\pi}_s^m > \pi_s^m$, i.e., if V_s is replaced by $\hat{V}_s = \{x \in \mathbb{R}_+^S : x_s \leq \hat{\pi}_s^m\}$, where $\hat{\pi}_s^m \geq \pi_s^m$, and the equality is only allowed to be strict for coordinates where $r_s < \pi_s^m$, and $\hat{R}$ is defined as $R = (\cap_{i=1}^S \hat{V}_s) \cap (\cap_{i=1}^S Q_s)$, then $\hat{R} = R$, and consequently, $\hat{r} = r$ where $\hat{r}$ is the unique maximal element in $\hat{R}$. To see this, assume that an element $v \in \hat{R}$ existed such that $v_s > \pi_s^m$ for at least one s. Then since $\hat{R}$ is convex there must also be an element, $w = \lambda r + (1-\lambda) v \in \hat{R}$, with $w_s \leq \pi_s^m$, for all s and $w_s = \pi_s^m$ for one coordinate such that $r_s < \pi_s^m$. But then $w \in R$, and it must then be that $r_s = \pi_s^m$, leading to a contradiction. Thus, no such element exists, so $\hat{R} = R$.

Now, from our discussion in Section 3.3, it follows that in an equilibrium in a homogeneous economy, all firms must charge the same markups in any state, $M_s(z) = \bar{M}_s$ for all z, and that any equilibrium must be efficient so that $C_s = \bar{A}_s = A_s(z)$ and $\alpha_s(z), = 1$ for all s for all z. What is not a priori clear is whether there may be multiple average markup vectors, $\bar{M}$, that constitute an equilibrium. We now show that this is not the case.

Given an equilibrium in a homogeneous economy, it follows from equation (19), and that $C_s = A_s$, that

$$\frac{1}{\bar{M}_s} = 1 - \frac{\pi_s}{A_s} = (1 - u_s). \tag{49}$$

Here $u_s = \frac{\pi_s}{A_s} \in \left[0, \frac{1}{\theta}\right]$ represents firm profits in state s as a fraction of total output.

It further follows from equation (**??**) that given such an average markup across industries, the monopolistic profits as a fraction of total output in one (zero-measure) industry, z, that deviates from the average markup function is $\hat{u}_s = \frac{\hat{\pi}_s^m}{A_s} = \zeta \bar{M}_s^{\theta-1} = \zeta(1-u_s)^{1-\theta}$. We note that $\hat{u}_s \geq u$ for all $u \in \left[0, \frac{1}{\theta}\right]$, and that the inequality is strict except for at $\hat{u} = u = \frac{1}{\theta}$.

Given the homogeneous behavior of all other industries, the firm optimization problem in (23-25) can be written

$$\hat{u} = \arg\max_{\hat{u}} \ \iota_j^T \Lambda_A^{-1} \Theta \Lambda_A \hat{u}, \quad \text{s.t.,} \tag{50}$$

$$\hat{u}_s \leq \zeta (1 - u_s)^{1-\theta}, \quad s = 1, \dots, S, \tag{51}$$

$$0 \leq \left(\Lambda_A^{-1} \Theta \Lambda_A - (N-1) I \right) \hat{u}, \tag{52}$$

where $\Lambda_A = diag(\bar{A}_1, \dots, \bar{A}_s)$. A necessary and sufficient condition for u to be an equilibrium is now that $\hat{u} = u$ in the above optimization problem.

Assume that we have found such a u (we know that there exists at least one such u from the existence theorem). If we can show that u is also the solution to the same program, but where (51) is replaced by $\hat{u}_s \leq \frac{1}{\theta}$ for all s, then we are done, since there is a unique solution for that optimization problem (as follows from an identical argument as the proof of Lemma 4).

An identical argument as in Lemma 4 implies that for each s, either (51) or (52) binds (or both). For any s such that (51) binds, it must further be that equilibrium markups in that state are monopolistic, i.e., $u = \frac{1}{\theta}$. Thus, relaxing the constraints for those s to $\hat{u}_s \leq \frac{1}{\theta}$ does not change the solution to the problem.

For any other s, where (51) does not bind and (52) binds, we note that since $u_s < \frac{1}{\theta}$, $u_s < \zeta(1-u_s)^{1-\theta}$, $\hat{u}_s$ is strictly lower than its bound imposed by (51) for such s. However, from the argument at the beginning of this lemma, it follows that relaxing the constraint for these coordinates does not change the solution, so we can relax the constraints to $\hat{u}_s \leq \frac{1}{\theta}$ for such s too. Thus, u is also a solution to the relaxed problem, and is therefore unique. We are done.

B Long Term Growth

When $g > 0$, we can still solve for time-invariant equilibria through appropriate normalizations. That is, we focus on equilibria which—except for the constant growth rate g—are time invariant in that outcomes are the same at t_1 and t_2 if the states are the same, i.e., if $s_{t_1} = s_{t_2}$. In such equilibria, outcomes *on the equilibrium path* can be written as:

$$C(t) = (1+g)^t C_{s_t}, \tag{53}$$

$$y(t) = (1+g)^t y_{s_t}, \tag{54}$$

$$w(t) = (1+g)^t w_{s_t}, \tag{55}$$

$$\pi(z,t) = (1+g)^t \pi_{s_t}(z), \tag{56}$$

$$c(z,t) = (1+g)^t c_{s_t}(z), \tag{57}$$

where variables on the right hand side are growth-normalized, time invariant, variables which only depend on the state, s_t. We want to emphasize that this formulation does not impose any restriction on *off-equilibrium path* behavior. Thus, the equilibria that we exhibit also exist in the broader class.

The focus on time invariant equilibria is natural, since we prove that optimizing firm behavior in one particular industry is endogenously time invariant provided that all other industries exhibit time-invariant behavior. Moreover, it is ensured that (at least) one time-invariant equilibrium exists (see Proposition 4). In such an economy we immediately obtain that markups are time-invariant

$$M(z,t) = M_{s_t}(z). \tag{58}$$

It follows from a standard transformation, using the utility representation (equation 2), that growth-normalized variables can be determined by solving the model for a non-growing economy with a growth-adjusted personal discount rate, i.e., with

$$\hat{\delta} \stackrel{\text{def}}{=} (1+g)^{1-\gamma}\delta. \tag{59}$$

Intuitively, the representative agent's trade-off between consumption in different times and states is affected in identical ways by changes in the growth rate and the subjective discount factor. Thus, the effective discount rate in a growing economy, $\hat{\delta}$, depends on long term growth rates. The importance of long-term growth rates for asset pricing was recently discussed in Parlour et al. (2011). In that paper, long-term growth rates are important because they determine how much investors care about rare disaster events in the far future.

References

Abreu, D., 1988, On the theory of infinitely repeated games with discounting, *Econometrica* 56, 383–96.

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi, 2011, The network origins of aggregate fluctuations, *Econometrica* forthcoming.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt, 2005, Competition and innovation: An inverted-u relationship, *The Quarterly Journal of Economics* 120, 701–728.

Bagwell, K., and R. Staiger, 1997, Collusion over the business cycle, *RAND Journal of Economics* 28, 82–106.

Benassy, J. P., 1978, The objective demand curve in general equilibrium with price makers, *Economic Journal* .

Bilbiie, F. O., F. Ghironi, and M. J. Melitz, 2008, Monopoly power and endogenous product variety: Distortions and remedies, NBER Working Papers 14383, National Bureau of Economic Research, Inc.

Bilbiie, F. O., F. Ghironi, and M. J. Melitz, 2012, Endogenous entry, product variety, and business cycles, *Journal of Political Economy* 120, 304 – 345.

Bils, M., P. J. Klenow, and B. A. Malin, 2012, Testing for Keynesian labor demand, Nber macroeconomics annual, National Bureau of Economic Research, Inc.

Calvo, G. A., 1983, Staggered prices in a utility-maximizing framework, *Journal of Monetary Economics* 12, 383–398.

Chevalier, J. A., and D. S. Scharfstein, 1995, Liquidity constraints and the cyclical behavior of markups, *American Economic Review* 85, 390–96.

Chevalier, J. A., and D. S. Scharfstein, 1996, Capital-market imperfections and countercyclical markups: Theory and evidence, *American Economic Review* 86, 703–25.

dal Bo, P., 2007, Tacit collusion under interest rate fluctuations, *The RAND Journal of Economics* 38, pp. 533–540.

De Loecker, J., 2011, Recovering markups from production data, *International Journal of Industrial Organization* 29, 350–355.

Dhingra, S., and J. Morrow, 2012, The impact of integration on productivity and welfare distortions under monopolistic competition, CEP Discussion Papers dp1130, Centre for Economic Performance, LSE.

Dixit, A. K., and J. E. Stiglitz, 1977, Monopolistic competition and optimum product diversity, *American Economic Review* 67, 297–308.

Edmond, C., V. Midrigan, and D. Y. Xu, 2012, Competition, markups, and the gains from international trade, Working Paper 18041, National Bureau of Economic Research.

Epifani, P., and G. Gancia, 2011, Trade, markup heterogeneity and misallocations, *Journal of International Economics* 83, 1–13.

Gabaix, X., 2011, The granular origins of aggregate fluctuations, *Econometrica* 79, 733–772.

Gabszewicz, J. J., and J. P. Vial, 1972, Oligopoly 'a la cournot' in general equilibrium analysis, *Journal of Economic Theory* 4, 381–400.

Gali, J., 1994, Monopolistic competition, business cycles, and the composition of aggregate demand, *Journal of Economic Theory* 63, 73–96.

Holmes, T. J., W.-T. Hsu, and S. Lee, 2004, A joint econometric model of macroeconomic and term structure dynamics, Working Paper 405, Univeristy of Minnesota.

Hsieh, C.-T., and P. J. Klenow, 2009, Misallocation and manufacturing tfp in china and india, *The Quarterly Journal of Economics* 124, 1403–1448.

Jaimovich, N., 2007, Firm dynamics and markup variations: Implications for sunspot equilibria and endogenous economic fluctuations, *Journal of Economic Theory* 137, 300–325.

Jaimovich, N., and M. Floetotto, 2008, Firm dynamics, markup variations and the business cycle, *Journal of Monetary Economics* 55, 1238–1252.

Jovanovic, B., 1987, Micro shocks and aggregate risk, *Quarterly Journal of Economics* 102, 395–409.

Kung, H., and L. Schmid, 2012, Innovation, growth, and asset prices, Working Paper 405, Universit of British Columbia and Duke University.

Kydland, F. E., and E. C. Prescott, 1982, Time to build and aggregate fluctuations, *Econometrica* 50, 1345–1370.

Lerner, A. P., 1934, The concept of monopoly and the measurement of monopoly power, *The Review of Economic Studies* 1, 157–175.

Long, B., and C. I. Plosser, 1983, Real business cycles, *Journal of Political Economy* 91, 39–69.

Marschak, T., and R. Selten, 1974, *General Equilibrium with price-making firms, Lecture Notes in Economics and Mathematical Systems* (Springer-Verlag, Berlin).

Nekarda, C. J., and V. A. Ramey, 2012, The cyclical behavior of the price-cost markup, Working paper, Board of Governors of the Federal Reserve System and UC Davis.

Opp, M. M., 2010, Tariff wars in the ricardian model with a continuum of goods, *Journal of International Economics* 80, 212–225.

Opp, M. M., C. Parlour, and J. Walden, 2012, Industrial asset pricing, Working paper, UC Berkeley, Haas School of Business.

Parlour, C., R. Stanton, and J. Walden, 2011, Revisiting asset pricing puzzles in an exchange economy, *Review of Financial Studies* 24, 629–674.

Peters, M., 2012, Heterogeneous mark-ups and endogenous misallocation, Working paper, Massachusetts institute of Technology.

Ravn, M., S. Schmitt-Grohe, and M. Uribe, 2006, Deep habits, *Review of Economic Studies* 73, 195–218.

Rotemberg, J. J., and G. Saloner, 1986, A supergame-theoretic model of price wars during booms, *American Economic Review* 76, 390–407.

Rotemberg, J. J., and M. Woodford, 1992, Oligopolistic pricing and the effects of aggregate demand on economic activity, *Journal of Political Economy* 100, 1153–1207.

Samuelson, P., 1949, *Foundations of Economic Analysis* (Harvard University Press, Cambridge).

Schmitt-Grohe, S., 1997, Comparing four models of aggregate fluctuations due to self-fulfilling expectations, *Journal of Economic Theory* 72, 96–147.

van Binsbergen, J. H., 2007, Deep Habits and the Cross Section of Expected Returns, *SSRN eLibrary* .

Woodford, M., 1986, Stationary sunspot equilibria: The case of small fluctuations around a deterministic steady state, Technical report, University of Chicago, Unpublished manuscript.

Woodford, W., 1991, Self-fulfilling expectations and fluctuations in aggregate demand, in N. Mankiw, and D. Romer, eds., *New Keynesian Economics* (MIT Press, Cambridge, MA).