Time Use and the Efficiency of Heterogeneous Markups

Albrecht, Brian C. and Phelan, Thomas and Pretnar, Nick

2 August 2023

Time Use and the Efficiency of Heterogeneous Markups*

Brian C. Albrecht† Thomas Phelan‡ Nick Pretnar§

August 2, 2023

Abstract

A recent literature has provided empirical evidence that markups are increasing and are heterogeneous across firms. In standard monopolistic competition models, such heterogeneity implies inefficiency even in the presence of free entry. We enrich the standard model of monopolistic competition with heterogeneous firms to incorporate off-market time use that is nonseparable with market consumption into the consumer problem. Within this framework the constancy of equilibrium markups is neither sufficient nor necessary for efficiency. Whether or not the competitive level of production and market concentration of firms are efficient depends on the degree to which consumption time and market goods are complements or substitutes. Such inefficiencies are the result of time use being misallocated toward home production at the expense of market production.

Keywords: monopolistic competition, markups, efficiency, time use, home production, elasticity of substitution, concentration, selection, love for variety, heterogeneous firms

JEL codes: D1, D4, D6, L1

*The views stated herein are those of the authors and are not necessarily those of the Federal Reserve Bank of Cleveland or the Board of Governors of the Federal Reserve System. We thank Kristian Behrens, David Berger, Mark Bils, Benjamin Bridgman, Baris Kaymak, Roberto Pinheiro, and James Traina for helpful discussions.

†International Center for Law & Economics. Email: mail@briancalbrecht.com.

‡Federal Reserve Bank of Cleveland. Email: tom.phelan@clev.frb.org.

§UCSB, LAEF Email: npretnar@ucsb.edu.

1 Introduction

Measures of markups find large heterogeneity across firms.[1] Are heterogeneous markups inefficient? In this paper we revisit the analyses of markups and efficiency in monopolistically competitive settings previously undertaken by Dixit and Stiglitz (1977), Zhelobodko et al. (2012), and Dhingra and Morrow (2019). We consider how the connection between markups and efficiency changes in a model in which consumers' off-market time use is not separable from market consumption in the home production sense; consumers choose how many hours to work and how much time to devote to home production. Within this context, we revisit several core questions. Namely, can efficient allocations be achieved even when markups are heterogeneous? And, how is markup heterogeneity related to the elasticity of substitution across varieties of market goods when off-market time and market consumption are non-separable?

Whether heterogeneous markups are efficient will depend on the cause of markups. In standard models of monopolistic competition, like the benchmark Dixit and Stiglitz (1977) environment with constant elasticity of substitution (CES) preferences, the distribution of markups is determined solely by preferences for varieties of market consumption. More recent papers, such as Zhelobodko et al. (2012) and Dhingra and Morrow (2019), enrich the consumer's preference structure beyond CES, exploring the implications for markups and welfare when elasticities of substitution are variable (VES). Meanwhile, Behrens et al. (2020) allow for different elasticities between sectors and quantify the welfare losses associated with markups. In each paper time use is inelastic, and CES preferences are both necessary and sufficient for markups to be constant and allocations to be efficient when cost structures are heterogeneous across firms.

We show that when firms are heterogeneous in costs and off-market time use is both elastic and non-separable with market consumption, CES preferences over market consumption are necessary but not sufficient to achieve allocative efficiency. Further, CES preferences over market consumption are necessary to ensure markups are constant, but constant markups are neither necessary nor sufficient to ensure allocations are efficient. The commonly posited relationship from the literature between the constancy of markups, elasticities of substitution, and efficiency is broken under a home production preference structure.

We also examine the positive role preferences play in generating general equilibrium outcomes, in particular, markups and markup dispersion. In this way

[1] See, e.g., Epifani and Gancia (2011), De Loecker et al. (2020), Peters (2020), and Edmond et al. (2022).

our work is close in spirit to that of Parenti et al. (2017). Both Parenti et al. (2017) and our paper posit that consumers face a richer decision-making problem than what is typically assumed in the monopolistic-competition literature. In Parenti et al. (2017) consumers have uncertainty over their love for variety which yields variable markups that can (possibly) be socially optimal. Alternatively, our approach assumes consumption requires time. This microfounded enrichment of the consumer's problem can generate heterogeneous markups in equilibrium that are efficient. This is because when consumers also require time in order to derive utility from a consumption activity, the shadow price of final consumption is fundamentally different from the posted, market price. We show, under certain conditions, that this model can sometimes lead to variable markups that are socially optimal. Particularly, if the cost of final consumption is the sum of the price of market inputs and the marginal cost of a consumer's off-market time, then markups are variable but also socially optimal.

By allowing for home production, and thus non-separable market consumption and off-market time, we introduce a new channel through which distortions may operate. We show that markups over the marginal cost of production, which monopolistically competitive firms choose when making pricing decisions, are *not* the same as the markups experienced by consumers when they choose to allocate time toward *using* market purchases in off-market activities. The latter account for both the degree to which market consumption and off-market time are complimentary or substitutable and how consumers value their off-market time. Meanwhile, markups as chosen by firms (what we call "posted" markups) are merely markups on a single, but not exclusive, input in consumers' home production processes. We show that these posted markups always exceed the "holistic" markups over marginal-cost pricing that consumers actually experience. This is because, in the face of increasing market prices, consumers have several intensive margin choices: 1) they can reallocate their off-market time toward utilization of different kinds of high- or low-cost market products; 2) they can work more or less and thus supply less or more time toward off-market, home-production activities.

The results in this paper also have several normative implications. First, our main result challenges the idea that variable markups are always distortionary.[2] Given that estimates of markup dispersion suggest it has increased over the last forty years (see, e.g., De Loecker et al. (2020, Figure 3) and Flynn et al. (2019, Figure 2)), that this increase has resulted in a decline in consumer surplus relative

[2]In many models, higher markup dispersion reduces welfare. In models such as Peters (2020) and Edmond et al. (2022), markup dispersion creates misallocation and ultimately reduces TFP relative to a benchmark with no dispersion.

to an environment where such an increase had *not* occurred should be re-examined. Second, the strict focus on posted, observable markups, as opposed to the holistic markups our model generates, may overstate both the welfare implications of markups (and, by extension, market concentration) and the degree to which they have effectively impacted consumers' budgets.

Our theoretical model reflects the fundamental idea first discussed in Becker (1965) and analyzed in depth in Ghez and Becker (1974) that utility derived from off-market time (i.e., "leisure") cannot be separated from utility derived from market consumption. That is, in order to enjoy consumption, consumers must allocate time toward doing so. The literature has often referred to this process as "home production," implying that there exist shadow "commodities" or "experiences" that yield utility but which we cannot directly measure, yet which are produced by combining inputs of time and market purchases (i.e., what is often referred to as "consumption") in order to yield some final output over which consumers have utility. In this sense time-use and market consumption are non-separable in preferences, since each market purchase is also associated with a unique time-use decision.

This modeling approach is motivated by strong empirical evidence that time-use and consumption are not separable in preferences.[3] Further, the literature has only recently begun to consider the implications of incorporating non-separable preferences for consumption and leisure time into studies involving broader outcomes that are only indirectly related to the study of consumer time-use itself. Boerma and Karabarbounis (2021) and Pretnar (2022) use models with consumption and time-use non-separabilities to measure welfare inequality, each coming to opposite conclusions which depend on modeling assumptions. Meanwhile, Bridgman et al. (2018) and Bednar and Pretnar (2023) study how accounting for time-use non-separabilities in home production affect implications for structural change, while Bridgman (2016) uses a model with consumption and time use non-separabilities to study how in-home productivities associated with using different types of market products have changed over time.

The findings throughout the macroeconomics literature provide evidence that allowing for non-separabilities between market purchases and off-market time may fundamentally change other economic inferences. In this paper we add to these findings by studying the welfare implications of consumption and time-use non-separabilities in a monopolistically competitive economy.

When one incorporates off-market time, it is natural to also allow the labor

[3]See, e.g., Aguiar and Hurst (2005), Aguiar and Hurst (2007), Pretnar (2022), and Fang et al. (2022).

supply to be elastic. To this point, the literature exploring models of monopolistic competition has sometimes allowed for elastic labor, though leisure preferences are almost always separable from consumption. Bilbiie et al. (2012) explore how the allocation of labor across sectors and the number of products and producers varies over the business cycle. Bilbiie et al. (2019) consider the role that elastic factors of production (labor and capital) play in amplifying distortions on a dynamic path along which firms may enter and exit. Boar and Midrigan (2022) also have additively separable consumption and time, though they focus on how markup distortions redistribute income from laborers to entrepreneurs. Meanwhile, Edmond et al. (2022) disentangle the degree to which markups, misallocation of factors of production, and inefficient entry contribute to welfare costs.

The paper proceeds as follows. Section 2 outlines the broad ingredients required for the benchmark model. All units are denominated in units of time. Consumers are identical but enjoy experiences associated with market products of different varieties. Firms, however, are not identical, as each faces a unique cost structure. Finally, we define the problem faced by agents in a decentralized setting before also describing the choices faced by a planner seeking to tax/subsidize labor, production, and profits in order to achieve allocative efficiency. In Section 3 we separately characterize the efficient and equilibrium allocations. Section 4 discusses the relationship between our results and models with no off-market time, while characterizing conditions under which equilibrium allocations are efficient. Section 5 concludes.

2 Model

We first describe the physical environment, starting with quite general formulations of the problems faced by consumers and firms, before turning to formal definitions of both the efficient allocation and the monopolistically competitive equilibrium with free entry.

2.1 Physical environment

In this section we describe the physical environment, specifying the preferences of agents, the technology available to firms, and aggregate resources in the economy. Subsequent sections will characterize efficient allocations and impose a particular market structure.

Consumers. There exists a unit mass of identical consumers who have preferences over a continuum of consumption experiences, indexed by a set $i \in \mathcal{I} \subseteq \mathbb{R}$.

We denote consumption experiences by c_i. A consumption *experience* is defined as the act of combining goods and services which are purchased on the market with off-market time in order to generate an experience which yields final utility. Consumers produce these experiences using a Becker (1965) production function which can take market quantities, q_i, and off-market time, n_i, as inputs. Our *experiences* are analogous to what Aguiar et al. (2012), Aguiar and Hurst (2013), and Aguiar and Hurst (2016) refer to, in different contexts, as "commodities" which generate utility but, which in order to be produced, require market inputs and time. Throughout this analysis we may also refer to these c_i's as "final consumption" that are distinct from q_i's which we sometimes call "market consumption."

Each q_i is a particular good or type of good that consumers buy on the market.[4] Therefore, each $c_i := c(q_i, n_i)$ can be thought of as an output of a particular home production process. We will assume that each consumption experience is produced with a production function that is constant elasticity of substitution (CES) in market quantities and off-market time, so that

$$c_i := c(q_i, n_i) = (\alpha q_i^{\xi} + (1-\alpha)(\zeta n_i)^{\xi})^{1/\xi} \tag{2.1}$$

where $\alpha \in [0, 1]$ represents the weight of market consumption in home production, $\zeta \geq 0$ is proportional to the in-home value of off-market time, and the elasticity of substitution between a particular market purchase and its associated allocation of off-market time is governed by $\xi \leq 1$. Note that this is object is distinct from both the elasticities of substitution across *final consumption* experiences *and* the elasticities of substitution across *market consumption*. The latter is an important object of study in models of monopolistic competition, such as Zhelobodko et al. (2012) and Dhingra and Morrow (2019) who show, to varying degrees, that markups are efficient if and only if the elasticity of substitution for market consumption is constant. Our setup, heretofore, places no restrictions on the constancy of this object. Further, we will show that in our setup it need not be constant in order for monopolistically competitive equilibria to be efficient.

Let u be a strictly positive, strictly increasing, and strictly concave function that takes as its argument consumption experiences. We will assume that $u(c_i) \equiv c_i^{\rho}$ for some parameter ρ governing the substitutability of final experiences satisfying $\max\{0, \xi\} \leq \rho < 1$. Consumer preferences are a linear aggregate of consumption experiences,

$$\mathcal{C} = \int_{\mathcal{I}} u(c_i)\mathrm{d}i. \tag{2.2}$$

[4]In the language of Dhingra and Morrow (2019), each q_i corresponds to a particular "variety." We first index quantities by the dummy variable i, and later index by marginal cost, κ.

Note that when $\alpha = 1$, consumption experiences coincide with market consumption and the specification of preferences in (2.2) is nested within the frameworks of Zhelobodko et al. (2012) and Dhingra and Morrow (2019). In this formulation consumers' love for variety is, more specifically, a love for a variety of consumption *experiences*, not just the goods and services which can be bought and sold on the market. That is, consumers want to engage in a variety of different activities with the goods and services they buy, and for that they have a love of variety.

Consumers desire to choose market purchases and off-market time to maximize $\mathcal{C}$. In this sense time-use and market consumption are non-separable in preferences. This is the key innovation of our paper as it pertains to the literature on monopolistic competition. Bilbiie et al. (2012) and Bilbiie et al. (2019) feature variable labor supply and leisure, but it is separable in preferences from market consumption. In Zhelobodko et al. (2012) and Dhingra and Morrow (2019) time-use is inelastic and does not enter into an agent's utility function.

We denote by $\mathcal{N} := \int_{\mathcal{I}} n_i \, \mathrm{d}i$ the total amount of time spent engaged in the production of experiences. Every consumer is endowed with $\overline{T}$ units of efficiency time. Consumers, generally speaking, may supply two types of labor: 1) start-up effort to firms in order to set them up; 2) variable labor hours in efficiency units of time which are the sole inputs to production. Let $\mathcal{E}$ denote the total amount of start-up effort a consumer supplies in efficiency units of labor, and let $\mathcal{L}$ denote the total amount of variable efficiency hours of labor supplied. The representative consumer's time-use constraint is then

$$\mathcal{E} + \mathcal{L} + \mathcal{N} \leq \overline{T}. \tag{2.3}$$

Firms. Following Dhingra and Morrow (2019) there is a continuum of possible firms which each produce a given variety of goods. Firms are ex-ante identical and must pay a fixed cost f_e in order to exist (these are akin to entry costs) and draw a marginal cost κ from some distribution G. We will consider two separate assumptions/cases on this distribution of productivity. In the *homogeneous firms* case, we assume that G is a point mass at some point $\overline{\kappa} > 0$. In contrast, in the *heterogeneous firms* case, we assume that G has continuous, positive density g defined on an interval $[\underline{\kappa}, \infty)$ for some minimal cost $\underline{\kappa} \in (0, \overline{\kappa})$. Upon drawing their productivity, the firms choose whether or not to pay an additional fixed cost f to remain in business. Both of these fixed costs are denominated in units of effective labor and are interpreted as real resource costs for setting up a firm (i.e., they are not costs imposed by, e.g., regulation). We index firms by κ and refer to this as the firm's "type." The output of a firm of type κ who employs ℓ units of effective labor is then

$$y(\kappa, \ell) = \ell/\kappa \tag{2.4}$$

so that κ is the marginal cost of the firm, and $1/\kappa$ is the firm's labor productivity.

Aggregate Resources. Since labor is the only input to production and all fixed costs are denominated in units of effective labor, and since the consumer is representative, the consumer's individual time-use constraint is also the aggregate resource constraint in this economy. If we denote by M_e the mass of firms that enter and draw a productivity κ, then total effective labor used in production is $\mathcal{L} = \int_{\underline{\kappa}}^{\overline{\kappa}} \ell(\kappa) M_e G(d\kappa)$. Total fixed costs paid by firms are then given by $\mathcal{E} = (f_e + fG(\overline{\kappa}))M_e$. To relate output and costs to this resource constraint we impose the following assumption on the uniqueness of productive outputs in the set of consumption experiences.

Assumption 1. *The experience set, $\mathcal{I}$, is identical to the production set, $[\underline{\kappa}, \overline{\kappa}]$, so that each q_i is associated with one and only one firm output, $q(\kappa) = y(\kappa, \ell)$. This implies that we can index consumer preferences over varieties by $\kappa \in [\underline{\kappa}, \overline{\kappa}]$.*

Assumption 1 states that each firm produces one and only one type of consumption good, with quantity denoted $q(\kappa)$, and consumers supply off-market time, $n(\kappa)$, toward engaging in consumption activities associated with that particular type-κ consumption good. We can then write $c(\kappa) = c(q(\kappa), n(\kappa))$, so that

$$\mathcal{C} = \int_{\underline{\kappa}}^{\overline{\kappa}} u(c(\kappa)) M_e G(d\kappa). \tag{2.5}$$

Since labor inputs to production can be written $\ell(\kappa) = \kappa q(\kappa)$, the economy-wide resource constraint can then be written

$$(f_e + fG(\overline{\kappa}))M_e + \int_{\underline{\kappa}}^{\overline{\kappa}} (\kappa q(\kappa) + n(\kappa)) M_e G(d\kappa) \leq \overline{T}. \tag{2.6}$$

Definition 2.1. *The problem of the planner is to maximize* (2.5) *subject to the resource constraint* (2.6)*.*

Note that because consumers are identical, the planner is constrained only by the aggregate resource constraint (2.6) and can safely ignore the separate time constraints of individual agents. If consumers were *not* identical, the planner would be additionally constrained by the entire set of all consumers' time constraints, since time endowments cannot be redistributed.

This completely describes the physical environment. Note that there are no prices or government transfers in the above, because we have yet to describe the nature of trade between agents.

2.2 Monopolistically competitive equilibrium

We now define the monopolistically competitive equilibrium with free entry. Consumers take prices as given in both product and labor markets, while firms behave in a monopolistically competitive fashion in the product market, in the sense that they take the prices of their rival firms as given but internalize the effect of their own price on consumer demand. Firms will enter if doing so gives them non-negative expected profits, and will operate if doing so provides them with non-negative operating profits.

Consumers. A consumer earns labor income from supplying time toward setting-up firms, $\mathcal{E}$, and variable hours toward production, $\mathcal{L}$. Hourly wages are normalized to unity. Consumers collectively own the firms in the economy, and so in principle may earn profits $\overline{\Pi}$ net of fixed costs. After invoking Assumption 1 the consumer's budget constraint is

$$\int_{\underline{\kappa}}^{\overline{\kappa}} p(\kappa)q(\kappa)M_e G(d\kappa) = \mathcal{E} + \mathcal{L} + \overline{\Pi} \tag{2.7}$$

where the left-hand side is total expenditure and $p(\kappa)$ are prices for market-traded goods $q(\kappa)$. Because the model is expressed in terms of efficiency units of time, each $p(\kappa)$ is technically the price of market-traded goods $q(\kappa)$ *relative* to the value of off-market time, which is equal to the hourly wage and normalized to unity. Combining (2.7) with (2.3) we get the Beckerian budget constraint:

$$\int_{\underline{\kappa}}^{\overline{\kappa}} (p(\kappa)q(\kappa) + n(\kappa))M_e G(d\kappa) = \overline{T} + \overline{\Pi}. \tag{2.8}$$

Definition 2.2. *The problem of a consumer receiving profits $\overline{\Pi}$ and facing a continuum of goods indexed by $[\underline{\kappa}, \overline{\kappa}]$, a mass of firms M_e, and prices p, is to maximize the objective (2.5) subject to the budget constraint (2.8).*

Firms. Each firm chooses their own price taking the prices of all other firms as given. We denote the demand functions implied by Definition 2.2 by $q(\kappa, p; \tilde{p})$, for a firm of type κ who chooses their price p given the set $\tilde{p}$ of all other prices. The problem of a firm who has paid both fixed costs f (operating cost) and f_e (entry cost) is then described in the following definition.

Definition 2.3. *Given the prices $\tilde{p}$ chosen by other firms, the problem of a firm of type κ who has paid both fixed costs is given by*

$$\pi(\kappa; \tilde{p}) := \max_{p \geq 0}(p - \kappa)q(\kappa, p; \tilde{p}), \tag{2.9}$$

so that the quantity $\pi(\kappa; \tilde{p})$ can be interpreted as ex-post operating profits.

Definition 2.3 describes the problem of a firm who has already paid the two possible fixed costs. A firm who has paid the entry cost, f_e, will pay the second, operating cost, f, if and only if $\pi(\kappa) \geq f$. We then have two separate conditions for firm entry, depending upon whether we are in the *homogeneous firms* case, or the *heterogeneous firms* case. In the homogeneous firms case everyone who pays f_e draws the same $\overline{\kappa}$ and chooses to operate, paying f. In the heterogeneous firms case we denote by $\overline{\kappa}$ the solution to $\pi(\overline{\kappa}) = f$, which characterizes the firm that is indifferent between operating and shutting down. Firms will be indifferent to entering (paying f_e) if and only if expected profits net of fixed costs are zero, or

$$\int_{\underline{\kappa}}^{\overline{\kappa}} \pi(\kappa) G(d\kappa) = f_e + f G(\overline{\kappa}).$$

Aggregate profits are

$$\overline{\Pi} := M_e \left(\int_{\underline{\kappa}}^{\overline{\kappa}} \pi(\kappa) G(d\kappa) - f_e - f G(\overline{\kappa}) \right) \tag{2.10}$$

and we have the following definition.

Definition 2.4. *A monopolistically competitive equilibrium consists of a mass of firms M_e, a cutoff value for productivity $\overline{\kappa}$, market quantities and off-market time $(q, n) = (q(\kappa), n(\kappa))_{\kappa \in [\underline{\kappa}, \overline{\kappa}]}$, and prices $p = (p(\kappa))_{\kappa \in [\underline{\kappa}, \overline{\kappa}]}$, such that:*

(i) given the cutoff $\overline{\kappa}$ and prices p, the market quantities q and off-market time n solve the consumer problem in Definition 2.2;

(ii) given prices p, mass of firms M_e, and cutoff $\overline{\kappa}$, for each $\kappa \in [\underline{\kappa}, \overline{\kappa}]$, $p(\kappa)$ solves the firm problem in Definition 2.3;

(iii) a firm of type κ produces if and only if $\pi(\kappa) \geq f$;

(iv) aggregate profits in (2.10) are zero, $\overline{\Pi} = 0$.

Remark 1. With homogeneous firms a non-negligible mass of firms could, in principle, be indifferent between producing and not producing. However, if we write η for the fraction of entering firms that operate, then expected profits are $\overline{\Pi} := M_e(\eta \pi(\overline{\kappa}) - (f_e + f\eta))$. Rearranging this expression for profits and setting to zero then gives $\eta(\pi(\overline{\kappa}) - f) = f_e$, and so the requirement $\pi(\overline{\kappa}) \geq f$ must be strict and hence $\eta = 1$. Intuitively, when productivity is known ex-ante, the decision to operate is redundant given that firms have chosen to enter.

Remark 2. Note that when $\rho = \xi$ preferences attain a special case of Bilbiie et al. (2019) in which consumption exhibits CES, and time use and consumption are additively separable. The $\alpha = 1$ case yields a CES version of Dhingra and Morrow (2019). When production is Leontief, the ratio n/q in that it does not depend on κ in the solution to the planner's problem or $p(\kappa)$ in the equilibrium allocation. Given the constancy of n/q, we can recover a disguised form of the CES environment of Dhingra and Morrow (2019) in which effective costs are not κ but instead $\kappa + 1/\zeta$. The difference between the production costs actually faced by firms, κ, and the cost of consumption experienced by the consumer will be key to our model's implications for the possible efficiency of variable markups.

3 Analysis

We now turn to characterizing both efficient and monopolistically competitive equilibrium allocations, while comparing them. We will show that accounting for off-market time use in a non-separable fashion affects implications pertaining to the efficiency of equilibrium markups and allocations because final consumption is achieved by not merely making a market purchase *but* from making such a purchase *and* allocating time toward *using* that purchase.

As we progress through our analyses, we will eventually distinguish between the markups on market quantities (i.e., q's) that the firms post versus total, what we call "holistic" markups, which comprise the markup of the value of final consumption experiences in equilibrium over the same value under marginal-cost pricing for q's. The former we will call "posted" markups, which to some extent are observable as a result of firm behavior, and the latter we will call "holistic" markups, which are the markups experienced by consumers. The total holistic price of a final consumption experience, from the perspective of the consumer, is thus a function of the posted price (and, therefore, posted markup) of the market good, the value of the consumer's off-market time, and the degree to which time and market purchases are complementary or substitutable within the act of *producing* in-home final consumption.

Consider now the implications of our model parameterization, where the elasticity of substitution across *experiences* is constant. In Dhingra and Morrow (2019), where firms are heterogeneous in costs/productivities and consumer time use is inelastic, CES preferences over *market purchases* are both sufficient and necessary for decentralized allocations from a monopolistically competitive equilibrium to be socially optimal. In our environment the elasticity of substitution across final consumption *experiences* is constant, but the prices of final experiences are only equivalent to the prices of market goods when consumers supply time inelastically (i.e., $\alpha = 1$). Thus, we seek to explore the effect of markups on the efficiency of allocations outside of this particular asymptotic case.

3.1 Efficient allocations

The problem of the planner is

$$\begin{aligned} &\max_{q,n,M_e,\overline{\kappa}} \int_{\underline{\kappa}}^{\overline{\kappa}} (\alpha q(\kappa)^{\xi} + (1-\alpha)(\zeta n(\kappa))^{\xi})^{\rho/\xi} M_e G(d\kappa) \\ &(f_e + fG(\overline{\kappa}))M_e + \int_{\underline{\kappa}}^{\overline{\kappa}} (\kappa q(\kappa) + n(\kappa))M_e G(d\kappa) \leq \overline{T}. \end{aligned} \tag{3.1}$$

where the objective function in (3.1) composes (2.1) with $u(x) = x^{\rho}$ for $\max\{0, \xi\} \leq \rho < 1$.

We first solve for off-market time in the above problem in order to reduce the problem to one of choosing consumption experiences, $c(\kappa)$, only. The reason for doing this is to show that the efficient price of final consumption embeds *both* the firm's marginal cost of production κ, as well as the degree to which market consumption and off-market time are complementary, ξ, and the intensity of final consumption in market goods, α, from the perspective of the consumer. The first-order conditions for n and q combine to give the efficient ratio of off-market time to market goods

$$n(\kappa)/q(\kappa) = \zeta^{\frac{\xi}{1-\xi}} [(1/\alpha - 1)\kappa]^{\frac{1}{1-\xi}}, \tag{3.2}$$

which satisfies the following comparative statics.

Lemma 3.1. *The efficient ratio of off-market time use to market consumption is concave in κ for $\xi \in (-\infty, 0)$, convex in κ for $\xi \in (0, \rho]$, and is strictly increasing in κ for all $-\infty < \xi \leq \rho$. When home production is Leontief, the ratio is $1/\zeta$ and is thus constant.*

Proof. See Appendix B. $\square$

Lemma 3.1 discusses how $n(\kappa)/q(\kappa)$ varies in κ, which will be informative for many of our later analyses. For high-cost/low-productivity varieties (high κ) consumers supply relatively more off-market time and engage in relatively less off-market consumption than for low-cost/high-productivity varieties. However, the rate at which the relative provision of time toward home production changes depends on whether n and q are complements or substitutes. When time and market consumption are complementary, the allocative difference between similar low-cost/high-productivity firms is greater than the same difference between high-cost/low-productivity firms. The opposite is true when time and market consumption are substitutes. In this case, as marginal costs rise consumers substitute time for market consumption at increasing rates. The result in Lemma 3.1 will be important for understanding how the total, time-inclusive cost of consumption experiences varies.

Using the expression in (3.2) for the efficient ratio of off-market time to market consumption, we can write

$$\kappa q(\kappa) + n(\kappa) = \frac{\kappa}{\alpha}\Big(\alpha + (1-\alpha)[(1/\alpha - 1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\Big)^{1-1/\xi} c(\kappa) \tag{3.3}$$

and so it is instructive to define a new function

$$\psi(\kappa) = \psi(\kappa;\alpha,\xi) := \frac{\kappa}{\alpha}\Big(\alpha + (1-\alpha)[(1/\alpha - 1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\Big)^{1-1/\xi} \tag{3.4}$$

that represents the total per unit resource cost of a consumption experience inclusive of off-market time along the upper envelope of the consumer's choice set.[5] The expression (3.4) is valid for $\xi > -\infty$, but when home production is Leontief, it can be shown that $\psi(\kappa) = \kappa + 1/\zeta$ since q is linear in n.

Using the envelope theorem, it is straightforward to show that the efficient ratio of market consumption, $q(\kappa)$, to total consumption experiences, $c(\kappa)$, is

$$q(\kappa)/c(\kappa) = \psi'(\kappa;\alpha,\xi). \tag{3.5}$$

The object in (3.5) is simply the efficient marginal cost of *using* a market product, $q(\kappa)$, to make an experience, $c(\kappa)$. Similarly, its inverse is a productivity term, characterizing the efficient in-home production level associated with utilization of the type-κ market product. Notice that for all κ, α, and ξ we have $\psi'(\kappa;1,\xi) \equiv \psi'(\kappa;\alpha,-\infty) \equiv 1$, so that market consumption and final consumption coincide both when off-market time and consumption are perfect complements and when the share of off-market time in consumption experiences vanishes. Further, for all κ and ξ we have $\psi(\kappa;0,\xi) = 1/\zeta$ and $\psi(\kappa;1,\xi) = \kappa$. Thus, when consumers only care about off-market time, the efficient value of one final consumption experience is just the inverse of the consumer's off-market, time-use productivity ζ, while when time is inelastic this value is equal to the marginal cost of production. Finally, in the Cobb-Douglas case we have $\psi(\kappa) = \alpha^{-\alpha}[(1-\alpha)\zeta]^{-(1-\alpha)}\kappa^{\alpha}$.[6]

Lemma 3.2. *The total cost function, $\psi(\kappa)$, is strictly increasing and concave in*

[5] The function $\psi(\kappa)$ is analogous to the same object in Becker (1965).

[6] Special cases are proved in Lemma A.2.

κ for all ξ, and the elasticity with respect to κ is

$$\epsilon(\kappa;\psi) = \frac{\kappa\psi'(\kappa)}{\psi(\kappa)} = \frac{1}{1 + (1/\alpha - 1)^{\frac{1}{1-\xi}} [\kappa\zeta]^{\frac{\xi}{1-\xi}}}. \tag{3.6}$$

This elasticity is bounded above by unity, increasing in κ when $\xi \in [-\infty, 0)$, decreasing in κ when $\xi \in (0, \rho]$, and constant and equal to α when $\xi = 0$.

Proof. See Appendix B. □

Equation (3.6) in Lemma 3.2 defines an elasticity, unique to this model, which describes how the total cost function, $\psi(\kappa)$, varies in κ. This elasticity will be important in order for us to understand how firms' markups over marginal cost are actually *experienced* (i.e., passed through) to the consumer. In models without consumption time, markups affect the cost associated with final consumption one-for-one. In this model, the shape of $\psi(\kappa)$ will determine the degree to which the imperfect substitution between market purchases and time use affects how consumers experience markups. The elasticity in (3.6) will thus help us characterize the *effective* efficient markup of final consumption experiences, which can then be compared to the *posted* markups over the marginal cost of production.

Note that ψ is always inelastic in κ.

Further, ψ is monotonically increasing (in levels), and so low-productivity (i.e., high κ) firms are also associated with a high final resource cost of *experienced* consumption. Since both $n(\kappa)$ and $q(\kappa)$ are "normal" goods, from the consumer's perspective, higher κ means higher input costs overall, given the resource cost of $n(\kappa)$ is fixed. But, the rate at which $\psi(\kappa)$ increases falls as κ rises, regardless of the sign of ξ. The final-cost differential for experiences associated with market outputs between two high-productivity (low κ) firms is greater than the final-cost differential for two experiences associated with market outputs from low-productivity firms (high κ). However, given the result from Lemma 3.1, the reasons driving this second-order phenomenon still pertain to the sign of ξ, which governs the rate at which time is allocated to different market purchases.

To illustrate, we consider two comparative statics: 1) the difference in final cost, $\psi(\kappa)$, between two, similar low-cost/high-productivity firms; 2) the difference in final cost, $\psi(\kappa)$, between two, similar high-cost/low-productivity firms. Let $d\kappa$ be the marginal cost differential between the two different pair-wise sets of firms from cases (1) and (2), and suppose this marginal cost differential is the same across both cases. In case (2) the final cost differential, $d\psi$, is less than that in case (1) due to the concavity of $\psi(\kappa)$. This is true regardless of the sign of ξ.

Now, returning to Lemma 3.1, note that when $\xi < 0$, the rate at which consumers substitute from market consumption towards off-market time slows as the marginal cost of producing a variety rises. Thus, when $\xi < 0$, differences in $d\psi$ at the high-cost/low-productivity end of the distribution of κ are diminished precisely *because* of consumption/time-use complementarities. Note that the power term in (3.4) is such that $1 - 1/\xi \to_+ 1$ as $\xi \to -\infty$, so that if $n(\kappa)/q(\kappa)$ were convex with $\xi < 0$ then we would expect $\psi(\kappa)$ to actually rise faster as κ rises. Since this is not the case we can conclude that the complementarities causing substitution from $q(\kappa)$ to $n(\kappa)$ to slow as κ rises are what primarily contribute to the concavity of $\psi(\kappa)$ when $\xi < 0$. Complementarities in home production thus imply that consumers derive utility from final consumption by substituting time away from low-κ final experiences at lower rates as the cost of production rises.

When $\xi \in (0, \rho]$, so that time use and market consumption are substitutes, Lemma 3.1 shows that $n(\kappa)/q(\kappa)$ is increasing and convex in κ. Consequently, substitution away from market consumption and toward off-market time accelerates as the cost of producing varieties rises. Because $1 - 1/\xi < 0$ when $\xi \in (0, \rho]$, this leads to smaller final cost differentials at the low end of the marginal cost distribution relative to differences at the high end of the marginal cost distribution.

We can now use the above to write both the objective and the constraint in the planner's problem described in (3.1) in terms of consumption experiences, $c(\kappa)$, rather than market consumption, $q(\kappa)$, and off-market time, $n(\kappa)$:

$$\begin{gathered} \max_{c, M_e, \overline{\kappa}} \int_{\underline{\kappa}}^{\overline{\kappa}} c(\kappa)^{\rho} M_e G(d\kappa) \\ (f_e + fG(\overline{\kappa}))M_e + \int_{\underline{\kappa}}^{\overline{\kappa}} \psi(\kappa) c(\kappa) M_e G(d\kappa) \leq \overline{T}. \end{gathered} \tag{3.7}$$

Writing the planner's problem like (3.7) shows that the efficient allocations coincide with a special case of Dhingra and Morrow (2019), in which preferences are CES over final consumption and the cost is not κ but instead $\psi(\kappa)$. To anticipate our later analyses, much of the interest (and difficulty) in the characterization of equilibrium markups and welfare arises from the fact that despite the isomorphism between the problem in (3.7) and planning problems without off-market time, the firms in our decentralized environment sell *market* consumption (not consumption *experiences*) and face only the *technological* costs κ of producing the variety (and not the *total* cost $\psi(\kappa)$). Proposition 3.3 characterizes efficient allocations of final consumption experiences $c(\kappa)$ and, by extension, market consumption $q(\kappa)$, the mass of firms M_e, and the productivity cutoff $\overline{\kappa}$. We characterize allocations for both the case with heterogeneous firms and the case with homogeneous firms.

Proposition 3.3 (Efficient Allocations)**.** *In the heterogeneous firms case there exists a unique solution to the equation*

$$f \times \int_{\underline{\kappa}}^{\overline{\kappa}} \psi(\kappa)^{-\frac{\rho}{1-\rho}} G(d\kappa) = (f_e + fG(\overline{\kappa})) \times \psi(\overline{\kappa})^{-\frac{\rho}{1-\rho}}, \tag{3.8}$$

which gives the cutoff productivity in the solution to the planner's problem. For this value of $\overline{\kappa}$, the efficient mass M_e of entering firms is given by

$$M_e = \frac{(1-\rho)\overline{T}}{f_e + fG(\overline{\kappa})}. \tag{3.9}$$

The efficient quantity of final consumption experiences of every variety is

$$c(\kappa) = \frac{\rho \overline{T} \psi(\kappa)^{-\frac{1}{1-\rho}}}{M_e \int_{\underline{\kappa}}^{\overline{\kappa}} \psi(\kappa)^{-\frac{\rho}{1-\rho}} G(d\kappa)} \tag{3.10}$$

while the efficient market consumption of each variety is $q(\kappa) = \psi'(\kappa)c(\kappa)$. In the homogeneous firms case, the mass of entering firms is

$$M_e = \frac{(1-\rho)\overline{T}}{f_e + f}, \tag{3.11}$$

each consumption experience is $\overline{c} = (f_e + f)/((1/\rho - 1)\psi(\overline{\kappa}))$, and market consumption of each variety is $\overline{q} = \psi'(\overline{\kappa})\overline{c}$.

Proof. See Appendix B. □

Some discussion around Proposition 3.3 is warranted. First, note that in equation (3.8), the integral $\int_{\underline{\kappa}}^{\overline{\kappa}} \psi(\kappa)^{-\frac{\rho}{1-\rho}} G(d\kappa)$ may be interpreted as an index that describes the aggregate, total resource cost of all consumption experiences. The left-hand side of (3.8) is then the total operating cost in units of final consumption experiences, while the right-hand side describes the cost of entry plus operation for the last firm to enter, denominated in units of final consumption experiences. Firms ought to produce $q(\kappa) > 0$ only if the final consumption value of their fixed costs is less than or equal to the total consumption value of all production, as is apparent by inspecting the left-hand side of (3.8). The presence of the total cost functions demonstrates how consumer behavior in home production impacts the efficient entry margin and firm profitability: $\psi(\overline{\kappa})$ needs to be low enough relative to economy-wide costs in order for entry to be profitable. Equation (3.9) also shows that the optimal fraction of the endowment of labor $\overline{T}$ devoted to setup costs is always $1 - \rho$, regardless of the distribution G, costs f and f_e, or the preference parameters ξ and α.

Equation (3.10) characterizes the level of final consumption experiences for each variety. Final consumption is proportional to $\psi(\kappa)^{-\frac{1}{1-\rho}}$, while market consumption differs from this by the factor $\psi'(\kappa)$, the marginal cost of home production. When $\alpha = 1$ we obtain the familiar constant elasticity of demand with respect to κ, as $c \equiv q \propto \kappa^{-\frac{1}{1-\rho}}$.

3.2 Monopolistically competitive allocations

We now turn to the characterization of monopolistically competitive equilibria. We will proceed in a manner similar to that adopted in Section 3.1 in order to highlight the similarities and differences between the efficient and equilibrium allocations. To this end we first characterize the off-market time chosen by the consumer in order to reduce the problem to one in which consumption experiences are the sole object of choice. From here we will then turn to an analysis of the pricing and entry decisions of the firms.

Faced with a continuum of firms distributed over the interval $[\underline{\kappa}, \overline{\kappa}]$ according to the CDF G, with pricing schedule p, the consumer problem is to choose market consumption and off-market time $(q, n) = (q(\kappa), n(\kappa))_{k\in[\underline{\kappa},\overline{\kappa}]}$ satisfying

$$V(p) = \max_{q,n} \int_{\underline{\kappa}}^{\overline{\kappa}} \left(\alpha q(\kappa)^{\xi} + (1-\alpha)(\zeta n(\kappa))^{\xi}\right)^{\rho/\xi} M_e G(d\kappa)$$
$$\int_{\underline{\kappa}}^{\overline{\kappa}} (p(\kappa)q(\kappa) + n(\kappa)) M_e G(d\kappa) = \overline{T} + \overline{\Pi}.$$

Taking first-order conditions for consumption and time and rearranging provides us with an analog of equation (3.2) for the decentralized economy,

$$n(\kappa)/q(\kappa) = \zeta^{\frac{\xi}{1-\xi}} [(1/\alpha - 1)p(\kappa)]^{\frac{1}{1-\xi}}. \tag{3.12}$$

Substituting the ratio (3.12) into the consumer problem gives

$$\begin{aligned} V(p) = \max_{q} \int_{\underline{\kappa}}^{\overline{\kappa}} c(\kappa)^{\rho} M_e G(d\kappa) \\ \int_{\underline{\kappa}}^{\overline{\kappa}} \psi(p(\kappa))c(\kappa) M_e G(d\kappa) = \overline{T} + \overline{\Pi}. \end{aligned} \tag{3.13}$$

The consumer problem in (3.13) is similar to the planner's problem in (3.7) except for a few key differences. First, the consumer obviously does not choose the mass of entering firms M_e or the cutoff $\overline{\kappa}$. Second, the resources devoted to setting up firms appear nowhere in the consumer's problem. However, in spite of these obvious differences, the function ψ capturing the role of off-market time in affecting quality and costs appears in both the centralized and decentralized environments. In the planner's problem the function ψ is evaluated at the technological cost κ, while in the equilibrium allocation it is evaluated at the (yet-to-be-determined) price p chosen by the firm.

The degree to which posted markups pass through to consumers depends on the structure of ψ. Firms, therefore, must take a consumer's known value of their off-market time, ζ, as given and then choose a markup that maximizes their profits subject to the knowledge of this ζ. But as long as ψ has curvature (i.e., off-market time and market consumption are neither perfect complements nor perfect substitutes in home production), the holistic markup experienced by the consumer will be different from the posted markup the firm sets. Home production thus augments effective markups, which may have both implications for whether or not they are economically efficient *and* theoretical implications for the relationship between efficiency, markups, and the structure of demand (i.e., constancy versus variability of the elasticity of substitution for market consumption).

After solving for the solution to the consumer's problem, $q(p)$, the firm's problem in Definition 2.3 may now be written

$$\pi(\kappa) = \max_{p\geq 0} \ (p-\kappa)q(p) =: \frac{(\overline{T}/M_e)}{\int_{\underline{\kappa}}^{\overline{\kappa}} \psi(p(\kappa))^{-\frac{\rho}{1-\rho}} G(d\kappa)} \times \widehat{\pi}(\kappa) \tag{3.14}$$

where the second equality defines $\widehat{\pi}(\kappa)$, which may be viewed as the profits of the firm up to a constant that is independent of the firm's choices. Now note that for any demand schedule q faced by the firm, the first-order condition of the firm's problem may be rearranged to obtain

$$p/\kappa = \frac{\epsilon(p;q)}{\epsilon(p;q)-1} \tag{3.15}$$

where

$$\epsilon(p;q) := -\frac{pq'(p)}{q(p)} \tag{3.16}$$

is the own-price elasticity of demand. Outside of special cases considered below, the firm's problem does not have a closed-form solution precisely because this price elasticity is, in general, not constant in market prices.

Proposition 3.4 characterizes the equilibrium markups under different parametric conditions. We identify three cases in which markups have a closed-form solution, and two cases in which markups are constant. Note that constancy of markups does not necessarily imply efficiency, because the posted markups are *not* the holistic markups which are actually internalized by the consumer. We will discuss this more below.

Proposition 3.4 (Markups)**.** *If $\xi \leq \rho$ then the price $p(\kappa)$ chosen by the firm is unique for all κ and increasing in κ. Further, the markup $m(\kappa) := p(\kappa)/\kappa$ always weakly exceeds $1/\rho$, and in the special cases in which $\xi \in \{-\infty, 0, \rho\}$, satisfies:*

(i) (Leontief) For $\xi = -\infty$ we have $m(\kappa) = 1/\rho + (1/\rho - 1)/[\zeta\kappa]$.

(ii) (Cobb-Douglas) For $\xi = 0$, we have $m(\kappa) = 1/\rho + (1/\rho - 1)(1/\alpha - 1)$.

(iii) (CES) For $\xi = \rho$, we have $m(\kappa) = 1/\rho$.

Proof. See Appendix C. □

Markups are thus only constant when time use and market consumption are unit elastic in terms of their input substitutability (i.e., Cobb-Douglas home production) or preferences take a Bilbiie et al. (2019) flavor with elastic leisure that is separable from market consumption (i.e., $\xi = \rho$). In the latter case we get the constant markup that we would expect under the Bilbiie et al. (2019) model with CES preferences but with heterogeneous firms as in Dhingra and Morrow (2019). To our knowledge, however, whether or not constant markups from CES preferences are efficient in an environment with endogenous labor and heterogeneous firms has yet to be explored. We consider their potential efficiency in the next section.

The envelope theorem implies that $\widehat{\pi}'(\kappa) = -q(p(\kappa))$, and so the elasticity of profits with respect to productivity may be written as a function of the markup,

$$\frac{\kappa\pi'(\kappa)}{\pi(\kappa)} = \frac{\kappa\widehat{\pi}'(\kappa)}{\widehat{\pi}(\kappa)} = -\frac{1}{m(\kappa) - 1}. \tag{3.17}$$

The price of market consumption does not admit a closed-form solution outside of the special cases stated in Proposition 3.4. Nonetheless, we can qualitatively describe how markups vary in marginal costs under different values of ξ.

Lemma 3.5. *The markup is weakly decreasing in κ when $\xi \in [-\infty, 0)$ and weakly increasing in κ when $\xi \in (0, \rho]$.*

Proof. See Appendix C. □

Lemma 3.5 states that markups decline in per-unit marginal costs when time use and market consumption are complementary, and they rise when time use and market consumption are substitutes. We now provide some intuition for the variation of markups in productivity, in the spirit of the discussion following Lemma 3.1 (variation of efficient n/q) and Lemma 3.2 (variation of efficient ψ).

First, consider the case when $\xi < 0$, so that time use and market consumption are complements. Lemma 3.5 tells us that differences in final experienced (effective) prices, $\psi(p(\kappa))$, at the high-end of the marginal-cost distribution are small precisely because of markup compression as costs rise and productivities fall. This is a bit counterintuitive, however. Consider, now, the result in Lemma 3.1: substitution away from market consumption and toward time use slows as κ rises when $\xi < 0$. This result would seem to suggest that firms at the high-end of the cost distribution would then have relatively higher pricing power because consumers always need some non-insignificant amount of market consumption to achieve production of final experiences, given the complementarities in $c(q, n)$. However, final experiences themselves are substitutes because $\rho \in (0, 1)$, so that rising $\psi(p(\kappa))$ in κ forces consumers to substitute away from final consumption of high ψ experiences with less productive inputs overall (i.e., lower levels of n and q).[7] This channel dominates, forcing low-productivity firms to markup at lower rates.

The opposite occurs when $\xi \in (0, \rho]$, so that time use and market consumption are substitutes. In this case low-productivity firms have the highest markups. Intuitively, these firms charge a higher markup in order to remain profitable precisely because consumers substitute time for their market consumption at increasing rates as costs rise. Nonetheless, the final costs faced by the consumer, $\psi(p(\kappa))$, do not accelerate as κ rises, because consumers deploy lower-cost inputs (i.e., their off-market time) toward production of those final experiences in place of heavily-marked-up market consumption.

The above analysis has characterized the behavior of firms in partial equilibrium. In order to characterize the monopolistically competitive equilibrium, we now impose two further conditions: 1) the marginal firm with productivity $\overline{\kappa}$ is indifferent between producing and not producing; and 2) firms make zero profits (net of entry costs) in expectation. The following Proposition 3.6 is the analogue of Proposition 3.3, insofar as it describes the equilibrium mass of entrants, the productivity cutoff level $\overline{\kappa}$, which Dhingra and Morrow (2019) refer to as productivity "selection," and final consumption.

Proposition 3.6 (Equilibrium Allocations)**.** *In the heterogeneous firms case there is a unique monopolistically competitive equilibrium, with cutoff productivity* $\overline{\kappa}$

[7] Note that this is because $p(\kappa)$ is increasing in κ despite the fact that markups *over* marginal costs may be either increasing or decreasing, depending on ξ.

satisfying

$$f \times \int_{\underline{\kappa}}^{\overline{\kappa}} \widehat{\pi}(\kappa) G(d\kappa) = (f_e + f G(\overline{\kappa})) \times \widehat{\pi}(\overline{\kappa}) \tag{3.18}$$

where $\widehat{\pi}$ is defined in equation (3.14). Given $\overline{\kappa}$ the mass of firms M_e in the environment with heterogeneity is

$$M_e = \frac{\overline{T}\, \widehat{\pi}(\overline{\kappa})}{f \int_{\underline{\kappa}}^{\overline{\kappa}} \psi(p(\kappa))^{-\frac{\rho}{1-\rho}} G(d\kappa)}. \tag{3.19}$$

Note that $\overline{\Pi} = 0$ in equilibrium. The equilibrium quantity of final consumption experiences of every variety is

$$c(p(\kappa)) = \frac{\overline{T} \psi(p(\kappa))^{-\frac{1}{1-\rho}}}{M_e \int_{\underline{\kappa}}^{\overline{\kappa}} \psi(p(\kappa))^{-\frac{\rho}{1-\rho}} G(d\kappa)}, \tag{3.20}$$

and $q(p(\kappa)) = \psi'(p(\kappa))c(p(\kappa))$ for market consumption. When firms are homogeneous and possess the same productivity level $\overline{\kappa}$ and thus the same price $\overline{p} = p(\overline{\kappa})$, the mass of firms is

$$M_e = \frac{(1 - \overline{\kappa}/\overline{p})\, \epsilon(\overline{p}; \psi)\, \overline{T}}{f_e + f}, \tag{3.21}$$

each consumption experience is $\overline{c} = (f_e + f)/((1 - \overline{\kappa}/\overline{p})\epsilon(\overline{p}; \psi)\psi(\overline{p}))$, and market consumption of each variety is $\overline{q} = \psi'(\overline{\kappa})\overline{c}$.

Proof. See Appendix C. □

We will now discuss how the selection and competition conditions characterizing general equilibrium outcomes in a competitive environment differ from efficient allocations. In the next section we will dive into the details as to exactly *how* (via sign) and *why* equilibrium allocations may be inefficient. Our points of comparison between Proposition 3.6 and Proposition 3.3 are the pairs of equations, (3.8) and (3.18), (3.9) and (3.19), and (3.10) and (3.20). Recall from previous discussion that $\widehat{\pi}(\kappa)$ represents operating profits in units of final consumption experiences (not efficiency units of time). Similar to (3.8) which equates total operating *costs* of final consumption associated with type-κ market goods with that of the last entering firm, (3.18) equates the total operating *profits* of all production with the operating profits of the last firm to enter. When $\psi(p(\kappa))^{-\frac{\rho}{1-\rho}}/\widehat{\pi}(\kappa)$ is constant in κ, efficient selection is achieved in a competitive environment. Comparing the concentrating conditions, (3.9) and (3.19), note that a constant fraction of time is devoted to aggregate production in an efficient environment, but in the competitive environment the competition level will depend on the degree of selection via the function $\widehat{\pi}(\overline{\kappa})$.

Briefly, we would like to comment on the equilibrium mass of firms in the homogeneous firms case, as given by (3.21). Note that when all firms are homogeneous the zero aggregate profit condition dictates that fixed costs for entry plus fixed costs for operation must equate with operating profits. For this reason both f_e and f enter into the expression in (3.21), as they do in the efficient allocation. However, whereas the elasticity of substitution directly determines the mass in the efficient allocation, it is apparent, both with heterogeneous and homogeneous firms, that markups will determine the market-clearing mass. This is somewhat trivial when firms are homogeneous since the model yields a single, economy-wide markup, whereas when firms are heterogeneous, markups for the last entrant (embedded in $\widehat{\pi}(\overline{\kappa})$) determine the equilibrium mass. It should then be clear by comparing Propositions 3.6 and 3.3 that as long as markups are *not* equal to $1/\rho$, the mass of firms (and thus degree of competition) in a competitive equilibrium will not necessarily be efficient.

Finally, note that, depending on a specification for $G(\cdot)$, the level of final consumption experiences at each variety may be higher or lower than the efficient allocation. This is because $\overline{\kappa}$ may be too low or too high in equilibium relative to the socially optimal outcome. We thus will not sign how the individual allocations of $c(p(\kappa))$ differ relative to their efficient counterparts. Still though, we will eventually show that total consumption utility, $\mathcal{C}$, is equivalent to the efficient level of utility in a particular case ($\xi = -\infty$) when markups are also variable.

4 Welfare analysis and discussion

Section 3.1 characterizes efficient allocations and Section 3.2 characterizes the allocations that are attained in monopolistically competitive equilibria. We now wish to compare and contrast these two. Are there "too many firms" operating in equilibrium? Is market concentration "too high" in equilibrium? Under what conditions on the structure of competition and home production are allocations first-best from a welfare standpoint?

It will become apparent that the fulcrum driving our results is the parameter ξ. As has already been seen in Proposition 3.4 and Lemma 3.5, this elasticity dictates the conditions under which markups are constant and how they vary in firm costs. It is also integral in determining whether selection and concentration are efficient, whether the elasticity of substitution for market purchases is constant or variable, and whether the economy itself achieves first-best outcomes. Table 1 summarizes our findings regarding efficiency and constancy of markups and elasticities in the

Table 1: Efficiency and Constancy

		Special Cases for ξ		
Object	Description	$-\infty$	0	ρ
$\overline{\kappa}$	Selection	Yes	Yes	No
M_e	Concentration	Yes	No*	No*
$\mathcal{C}$	Welfare	Yes	No	No
Economy First-best?		Yes	No	No
Markups Constant?		No	Yes	Yes
CES in q's?		Yes	Yes	Yes

* We can only explicitly show this when firms are homogeneous.

special cases we have heretofore considered. In the pages that follow we will first compare efficient allocations with equilibrium allocations before moving on to a culminating discussion in Section 4.1 which relates allocative efficiency back to elasticities of substitution. Our goal in Section 4.1 will be to compare our results to the notion, widely held in the literature, that CES preferences are both necessary and sufficient to ensure that both markups are constant and allocations are efficient.

Preliminary to our welfare discussion, we must first introduce a useful mathematical object. The equilibrium cutoff value of productivity is given by (3.18), while the efficient cutoff value of productivity is given by (3.8). Defining

$$J(\overline{\kappa}; F) := \int_{\underline{\kappa}}^{\overline{\kappa}} (F(\kappa)/F(\overline{\kappa}))G(d\kappa) \tag{4.1}$$

for an arbitrary function F, note that the efficient cutoff is characterized by the solution to $J(\overline{\kappa}; F) = f_e/f + G(\overline{\kappa})$ for $F(\kappa) := \psi(\kappa)^{-\frac{\rho}{1-\rho}}$, and the equilibrium cutoff is a root of the same equation for $F(\kappa) := \widehat{\pi}(\kappa)$. The equilibrium cutoff productivity will therefore be inefficiently high if $J(\kappa; \pi) \leq J(\kappa; \psi^{-\frac{\rho}{1-\rho}})$ for all κ.

Proposition 4.1 (Selection)**.** *The equilibrium cutoff $\overline{\kappa}$ is inefficiently high if $\xi \in (-\infty, 0)$, inefficiently low if $\xi \in (0, \rho]$, and efficient if $\xi \in \{-\infty, 0\}$.*

Proof. See Appendix D. □

Proposition 4.1 is our first indication that when time use is elastically supplied and (sometimes) non-separable with market consumption, the link between markup constancy and efficiency is broken. Further, Proposition 4.1 together with Proposition 3.4 say that constancy of markups is *neither* necessary nor sufficient for efficient selection and, thus, first-best outcomes to be achieved. This is because when home production is Cobb-Douglas, markups are constant and selection is efficient, but markups are variable when home production is Leontief and selection is still efficient. Further, when $\xi = \rho$ time use and market consumption are strongly separable in preferences, and yet selection is still inefficiently low even though markups are constant. In this latter case consumers supply excess time toward home production, so that high-cost/low-productivity firms cannot profitably enter and less varieties are produced. Thus, consumers' preferences for engaging in off-market activities ensure they have less options (in terms of market goods) with which to *engage* in such activities.

We have thus far discussed how our model breaks the link between constant *posted* markups and efficiency. But the model generates two different kinds of markups: posted markups, chosen by the firm, and holistic markups, which reflect the total value of consumption over marginal-cost pricing that the consumer experiences when choosing to allocate time in order to *use* market purchases. Definition 4.2 defines these holistic markups.

Definition 4.2. *The holistic markup is* $\varphi(\kappa) := \psi(p(\kappa))/\psi(\kappa)$.

Definition 4.2 describes how much the final cost of consumption experiences, from the consumer's perspective, is marked up over an economic environment where firms charge marginal cost for all inputs. The posted markup gives the familiar price over the cost of market consumption, while the holistic markup is the price of final consumption experiences given market prices relative to the technological cost of final consumption.

Note that the object $\psi(\kappa)$ is *not* a marginal cost *per-se* from the perspective of an in-home producer: in a competitive equilibrium the marginal cost of home production is $1/\psi'(p)$ because consumers must take market prices of inputs, q, as given. Rather, $\psi(\kappa)$ is the cost a consumer would theoretically incur if firms set posted prices equal to their marginal costs of production.

Lemma 4.3. *If the equilibrium cutoff* $\overline{\kappa}$ *is efficient then holistic markups* $\varphi(\kappa)$ *are constant.*

Proof. See Appendix D. □

Lemma 4.3 establishes that constancy of holistic markups is necessary for selection to be efficient, despite the fact that posted markups (i.e., what the firms actually choose) are non-constant in the Leontief case and constant but different from $1/\rho$ in the Cobb-Douglas case. This is because when $\xi \in \{-\infty, 0\}$ preferences in q net of time use are CES, but firms still must internalize either the consumer's off-market productivity, ζ (in the Leontief case) and/or the consumer's time-intensity of home production, α (in the Cobb-Douglas case). What is further notable is that constancy of posted markups is *not* sufficient for either efficient selection or constant holistic markups, because the $\xi = \rho$ case provides a nice counterexample to such a claim.

Comparing holistic versus posted markups has implications for how we understand whether monopolistic competition is indeed efficient. One question which inspired our exploration of a model with consumption time is as follows: if we account both for consumers' market-purchasing decisions and their associated time-allocation decisions toward utilizing their market purchases, how are our inferences regarding markups affected? That is, we can hypothetically measure firms' posted markups in data, but these may not be the actual or holistic markup after accounting for the time-use trade-offs consumers face when making market purchase decisions. When accounting for these trade-offs, as we do in this paper, how do holistic markups experienced by consumers compare to firms' posted markups? Lemma 4.4 provides an answer, showing that the markups we measure based on firm behavior are *not* the same as the markups over hypothetical, marginal-cost pricing that consumers experience.

Lemma 4.4. *Posted markups always weakly exceed holistic markups and coincide if and only if* $\alpha = 1$.

Proof. See Appendix D. □

We now turn to a discussion of the mass of firms. Proposition 4.5 shows that in the case of homogeneous firms, we can unambiguously sign the difference between the equilibrium and efficient number of firms. Intuitively, from the point of view of the planner, it is as if the consumers are devoting "too much" time to off-market time, to the detriment of profits and the incentives for firm entry.

Proposition 4.5 (Concentration with Homogeneous Firms)**.** *In the homogeneous firms case the equilibrium mass of firms is lower than the efficient mass of firms unless off-market time and market consumption are perfect complements* ($\xi = -\infty$).

Proof. See Appendix D. □

Generally speaking, for heterogeneous firms, comparing the efficient mass to the equilibrium mass is not so straightforward. Even without off-market time use Dhingra and Morrow (2019) acknowledge that knowing the cost distribution function $G(\kappa)$ is required in order to sign the difference between efficient and equilibrium M_e. Nocco et al. (2013) (cited in Dhingra and Morrow (2019)) show that even if demand is linear and the cost distribution well-specified (Pareto), the difference cannot be signed. In our model, though, there is one particular case wherein it is straightforward to compare efficient and equilibrium concentration levels. Proposition 4.6 characterizes this.

Proposition 4.6 (Concentration with Heterogeneous Firms). *When firms are heterogeneous in κ and off-market time and market consumption are perfect complements, the equilibrium mass of firms is efficient.*

Proof. See Appendix D. □

The results in Propositions 4.5 and 4.6 follow directly from the fact that selection is efficient with Leontief home production. The fact that final costs are just shifted by a constant, $1/\zeta$, is crucial to understanding Propositions 4.5 and 4.6. In this case only, efficient final costs are just $\kappa + 1/\zeta$ and holistic markups are just $1/\rho$, which is equivalent to the constant markup from Dhingra and Morrow (2019) in an environment with inelastic time use.

It is rather intuitive why efficiency attains in the Leontief case when q/n is exogenous. The basic insight is that this case is essentially a disguised version of the CES case considered in Dhingra and Morrow (2019), in which the technological cost is not κ but $\kappa + 1/\zeta$. To understand how this intuition is borne out of our model, note that under Leontief home production we simply add and subtract $1/\zeta$ from the profit margin of the firm to obtain

$$\widehat{\pi}(\kappa) = \max_{\psi(p)\geq 0} \left(\psi(p) - \psi(\kappa)\right)\psi(p)^{-\frac{1}{1-\rho}}$$

which shows that the firm charges a constant markup of the total price over the total cost. The results of Dhingra and Morrow (2019) pertaining to efficiency of markups, selection, and concentration, then all follow immediately. This simple observation is particularly noteworthy because the markups in this case are *not* constant across firms. This is apparent by inspecting condition (i) of Proposition 3.4, which depends on κ.

By now it is obvious that the conventional wisdom from the literature with regards to the relationship between markup constancy and efficiency breaks in our model. Constant markups are neither necessary nor sufficient to ensure allocations are efficient. We have shown the existence of a parameterization of our model ($\xi = -\infty$) in which the economy achieves efficient selection and concentration, and thus efficient allocations of final consumption. Further, within this allocation markups are variable and decreasing in marginal costs (increasing in productivity). A consensus seems to suggest an increasing empirical relationship between markups and productivities.[8] Indeed, such variable markups may be efficient if consumers' time allocation decisions are accounted for, and the total value of final consumption is linear in a consumer's marginal cost of off-market time utilization, $1/\zeta$ (i.e., at $\xi = -\infty$).

4.1 Efficiency and demand elasticities

In the foregoing analyses we have characterized efficient and equilibrium allocations in an environment in which consumers value variety in market consumption but also have some capacity to substitute off-market time for market consumption. We now wish to relate this once more to the analysis of Dhingra and Morrow (2019), who allow for an arbitrary utility function over consumption but do not allow for off-market time. Have we simply considered a disguised form of their environment for a different choice of utility? That is, does incorporating off-market time simply change demand elasticities but not the interpretation of these elasticities? We will show by way of an example that the answer to both of these questions is unambiguously "no." The environments appear to be fundamentally different. To show this we will produce the exact same demand schedules in Dhingra and Morrow (2019), but show that whereas theirs are inefficient, ours are efficient.

Consider an environment in which firms are homogeneous and consumers do not value off-market time and have utility function $u(q) = q^\rho - \eta q$ for some $\eta > 0$. The elasticity of utility is then

$$\epsilon(q; u) = \frac{qu'(q)}{u(q)} = \rho - \frac{(1-\rho)\eta}{q^{\rho-1} - \eta}$$

and so $\epsilon'(q; u) < 0$. The results of Dhingra and Morrow (2019) (see Section B in page 211) then imply that there is too much entry in equilibrium. Given a multiplier λ on the budget constraint, the first-order conditions of the consumer for demand of product type κ are $\rho q(\kappa)^{\rho-1} = \lambda p(\kappa) + \eta$, which rearrange to give

$$q(\kappa) = (\lambda/\rho)^{-\frac{1}{1-\rho}}(p(\kappa) + \eta/\lambda)^{-\frac{1}{1-\rho}}, \tag{4.2}$$

with λ and the mass of firms M_e then determined by the zero profits condition and the consumer's budget constraint. The point of this calculation is that the equilibrium demand schedules in the economy with $u(q) = q^\rho - \eta q$ and no home production are identical to those in an economy with Leontief preferences and home production in which $\zeta = \lambda/\eta$ (although $\overline{T}$ will differ). However, the equilib-

[8] See, e.g., Edmond and Veldkamp (2009); Berry et al. (2019); Peters (2020).

rium in the economy with time constraints is efficient while the equilibrium in the above example economy is not. How can this be? Note that in the latter economy, the $1/\zeta$ term appearing in the demand schedule is exogenous and represents the real resource cost of time use, while the η/λ term appearing in the demand schedule of the former economy is endogenous and can be affected by taxes or subsidies on entry.

While Dhingra and Morrow (2019) focus on the relationship between markups and their utility elasticity defined above, their results speak to the efficiency of equilibrium allocations when the elasticity of substitution is either constant (CES) or variable (VES) across product types. Indeed, the relationship between the elasticity of substitution for market consumption and markups is an important feature of models with monopolistic competition. In the original Dixit and Stiglitz (1977) with homogeneous firms, this elasticity is a constant function of average aggregate markups. Dhingra and Morrow (2019) show that when firms are heterogeneous in costs (productivities) then competitive markups are only efficient if and only if this elasticity is constant and identical for substitution between all pairs of products (CES). Both the original Dixit and Stiglitz (1977) and Dhingra and Morrow (2019), however, consider environments with inelastic time use. Bilbiie et al. (2012) were the first to introduce labor into a model of monopolistic competition to study business cycle dynamics around the relationship between new varieties and firm entry. In Bilbiie et al. (2019) the authors build upon their analyses in Bilbiie et al. (2012) by exploring the welfare implications of a monopolistic competition model with labor. Several of their results are important for our analyses, primary of which is to show, as in Dhingra and Morrow (2019), that CES preferences over market quantities achieve allocative efficiency. Two fundamental differences between the setup in Bilbiie et al. (2019) and ours is that we assume time use is not necessarily separable with market consumption while *also* allowing for consumers to make multiple off-market time-use decisions.

Given these differences our approach thus warrants further analysis pertaining to the elasticities of substitution for *market consumption* that come out of our model. When it comes to elasticities of substitution, the literature has established a relationship between markups and the elasticity of substitution for *market* consumption. This object need not be constant in our model, despite having a constant elasticity of substitution for *final* consumption (i.e., $1/(1-\rho)$). Let $\sigma(p, p') = \ln(p'/p)$ (i.e., the log of the price ratios of two different firms' chosen $p' \neq p$, where the firm associated with price p' has cost κ'). The elasticity of substitution is defined as

$$\mu\big(p, p', \sigma(p, p')\big) := \frac{\partial \ln(q(p)/q(p'))}{\partial \sigma}.$$

We allow for μ to depend not only on relative prices, σ, but also the price levels, p and p'. It actually turns out that price *levels*, not just *relatives*, will matter for this elasticity except under very specific parameterizations. Thus, our model, under various parameter conditions, allows for variable elasticities of substitution (VES) across varieties.

The literature has established that when time use is either inelastic or elastic and additively separable with market consumption, CES preferences are both necessary and sufficient to generate efficient markups, which themselves are constrained to be constant. We seek to understand how the relationship between the elasticity of substitution and markups affects allocative efficiency when relaxing the assumptions of inelastic time use and additive separability between time use and market consumption. Proposition 4.7 establishes conditions under which the elasticity of substitution is constant. It turns out that the relationship between ξ and ρ is important for determining whether preferences in $q(p)$ are VES or CES. There are three specific cases in which preferences are CES, where two of those three (when $\xi = -\infty$ or ρ) yield elasticities of substitution that are identical to those in Dixit and Stiglitz (1977) and Dhingra and Morrow (2019).

Proposition 4.7. *For $\alpha \in (0, 1)$, preferences are VES in market consumption, $q(p)$, except in the special cases in which $\xi \in \{-\infty, 0, \rho\}$, when we have CES preferences, and the elasticities of substitution are:*

(i) For $\xi = -\infty$, $\mu = \frac{1}{1-\rho}$.

(ii) For $\xi = 0$, $\mu = \frac{\rho+1}{1-\rho}$.

(iii) For $\xi = \rho$, $\mu = \frac{1}{1-\rho}$.

Proof. See Appendix D. $\square$

When home production is Leontief or $\xi = \rho$, the elasticities of substitution for q's are constant and equivalent to those for c's. This is somewhat surprising because markups are non-constant but efficient when $\xi = -\infty$, even though they are constant and inefficient when $\xi = \rho$. Further, holistic markups are constant when $\xi = -\infty$ or $\xi = 0$, even though at $\xi = 0$ the elasticity of substitution for q's is not equivalent to the same elasticities for c's: it is shifted by the additive constant $\rho/(1-\rho)$. Still, it appears that CES preferences in q with off-market time use are still *necessary* to achieve efficiency (see $\xi \in \{-\infty, 0\}$ cases), but they are not *sufficient* (see $\xi = \rho$ case).

5 Conclusion

This paper has presented a parsimonious model that shows how the welfare effects of markups depend upon the extent to which consumers value off-market time. The key insight is that in our model, markups (which we have termed "posted" markups) differ from a more holistic definition of markups that incorporates the fact that the price paid by the consumer is not the sole economic cost they incur when consuming a good. For the case of perfect complements between market consumption and off-market time, we have shown that heterogeneous markups can be consistent with a first-best allocation.

We believe that our work ought to encourage researchers to continue to consider how accounting for different household decision structures affect inferences regarding efficiency when faced with market structures that allow for prices to exceed marginal costs. Further, from a policy standpoint, the notion that markup heterogeneity in and of itself leads to welfare losses should be reconsidered. We have shown that there exists an economy in which variable markups are indeed efficient. The recent work of Parenti et al. (2017) represents a different enrichment of the standard consumer problem in which efficiency is also consistent with heterogeneous markups. It thus remains for researchers to determine which decision structures themselves are most plausible when assessing what such structures imply for the efficiency of environments in which firms operate with imperfect competition.

References

Mark Aguiar and Erik Hurst. Consumption versus expenditure. *Journal of Political Economy*, 113(5), 2005.

Mark Aguiar and Erik Hurst. Life-cycle prices and production. *The American Economic Review*, 97(5), 2007.

Mark Aguiar and Erik Hurst. Deconstructing life cycle expenditure. *Journal of Political Economy*, 121(3), 2013.

Mark Aguiar and Erik Hurst. The macroeconomics of time allocation. In *Handbook of Macroeconomics*, volume 2, pages 203–253. Elsevier, 2016.

Mark Aguiar, Erik Hurst, and Loukas Karabarbounis. Recent developments in the economics of time use. *Annual Review of Economics*, 4:373–397, 2012.

Gary S. Becker. A Theory of the Allocation of Time. *The Economic Journal*, 75 (299):493–517, 1965. doi:10.2307/2228949.

W.L. Bednar and Nick Pretnar. Structural change with time to consume. *Working Paper*, 2023.

Kristian Behrens, Giordano Mion, Yasusada Murata, and Jens Suedekum. Quantifying the Gap Between Equilibrium and Optimum under Monopolistic Competition. *The Quarterly Journal of Economics*, 135(4):2299–2360, November 2020. doi:10.1093/qje/qjaa017.

Steven Berry, Martin Gaynor, and Fiona Scott Morton. Do increasing markups matter? lessons from empirical industrial organization. *Journal of Economic Perspectives*, 33(3):44–68, 2019.

Florin O. Bilbiie, Fabio Ghironi, and Marc J. Melitz. Endogenous entry, product variety, and business cycles. *Journal of Political Economy*, 120(2):304–345, April 2012.

Florin O. Bilbiie, Fabio Ghironi, and Marc J. Melitz. Monopoly Power and Endogenous Product Variety: Distortions and Remedies. *American Economic Journal: Macroeconomics*, 11(4):140–174, October 2019. doi:10.1257/mac.20170303.

Corina Boar and Virgiliu Midrigan. Markups and Inequality, 2022.

Job Boerma and Loukas Karabarbounis. Inferring inequality with home production. *Econometrica (forthcoming)*, 2021.

Benjamin Bridgman. Home productivity. *Journal of Economic Dynamics and Control*, 71:60–76, 2016.

Benjamin Bridgman, Georg Duernecker, and Berthold Herrendorf. Structural transformation, marketization, and household production around the world. *Journal of Development Economics*, 133:102–126, 2018.

Jan De Loecker, Jan Eeckhout, and Gabriel Unger. The Rise of Market Power and the Macroeconomic Implications*. *The Quarterly Journal of Economics*, 135(2):561–644, May 2020. doi:10.1093/qje/qjz041.

Swati Dhingra and John Morrow. Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity. *Journal of Political Economy*, 127(1): 196–232, September 2019. doi:10.1086/700732.

Avinash K. Dixit and Joseph E. Stiglitz. Monopolistic Competition and Optimum Product Diversity. *The American Economic Review*, 67(3):297–308, 1977. ISSN 0002-8282.

Chris Edmond and Laura Veldkamp. Income dispersion and counter-cyclical markups. *Journal of Monetary Economics*, 56(6):791–804, September 2009.

Chris Edmond, Virgiliu Midrigan, and Daniel Yi Xu. How Costly Are Markups? *Journal of Political Economy*, November 2022. doi:10.1086/722986.

Paolo Epifani and Gino Gancia. Trade, markup heterogeneity and misallocations. *Journal of International Economics*, 83(1):1–13, January 2011.

Lei Fang, Anne Hannusch, and Pedro Silos. Bundling time and goods: Implications for hours dispersion. *Working Paper*, 2022.

Zach Flynn, Amit Gandhi, and James Traina. Measuring markups with production data, 2019.

Gilbert R. Ghez and Gary S. Becker. *The Allocation of Time and Goods Over the Life Cycle*. Columbia University Press: New York, 1974.

Antonella Nocco, Gianmarco I.P. Ottaviano, and Matteo Salto. Monopolistic competition and optimum product selection: Why and how heterogeneity matters. *Discussion Paper no. 1206, Centre Econ. Performance, London School of Economics*, 2013.

Mathieu Parenti, Philip Ushchev, and Jacques-Francoise Thisse. Toward a theory of monopolistic competition. *Journal of Economic Theory*, 167:86–115, January 2017.

Michael Peters. Heterogeneous markups, growth, and endogenous misallocation. *Econometrica*, 88(5):2037–2073, September 2020.

Nick Pretnar. Measuring inequality with consumption time. *Working Paper*, 2022.

Evgeny Zhelobodko, Sergey Kokovin, Mathieu Parenti, and Jacques-François Thisse. Monopolistic Competition: Beyond the Constant Elasticity of Substitution. *Econometrica*, 80(6):2765–2784, 2012. doi:10.3982/ecta9986.

A Preliminary algebra

In this appendix we derive some preliminary observations and algebra that are used throughout the paper. First recall the definition of the total cost of a consumption experience given in the main text,

$$\psi(\kappa; \alpha, \xi) = \frac{\kappa}{\alpha}\Big(\alpha + (1-\alpha)[(1/\alpha - 1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\Big)^{1-1/\xi} \tag{A.1}$$

and the ratio of equilibrium market consumption to consumption experiences,

$$\psi'(\kappa; \alpha, \xi) = \Big(\alpha + (1-\alpha)[(1/\alpha - 1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\Big)^{-1/\xi}. \tag{A.2}$$

Lemma A.1. *For any α, ζ or κ we have*

$$\lim_{z\to 0}\left(\alpha + (1-\alpha)[(1/\alpha-1)\zeta\kappa]^{z}\right)^{1/z} = [(1/\alpha - 1)\zeta\kappa]^{1-\alpha}.$$

Proof. we first use l'Hopital's rule to obtain

$$\begin{aligned}
&\lim_{z\to 0}\frac{1}{z}\ln\left(\alpha + (1-\alpha)[(1/\alpha-1)\zeta\kappa]^{z}\right)\\
&= \lim_{z\to 0}(1-\alpha)\ln[(1/\alpha-1)\zeta\kappa]\frac{e^{z\ln[(1/\alpha-1)\zeta\kappa]}}{\alpha + (1-\alpha)e^{z\ln[(1/\alpha-1)\zeta\kappa]}}\\
&= \lim_{z\to 0}(1-\alpha)\ln[(1/\alpha-1)\zeta\kappa]
\end{aligned} \tag{A.3}$$

from which the limit in the lemma follows by raising to the exponential. □

Lemma A.1 gives the following expressions for the total cost ψ and marginal total ψ' in the Cobb-Douglas and Leontief cases.

Lemma A.2. *When $\xi = -\infty$ or $\xi = 0$, the total cost ψ and marginal cost ψ' become*

$$\begin{aligned}
\psi(\kappa; \alpha, 0) &= \alpha^{-\alpha}[(1-\alpha)\zeta]^{-(1-\alpha)}\kappa^{\alpha}\\
\psi'(\kappa; \alpha, 0) &= [(1/\alpha - 1)\zeta\kappa]^{-(1-\alpha)}\\
\psi(\kappa; \alpha, -\infty) &= \kappa + 1/\zeta\\
\psi'(\kappa; \alpha, -\infty) &= 1.
\end{aligned} \tag{A.4}$$

Proof. Calculation is straightforward by noting that $n = q/\zeta$ in the Leontief case and $n = (1/\alpha - 1)\kappa q$ in the Cobb-Douglas case. Then use (3.3), (3.4), and (3.5) to get the answer. □

Recalling the definition $\widehat{\pi}(\kappa; \alpha, \xi) := \max_{p\geq 0}\ (p-\kappa)\psi'(p; \alpha, \xi)\psi(p; \alpha, \xi)^{-\frac{1}{1-\rho}}$, we have the following.

Lemma A.3 (Special cases for $\widehat{\pi}$)**.** *When $\xi = 0$, the function $\widehat{\pi}$ becomes*

$$\widehat{\pi}(\kappa; \alpha, 0) = E(\alpha, \zeta, \rho)\kappa^{-\frac{\rho\alpha}{1-\rho}} \tag{A.5}$$

for some constant $E(\alpha, \zeta, \rho)$, while for $\xi = -\infty$ we have

$$\widehat{\pi}(\kappa; \alpha, -\infty) = \rho^{\frac{1}{1-\rho}}(1/\rho - 1)(\kappa + 1/\zeta)^{-\frac{\rho}{1-\rho}}. \tag{A.6}$$

Proof. Omitting arguments for ease of notation, when $\xi = 0$ we have

$$\begin{aligned}
\widehat{\pi}(\kappa; \alpha, 0) &= \max_{p \geq 0} \, (p - \kappa)[(1/\alpha - 1)\zeta p]^{-(1-\alpha)}\left(\alpha^{-\alpha}[(1-\alpha)\zeta]^{-(1-\alpha)}p^{\alpha}\right)^{-\frac{1}{1-\rho}} \\
&= D(\alpha, \zeta, \rho) \max_{p \geq 0} \, (p - \kappa)p^{-\frac{\rho\alpha}{1-\rho} - 1}
\end{aligned}$$

where $D(\alpha, \zeta, \rho) := \alpha^{\frac{1}{1-\rho}}[(1/\alpha - 1)\zeta]^{\frac{\rho(1-\alpha)}{1-\rho}}$ is a constant that is irrelevant to the firm's pricing decision. The first-order condition for the price is

$$0 = -\left(\frac{\rho\alpha}{1-\rho} + 1\right)(p - \kappa)p^{-\frac{\rho\alpha}{1-\rho} - 2} + p^{-\frac{\rho\alpha}{1-\rho} - 1}$$

which then simplifies to $(\rho\alpha/(1-\rho) + 1)(p - \kappa) = p$, and hence

$$p/\kappa = 1/\rho + (1/\rho - 1)(1/\alpha - 1).$$

Substitution then gives $p - \kappa = [(1/\rho - 1)/\alpha]\kappa$

$$\begin{aligned}
\widehat{\pi}(\kappa) &= D(\alpha, \zeta, \rho)[(1/\rho - 1)/\alpha]\kappa((1/\rho + (1/\rho - 1)(1/\alpha - 1))\kappa)^{-\frac{\rho\alpha}{1-\rho} - 1} \\
&=: E(\alpha, \zeta, \rho)\kappa^{-\frac{\rho\alpha}{1-\rho}}
\end{aligned}$$

for some (again unimportant) constant $E(\alpha, \zeta, \rho)$. For the Leontief case, we have

$$\widehat{\pi}(\kappa; \alpha, -\infty) := \max_{p \geq 0} \, (p - \kappa)(p + 1/\zeta)^{-\frac{1}{1-\rho}}.$$

The first-order condition is then

$$0 = -\frac{1}{1-\rho}(p - \kappa)(p + 1/\zeta)^{-\frac{1}{1-\rho} - 1} + (p + 1/\zeta)^{-\frac{1}{1-\rho}}$$

which simplifies to $p + 1/\zeta = (\kappa + 1/\zeta)/\rho$. Substitution then gives the claimed expression for $\widehat{\pi}$. $\square$

B Efficient allocations

We now record the proofs for all claims pertaining to the characterization of efficient allocations.

Proof of Lemma 3.1. The derivatives are

$$\begin{aligned}
\frac{\partial(n/q)}{\partial\kappa} &= \frac{\zeta^{\frac{\xi}{1-\xi}}(1/\alpha-1)}{1-\xi}[(1/\alpha-1)\kappa]^{\frac{\xi}{1-\xi}} \\
\frac{\partial^2(n/q)}{\partial\kappa^2} &= \frac{\xi\,\zeta^{\frac{\xi}{1-\xi}}(1/\alpha-1)^2}{(1-\xi)^2}[(1/\alpha-1)\kappa]^{\frac{2\xi+1}{1-\xi}}
\end{aligned}$$

where $\partial(n/q)/\partial\kappa > 0$ for all ξ and $\partial^2(n/q)/\partial\kappa^2 < 0$ only when $\xi < 0$. When home production is Leontief, $q = \zeta n$ and the ratio follows. □

Proof of Lemma 3.2. Explicit calculation gives

$$\begin{aligned}
\psi'(\kappa) = \frac{1}{\alpha}&\Big(\alpha + (1-\alpha)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\Big)^{1-1/\xi} \\
&+ (1-1/\xi)\times\frac{\xi}{1-\xi}\times(1-\alpha)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\kappa^{-1} \\
&\times\frac{\kappa}{\alpha}\Big(\alpha + (1-\alpha)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\Big)^{-1/\xi} \\
&= \frac{1}{\alpha}\Big(\alpha + (1-\alpha)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\Big)^{1-1/\xi} \\
&- (1-\alpha)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\frac{1}{\alpha}\Big(\alpha + (1-\alpha)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\Big)^{-1/\xi}
\end{aligned}$$

which yields

$$\psi'(\kappa) = \Big(\alpha + (1-\alpha)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\Big)^{-1/\xi} > 0$$

always. For the second-order condition we have

$$\begin{aligned}
\psi''(\kappa) = -\frac{1}{\xi}&\Big(\alpha + (1-\alpha)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\Big)^{-1/\xi-1} \\
&\times\frac{\xi}{1-\xi}\times(1-\alpha)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\times\kappa^{-1} < 0
\end{aligned}$$

always. Using the above expression for $\psi'(x)$, we have

$$\frac{\psi(\kappa)}{\psi'(\kappa)} = \kappa + (1/\alpha-1)\kappa[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}$$

which gives the claimed expression for the elasticity. When $\xi = -\infty$ the elasticity is just $\kappa/(\kappa+1/\zeta)$, which is still increasing in κ. At $\xi=0$ it is readily apparent, by direct calculation, that $\epsilon(\kappa;\psi)=\alpha$. $\square$

Lemma B.1. *The elasticity of the marginal total cost with respect to productivity is given by* $\epsilon(\kappa;\psi')=(\epsilon(\kappa;\psi)-1)/(1-\xi)$.

Proof. Using the expression for the second derivative derived in the proof of Lemma 3.2, we have

$$\epsilon(\kappa;\psi')=\frac{\kappa\psi''(\kappa)}{\psi'(\kappa)}=-\frac{1}{1-\xi}\left(\frac{(1/\alpha-1)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}}{1+(1/\alpha-1)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}}\right)$$

which simplifies as claimed. $\square$

Proof of Proposition 3.3. The first-order conditions for consumption experiences in the problem (3.7) are $\rho c(\kappa)^{\rho-1}=\lambda_p\psi(\kappa)$ or

$$c(\kappa)=(\rho/\lambda_p)^{\frac{1}{1-\rho}}\psi(\kappa)^{-\frac{1}{1-\rho}}. \tag{B.1}$$

The multiplier λ_p is obtained by substituting (B.1) into the resource constraint. We define $I(\overline{\kappa},M_e):=\overline{T}-(f_e+fG(\overline{\kappa}))M_e$. This is the value of all remaining resources after efficient entry margins have been established. Note that if (3.9) holds, then $I(\overline{\kappa},M_e)=\rho\overline{T}$. From $I(\overline{\kappa},M_e)$ efficient allocations of final consumption experiences will satisfy the following resource constraint:

$$I(\overline{\kappa},M_e)=\int_{\underline{\kappa}}^{\overline{\kappa}}\psi(\kappa)c(\kappa)M_eG(d\kappa)=(\rho/\lambda_p)^{\frac{1}{1-\rho}}\int_{\underline{\kappa}}^{\overline{\kappa}}\psi(\kappa)^{-\frac{\rho}{1-\rho}}M_eG(d\kappa).$$

Given (3.5) the efficient level of market consumption, $q(\kappa)$, is obvious. The expression (B.1) then becomes that in (3.10). Substitution of (3.10) into (3.7) gives the planner's problem for a particular M_e and $\overline{\kappa}$. Denoting the last quantity by $W(M_e,\overline{\kappa})$, we have

$$W(M_e,\overline{\kappa})=\int_{\underline{\kappa}}^{\overline{\kappa}}c(\kappa)^{\rho}M_eG(d\kappa)=I(\overline{\kappa},M_e)^{\rho}M_e^{1-\rho}\left(\int_{\underline{\kappa}}^{\overline{\kappa}}\psi(\kappa)^{-\frac{\rho}{1-\rho}}G(d\kappa)\right)^{1-\rho}.$$

Taking logs and dividing by ρ then gives the problem

$$\max_{M_e,\overline{\kappa}}\left\{\begin{aligned}&\ln(\overline{T}-(f_e+fG(\overline{\kappa}))M_e)+(1/\rho-1)\ln M_e\\&+(1/\rho-1)\ln\left(\int_{\underline{\kappa}}^{\overline{\kappa}}\psi(\kappa)^{-\frac{\rho}{1-\rho}}G(d\kappa)\right)\end{aligned}\right\}. \tag{B.2}$$

The first-order condition with respect to M_e then implies that the optimal mass of firms is given by (3.9). Finally, the problem of the planner choosing the cutoff is then equivalent to

$$\max_{\overline{\kappa}>0}\left\{\ln\left(\int_{\underline{\kappa}}^{\overline{\kappa}}\psi(\kappa)^{-\frac{\rho}{1-\rho}}G(d\kappa)\right)-\ln(f_e+fG(\overline{\kappa}))\right\}.$$

The optimal cutoff $\overline{\kappa}$ is then characterized by the first-order condition (3.8). Finally, clearly with homogeneous firms $G(\overline{\kappa})=1$, and the expressions for $\overline{c}$ and $\overline{q}$ are immediate. □

C Equilibrium allocations

We now record the proofs for all claims pertaining to the characterization of the monopolistically competitive equilibrium allocations. We first use Lemma B.1 to characterize the price elasticity of demand in terms of the cost elasticity.

Lemma C.1. *The price elasticity of demand is given by*

$$\epsilon(p;q)=\frac{1}{1-\xi}\left(1+\frac{(\rho-\xi)}{(1-\rho)}\epsilon(p;\psi)\right). \tag{C.1}$$

Consequently, $\epsilon(p;q)$ is increasing in p when $\xi\in(-\infty,0)$ and decreasing in p when $\xi\in(0,\rho)$.

Proof. Demand is a multiple of $\psi'(p)\psi(p)^{-\frac{1}{1-\rho}}$, and so we compute

$$\frac{d}{d\kappa}\left[\psi'(p)\psi(p)^{-\frac{1}{1-\rho}}\right]=\psi''(p)\psi(p)^{-\frac{1}{1-\rho}}-\frac{1}{1-\rho}[\psi'(p)]^2\psi(p)^{-\frac{1}{1-\rho}-1}.$$

Dividing by $\psi'(p)\psi(p)^{-\frac{1}{1-\rho}}p^{-1}$ and using Lemma B.1, we have

$$\epsilon(p;q)=-\left(\frac{\epsilon(p;\psi)-1}{1-\xi}-\frac{\epsilon(p;\psi)}{1-\rho}\right)$$

which gives the result. □

Proof of Proposition 3.4. Using Lemma C.1, the first-order condition (3.15) of the firm becomes

$$\frac{p}{p-\kappa}=\epsilon(p;q)=\frac{1}{1-\xi}\left(1-\frac{(\xi-\rho)}{(1-\rho)}\epsilon(p;\psi)\right)$$

which is equivalent to the equation $\Xi(p;\kappa) = 0$ where

$$\Xi(p;\kappa) = -\frac{(1-\rho)}{(1-\xi)} + \frac{(1-\rho)p}{p-\kappa} - \epsilon(p;\psi)\frac{(\rho-\xi)}{(1-\xi)}. \tag{C.2}$$

Note that because $\xi \leq \rho$ and $\epsilon(p;\psi) \leq 1$ for all $p \geq 0$, any root of (C.2) weakly exceeds the solution to $1 + (\rho-\xi)/(1-\rho) = (1-\xi)p/(p-\kappa)$, which is precisely $p = \kappa/\rho$. If $\xi < 0$, then the right-hand side of (C.2) is decreasing and tends to $1 + (\xi - 1)\rho/(1-\rho) < 1$ as $p \to \infty$, and so there exists a unique solution to the equation $\Xi(p;\kappa) = 0$. For $\xi > 0$, we can rearrange the first-order condition to

$$\epsilon(p;\psi)^{-1} = \frac{(\rho-\xi)/(1-\rho)}{(1-\xi)\kappa/(p-\kappa)-\xi} = \frac{(p-\kappa)(\rho-\xi)}{(\kappa-\xi p)(1-\rho)}.$$

Ignoring irrelevant constants and taking logarithms, at the points at which the first-order condition is satisfied, the objective of the firm agrees with the function

$$\begin{aligned} Y(p) &:= (1-\rho)\ln(p-\kappa) - \ln p + (\rho/\xi - 1)\ln\left(\frac{p-\kappa}{\kappa-\xi p}\right) \\ &= \rho(1/\xi - 1)\ln(p-\kappa) - \ln p + (1-\rho/\xi)\ln(\kappa - \xi p). \end{aligned}$$

Note that since $\xi \leq \rho$, the derivative satisfies

$$Y'(p) = \frac{\rho(1/\xi-1)}{p-\kappa} - 1/p + \frac{\rho-\xi}{\kappa-\xi p} \geq \left(\frac{\rho(1/\xi-1)p}{p-\kappa} - 1\right)p^{-1}$$

and so it will suffice to show that the right-hand side of this last inequality is positive for all candidate solutions. Using (C.2), we have

$$\rho/\xi + (\rho/\xi)\frac{(\rho-\xi)}{(1-\rho)}\epsilon(p;\psi) = \frac{\rho(1/\xi-1)p}{p-\kappa}$$

and hence $\rho(1/\xi - 1)p/(p-\kappa) \geq \rho/\xi \geq 1$, as desired. There is therefore a unique maximum in the firm's problem and it is the largest solution to the first-order conditions. The comparative statics with respect to κ follow from Topkis' theorem, while the case-by-case analysis follows from direct calculations. □

Proof of Lemma 3.5. This follows Topkis' theorem, the first-order condition (3.15), and the comparative statics in Lemma C.1. □

Lemma C.2. *If $F, H : [\underline{\kappa}, \infty) \to \mathbb{R}$ are two smooth, decreasing, positive-valued functions satisfying $F'(x)/F(x) \geq H'(x)/H(x)$ for all $x \in [\underline{\kappa}, \infty)$, then $J(\kappa; F) \leq J(\kappa; H)$ for all $\kappa \in [\underline{\kappa}, \infty)$.*

Proof. First note that $J(\underline{\kappa}; F) = J(\underline{\kappa}; G) = 0$ and that the derivative of J with respect to productivity is

$$\begin{aligned} J'(\overline{\kappa}; F) &= g(\overline{\kappa}) - F'(\overline{\kappa}) \int_{\underline{\kappa}}^{\overline{\kappa}} (F(\kappa)/F(\overline{\kappa})^2) G(d\kappa) \\ &= g(\overline{\kappa}) - (F'(\overline{\kappa})/F(\overline{\kappa})) J(\overline{\kappa}; F). \end{aligned}$$

The desired conclusion then follows from the non-negativity of $J(\overline{\kappa}; F)$ together with the assumption $-F'(x)/F(x) \leq -H'(x)/H(x)$. $\square$

Lemma C.3 and Lemma C.4 are technical observations that will be used in the proof of Proposition 4.1.

Lemma C.3. *If $\xi < 0$, then the inequality*

$$(1 - \rho/\xi + z)(1 + z)^{\xi/(1-\xi)} > \rho^{\frac{1}{1-\xi}}(1 - 1/\xi). \tag{C.3}$$

holds for all $\rho \in (0, 1)$ and $z \geq 0$.

Proof. We will prove (C.3) by showing that it holds for $z = 0$ and that the left-hand side is increasing. For $z = 0$, (C.3) becomes $1 - \rho/\xi > \rho^{\frac{1}{1-\xi}}(1 - 1/\xi)$, which for $\xi < 0$ is equivalent to $\xi - \rho < \rho^{\frac{1}{1-\xi}}(\xi - 1)$ or $\xi < \rho - \rho^{\frac{1}{1-\xi}} + \rho^{\frac{1}{1-\xi}}\xi$. This last inequality is an equality at $\rho = 1$ and the derivative with respect to ρ of the right-hand side is $1 - \rho^{\frac{\xi}{1-\xi}}/(1-\xi) + \rho^{\frac{\xi}{1-\xi}}\xi/(1-\xi) = 1 - \rho^{\frac{\xi}{1-\xi}} < 0$ for all $\rho \in (0, 1]$, which shows that equation (C.3) holds for $z = 0$. We now take logarithms of the left-hand side of (C.3) and take the derivative with respect to z to obtain

$$\frac{1}{1 - \rho/\xi + z} + \frac{(1 + z)^{-1}}{1/\xi - 1}. \tag{C.4}$$

The expression in (C.4) is positive for $z \geq 0$ if and only if $1 - 1/\xi + z(1 - 1/\xi) > 1 - \rho/\xi + z$, or $-z/\xi > (1 - \rho)/\xi$, which is always true. $\square$

Lemma C.4. *If $\xi \in (0, \rho)$ then the inequality*

$$(1 + z)^{\xi/(1-\xi)}[\rho - \xi(1 + z)] < \rho^{1/(1-\xi)}(1 - \xi) \tag{C.5}$$

holds for all $z \geq 0$ and $\rho \in (0, 1)$.

Proof. When $z = 0$, (C.5) becomes $0 < \rho^{1/(1-\xi)}(1 - \xi) + \xi - \rho$. For $\rho = 1$ this last inequality holds with equality, and the derivative of the right-hand side with respect to ρ is $\rho^{\xi/(1-\xi)} - 1$, which is negative for all $\rho \in (0, 1)$. This ensures that the inequality (C.5) holds for all $\rho \in (0,1)$ when $z = 0$. To establish (C.5) for all $z \geq 0$, note that we may rewrite it as

$$0 < \rho^{\xi/(1-\xi)}(1-\xi) + (1+z)^{1/(1-\xi)}(\xi/\rho) - (1+z)^{\xi/(1-\xi)}$$

The derivative of the right-hand side with respect to z is then

$$(1+z)^{1/(1-\xi)-1}(\xi/\rho)/(1-\xi) - (1+z)^{\xi/(1-\xi)-1}\xi/(1-\xi),$$

which is non-negative if and only if $\xi \leq (1+z)(\xi/\rho)$, which is true for all $z \geq 0$. □

Proof of Proposition 3.6. Indifference at the cutoff is given by $\pi(\overline{\kappa}) = f$, and aggregate profits per firm net of fixed costs are given by $\int_{\underline{\kappa}}^{\overline{\kappa}} \pi(\kappa)G(d\kappa) - f_e - fG(\overline{\kappa})$. Combining these two conditions and using the fact that $\widehat{\pi}(\kappa)/\pi(\kappa)$ is independent of κ, we obtain equation (3.18). To establish that there exists a unique solution to equation (3.18), first note that $\widehat{\pi}'(\kappa) < 0$ everywhere by the envelope theorem. The desired $\overline{\kappa}$ is a positive root of the function

$$H(\overline{\kappa}) := \frac{f\int_{\underline{\kappa}}^{\overline{\kappa}} \widehat{\pi}(\kappa)G(d\kappa)}{\widehat{\pi}(\overline{\kappa})(f_e + fG(\overline{\kappa}))} - 1.$$

Note that $H(\underline{\kappa}) = -1$ and $\limsup_{\overline{\kappa}\to\infty} H(\overline{\kappa}) = \infty$, because the numerator is increasing and the denominator is bounded above by $\widehat{\pi}(\overline{\kappa})(f_e + f)$ and therefore vanishes as $\overline{\kappa} \to \infty$. Now consider the function $L(\overline{\kappa}) := f\int_{\underline{\kappa}}^{\overline{\kappa}} \widehat{\pi}(\kappa)G(d\kappa) - \widehat{\pi}(\overline{\kappa})(f_e + fG(\overline{\kappa}))$ and note that

$$L'(\overline{\kappa}) = f\widehat{\pi}(\overline{\kappa})g(\overline{\kappa}) - f\widehat{\pi}(\overline{\kappa})g(\overline{\kappa}) - \widehat{\pi}'(\overline{\kappa})(f_e + fG(\overline{\kappa})) > 0$$

which establishes uniqueness. To determine the equilibrium mass of firms in the environment with firm heterogeneity, note that by (3.14), the profits of a firm with productivity $\kappa \in [\underline{\kappa}, \overline{\kappa}]$ are equal to

$$\pi(\kappa) = \frac{(\overline{T}/M_e)(p(\kappa) - \kappa)}{\int_{\underline{\kappa}}^{\overline{\kappa}} \psi(p(\kappa))^{-\frac{\rho}{1-\rho}}G(d\kappa)}\psi'(p(\kappa))\psi(p(\kappa))^{-\frac{1}{1-\rho}}.$$

It follows that $\pi(\overline{\kappa}) = f$ if and only if the mass of firms satisfies (3.19). Finally, note that when firms are homogeneous aggregate profits must equate with the sum of both fixed and operating costs, $f + f_e$. Because $q(\overline{p}) = \frac{\overline{T}}{M_e\overline{p}}(1 + (1/\alpha - 1)^{1/(1-\xi)}[\zeta\overline{p}]^{\xi/(1-\xi)})$ profit maximization clearly yields (3.21).

Finally, to obtain $c(p(\kappa))$, note that the first-order conditions for the problem (3.13) give the demand for each variety as a function of the price, $c(p) =$

$(\rho/\lambda)^{\frac{1}{1-\rho}}\psi(p)^{-\frac{1}{1-\rho}}$, where λ is the multiplier on the budget constraint. The market consumption per unit of final consumption demanded from the firm is $q(p)/c(p) = \psi'(p)$. The multiplier λ is then obtained by substituting consumption into the budget constraint and rearranging, which gives (3.20). For $\bar{c}$ note that $G(\overline{\kappa})$ is a point mass, and the result follows. $\square$

D Welfare analyses

Proof of Proposition 4.1. The claims regarding $\xi \in \{-\infty, 0\}$ follow from the explicit calculations stated in Lemma A.3, because in this case $\psi(\kappa)^{-\frac{\rho}{1-\rho}}/\widehat{\pi}(\kappa)$ is constant in κ. For the remaining claims, by (3.17) and Lemma C.2, it will suffice to establish that for $\xi \in (-\infty, 0)$, we have

$$-\frac{1}{p(\kappa)-\kappa} = \frac{\widehat{\pi}'(\kappa)}{\widehat{\pi}(\kappa)} \geq \frac{(d/d\kappa)\psi(\kappa)^{-\frac{\rho}{1-\rho}}}{\psi(\kappa)^{-\frac{\rho}{1-\rho}}} = -\frac{\psi'(\kappa)/\psi(\kappa)}{1/\rho - 1} \tag{D.1}$$

for all $\kappa \geq 0$, and that the reverse inequality in (D.1) holds for $\xi \in (0, \rho)$. The inequality (D.1) may then be written

$$p(\kappa)/\kappa \geq \frac{1/\rho - 1}{\epsilon(\kappa; \psi)} + 1 =: \widehat{p}(\kappa)/\kappa \tag{D.2}$$

where the equality on the right-hand side of (D.2) defines $\widehat{p}(\kappa)$. For ease of notation, in this proof we will write

$$z(\kappa) := (1-\rho)(1/\alpha - 1)^{\frac{1}{1-\xi}}[\zeta\kappa]^{\frac{\xi}{1-\xi}} \tag{D.3}$$

for $\kappa \geq 0$. As shown in the proof of Proposition 3.4, the price chosen by the firm is a solution to $\Xi(p; \kappa) = 0$, where Ξ is defined in equation (C.2). We then want to show that $\Xi(\widehat{p}(\kappa); \kappa) > 0$ for all $\kappa > 0$ when $\xi < 0$, and $\Xi(\widehat{p}(\kappa); \kappa) < 0$ for all $\kappa > 0$ when $\xi \in (0, \rho)$. Substitution gives

$$\Xi(\widehat{p}(\kappa); \kappa) = -\frac{(1-\rho)\xi}{(1-\xi)} + \rho\epsilon(\kappa; \psi) - \epsilon(\widehat{p}(\kappa); \psi)\frac{(\rho - \xi)}{(1-\xi)}. \tag{D.4}$$

We first suppose that $\xi < 0$, in which case $\Xi(\widehat{p}(\kappa); \kappa) > 0$ is equivalent to

$$(\rho - \xi)\epsilon(\widehat{p}(\kappa); \psi) < \rho(1-\xi)\epsilon(\kappa; \psi) - (1-\rho)\xi. \tag{D.5}$$

When $\xi < 0$, both sides of (D.5) are positive and so taking the reciprocal and simplifying gives

$$\epsilon(\widehat{p}(\kappa);\psi)^{-1} - 1 > \frac{\rho(1-\xi)[1-\epsilon(\kappa;\psi)]}{\rho(1-\xi)\epsilon(\kappa;\psi) - (1-\rho)\xi}$$
$$= \frac{\rho(1-\xi)[\epsilon(\kappa;\psi)^{-1} - 1]}{\rho - \xi - (1-\rho)\xi[\epsilon(\kappa;\psi)^{-1} - 1]}.$$

Using the expression for the elasticity in Lemma 3.2 and simplifying gives

$$(1+z(\kappa))^{\xi/(1-\xi)} = [\rho\widehat{p}(\kappa)/\kappa]^{\xi/(1-\xi)} > \frac{\rho^{\frac{1}{1-\xi}}(1-\xi)}{\rho - \xi - (1-\rho)\xi(1/\alpha - 1)^{\frac{1}{1-\xi}}[\zeta\kappa]^{\frac{\xi}{1-\xi}}}$$
$$= \frac{\rho^{\frac{1}{1-\xi}}(1-1/\xi)}{1-\rho/\xi + z(\kappa)},$$

which is true by Lemma C.3. For $\xi > 0$, the desired inequality $\Xi(\widehat{p}(\kappa);\kappa) < 0$ is equivalent to

$$(1-\rho)\xi + (\rho - \xi)\epsilon(\widehat{p}(\kappa);\psi) > \rho(1-\xi)\epsilon(\kappa;\psi). \tag{D.6}$$

Since $\xi \in (0,\rho)$, both sides of (D.6) are positive and so taking the reciprocal gives

$$\frac{\rho(1-\xi)}{(1-\rho)\xi + (\rho-\xi)\epsilon(\widehat{p}(\kappa);\psi)} < \frac{1}{\epsilon(\kappa;\psi)}.$$

This is equivalent to

$$\frac{\rho(1-\xi)\epsilon(\widehat{p}(\kappa);\psi)}{(1-\rho)\xi + (\rho-\xi)\epsilon(\widehat{p}(\kappa);\psi)} < \frac{\epsilon(\widehat{p}(\kappa);\psi)}{\epsilon(\kappa;\psi)}$$
$$\frac{\epsilon(\kappa;\psi)}{\epsilon(\widehat{p}(\kappa);\psi)} < \frac{(1-\rho)\xi + (\rho-\xi)\epsilon(\widehat{p}(\kappa);\psi)}{\rho(1-\xi)\epsilon(\widehat{p}(\kappa);\psi)}$$

Simplifying further then gives

$$\frac{\epsilon(\widehat{p}(\kappa);\psi)^{-1} - \epsilon(\kappa;\psi)^{-1}}{\epsilon(\kappa;\psi)^{-1}} < \frac{\xi(1-\rho)}{\rho(1-\xi)}\big(\epsilon(\widehat{p}(\kappa);\psi)^{-1} - 1\big)$$
$$\widehat{p}(\kappa)^{\xi/(1-\xi)} - \kappa^{\xi/(1-\xi)} < \frac{\xi(1-\rho)}{\rho(1-\xi)}\widehat{p}(\kappa)^{\xi/(1-\xi)}\epsilon(\kappa;\psi)^{-1}.$$

Further rearrangement gives

$$[\widehat{p}(\kappa)/\kappa]^{\xi/(1-\xi)}\left[1 - \frac{\xi(1-\rho)}{\rho(1-\xi)}\epsilon(\kappa;\psi)^{-1}\right] < 1$$
$$[\rho\widehat{p}(\kappa)/\kappa]^{\xi/(1-\xi)}\big[1 - \xi - (\xi/\rho)(1-\rho)\epsilon(\kappa;\psi)^{-1}\big] < \rho^{\xi/(1-\xi)}(1-\xi)$$

which is equivalent to $(1+z(\kappa))^{\xi/(1-\xi)}[\rho-\xi-\xi z(\kappa)] < \rho^{1/(1-\xi)}(1-\xi)$, and is therefore true by Lemma C.4.
Finally, at the limit when $\xi=\rho$, (D.6) simply becomes

$$(1-\rho)\rho > \rho(1-\rho)\epsilon(\kappa;\psi)$$

which is always true since $\epsilon(\kappa;\psi)<1$ always. □

Proof of Lemma 4.3. First, when $\xi=-\infty$, $p(\kappa)=\kappa(1/\rho+(1/\rho-1)/\zeta)$ and $\psi(x)=x+1/\zeta$ for arbitrary x, yielding $\varphi=1/\rho$. Second, when $\xi=0$, $p(\kappa)$ is linear in κ and $\varphi=m^{\alpha-1}$, where m is the posted markup which is constant, as described in Proposition 3.4. □

Proof of Lemma 4.4. Noting that $p(\kappa)=m(\kappa)\kappa$, the holistic markup is

$$\varphi(\kappa)=\frac{\psi(m(\kappa)\kappa)}{\psi(\kappa)}=m(\kappa)\frac{\left(\alpha+(1-\alpha)[(1/\alpha-1)\zeta m(\kappa)\kappa]^{\frac{\xi}{1-\xi}}\right)^{1-1/\xi}}{\left(\alpha+(1-\alpha)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\right)^{1-1/\xi}}. \tag{D.7}$$

Clearly from the proof of Lemma 4.3 $\varphi=1/\rho<m$ when $\xi=-\infty$ and $\varphi=m^{\alpha-1}<m$, since $\alpha\in(0,1)$ when $\xi=0$. When $\xi<0$, note that $m>1$ implies $m\kappa>\kappa$ and $(m\kappa)^{\frac{\xi}{1-\xi}}<\kappa^{\frac{\xi}{1-\xi}}$ since $\xi/(1-\xi)<0$. It is thus readily apparent that $\varphi<m$. When $0<\xi\le\rho$, we have that $m\kappa>\kappa$ implies

$$\begin{aligned}&\alpha+(1-\alpha)[(1/\alpha-1)\zeta m\kappa]^{\frac{\xi}{1-\xi}}>\alpha+(1-\alpha)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\\ \Rightarrow\quad&\left(\alpha+(1-\alpha)[(1/\alpha-1)\zeta m\kappa]^{\frac{\xi}{1-\xi}}\right)^{1-1/\xi}<\left(\alpha+(1-\alpha)[(1/\alpha-1)\zeta\kappa]^{\frac{\xi}{1-\xi}}\right)^{1-1/\xi}\end{aligned}$$

since $1-1/\xi<0$, and the result is attained. □

Proof of Proposition 4.5. In view of Proposition 3.3 and Proposition 3.6, there will be too few firms in equilibrium precisely when

$$1>\frac{1-\kappa/p(\kappa)}{1-\rho}\times\epsilon(p(\kappa);\psi). \tag{D.8}$$

Note that by (C.2), $p(\kappa)$ satisfies

$$1-\kappa/p(\kappa)=1-\xi-(1-\kappa/p(\kappa))\frac{(\rho-\xi)}{(1-\rho)}\epsilon(p;\psi) \tag{D.9}$$

and more simplification gives

$$\frac{\kappa/p(\kappa)-\xi}{\rho-\xi}=\frac{1-\kappa/p(\kappa)}{1-\rho}\times\epsilon(p;\psi)$$

which is the right-hand side of (D.8). The equilibrium mass of firms will therefore be too low if $\kappa/p(\kappa)-\xi<\rho-\xi$, which is always true for finite ξ by Proposition 3.4. When $\xi=-\infty$, $\epsilon(\overline{p};\psi)=1$, $\overline{\kappa}/\overline{p}=\rho$, and the efficient and equilibrium M_e are identical. □

Proof of Proposition 4.6. Here, $\xi=-\infty$. Note that in equilibrium $\psi(p(\kappa))=\frac{1}{\rho}(\kappa+1/\zeta)$, and $\widehat{\pi}(\kappa)=(1-\rho)\big[(1/\rho)(\kappa+1/\zeta)\big]^{-\frac{\rho}{1-\rho}}$, so that (3.19) becomes

$$M_e=\frac{\overline{T}(1-\rho)\,\big[(1/\rho)(\overline{\kappa}+1/\zeta)\big]^{-\frac{\rho}{1-\rho}}}{f\int_{\underline{\kappa}}^{\overline{\kappa}}\big[(1/\rho)(\kappa+1/\zeta)\big]^{-\frac{\rho}{1-\rho}}G(d\kappa)},$$

which reduces to

$$M_e=\frac{\overline{T}(1-\rho)\,\psi(\overline{\kappa})^{-\frac{\rho}{1-\rho}}}{f\int_{\underline{\kappa}}^{\overline{\kappa}}\psi(\kappa)^{-\frac{\rho}{1-\rho}}G(d\kappa)}. \tag{D.10}$$

Setting (D.10) equal to (3.9) (the efficient mass) and rearranging gives the expression for (3.8), the efficient cutoff. This can only be true if the equilibrium mass when $\xi=-\infty$ is itself efficient. □

Proof of Proposition 4.7. Note that when home production is Leontief, $q(p)=\zeta n(p)$, and $\mathcal{C}=\int_{\underline{\kappa}}^{\overline{\kappa}}q(p(\kappa))^{\rho}M_eG(d\kappa)$, so that the elasticity of substitution is clearly just $1/(1-\rho)$. For finite ξ we start with the ratio

$$\frac{q(p)}{q(p')}=\frac{\psi'(p)}{\psi'(p')}\left(\frac{\psi(p)}{\psi(p')}\right)^{-\frac{1}{1-\rho}}. \tag{D.11}$$

In a slight abuse of notation, note that $\psi'(p')$ and $\psi'(p)$ can be written

$$\psi'(p)\equiv\psi'((p'/p)^{-1};p')=(p')^{\frac{1}{\xi-1}}\left(\alpha(p')^{\frac{\xi}{\xi-1}}+(1-\alpha)[(1/\alpha-1)\zeta(p'/p)^{-1}]^{\frac{\xi}{1-\xi}}\right)^{-1/\xi}$$
$$\psi'(p')\equiv\psi'(p'/p;p)=p^{\frac{1}{\xi-1}}\left(\alpha p^{\frac{\xi}{\xi-1}}+(1-\alpha)[(1/\alpha-1)\zeta(p'/p)]^{\frac{\xi}{1-\xi}}\right)^{-1/\xi}.$$

A similar expression can be derived for ψ:

$$\psi(p)\equiv\psi((p'/p)^{-1};p')=\frac{1}{\alpha}\left(\frac{p'}{p}\right)^{-1}\left(\alpha(p')^{\frac{\xi}{\xi-1}}+(1-\alpha)[(1/\alpha-1)\zeta(p'/p)^{-1}]^{\frac{\xi}{1-\xi}}\right)^{\frac{\xi-1}{\xi}}$$
$$\psi(p')\equiv\psi(p'/p;p)=\frac{1}{\alpha}\left(\frac{p'}{p}\right)\left(\alpha p^{\frac{\xi}{\xi-1}}+(1-\alpha)[(1/\alpha-1)\zeta(p'/p)]^{\frac{\xi}{1-\xi}}\right)^{\frac{\xi-1}{\xi}}.$$

Equation (D.11) can then be written

$$\frac{q(p)}{q(p')} = \frac{\psi'((p'/p)^{-1};p')}{\psi'(p'/p;p)}\left(\frac{\psi((p'/p)^{-1};p')}{\psi(p'/p;p)}\right)^{-\frac{1}{1-\rho}}. \tag{D.12}$$

Replacing $\ln(p'/p) = \sigma$ in (D.12) and taking logs, note that we can split the right-hand side into two pieces:

$$\begin{aligned}
\ln\left(\frac{\psi'(\sigma^{-1};p')}{\psi'(\sigma;p)}\right) &= \frac{1}{\xi-1}\sigma - \frac{1}{\xi}\ln\left(\alpha(p')^{\frac{\xi}{\xi-1}} + (1-\alpha)[(1/\alpha-1)\zeta e^{-\sigma}]^{\frac{\xi}{1-\xi}}\right) \\
&\quad + \frac{1}{\xi}\ln\left(\alpha p^{\frac{\xi}{\xi-1}} + (1-\alpha)[(1/\alpha-1)\zeta e^{\sigma}]^{\frac{\xi}{1-\xi}}\right) \\
\frac{1}{\rho-1}\ln\left(\frac{\psi(\sigma^{-1};p')}{\psi(\sigma;p)}\right) &= \frac{2}{1-\rho}\sigma \\
&\quad - \frac{\xi-1}{\xi(1-\rho)}\ln\left(\alpha(p')^{\frac{\xi}{\xi-1}} + (1-\alpha)[(1/\alpha-1)\zeta e^{-\sigma}]^{\frac{\xi}{1-\xi}}\right) \\
&\quad + \frac{\xi-1}{\xi(1-\rho)}\ln\left(\alpha p^{\frac{\xi}{\xi-1}} + (1-\alpha)[(1/\alpha-1)\zeta e^{\sigma}]^{\frac{\xi}{1-\xi}}\right).
\end{aligned}$$

Let $\widetilde{\mu}(\sigma;p) = \ln\left(\alpha p^{\frac{\xi}{\xi-1}} + (1-\alpha)[(1/\alpha-1)\zeta e^{\sigma}]^{\frac{\xi}{1-\xi}}\right)$. Combining the two pieces and their like terms gives

$$\ln\left(\frac{q(p)}{q(p')}\right) = \frac{2\,\xi-\rho-1}{(\xi-1)(1-\rho)}\sigma + \frac{\rho-\xi}{\xi(1-\rho)}\widetilde{\mu}(\sigma^{-1};p') + \frac{\xi-\rho}{\xi(1-\rho)}\widetilde{\mu}(\sigma;p).$$

Generally speaking, the elasticity of substitution is

$$\mu\big(p,p',\sigma(p,p')\big) = \frac{2\,\xi-\rho-1}{(\xi-1)(1-\rho)} - \frac{\rho-\xi}{\xi(1-\rho)}\frac{\partial\widetilde{\mu}(\sigma^{-1};p')}{\partial\sigma^{-1}} + \frac{\xi-\rho}{\xi(1-\rho)}\frac{\partial\widetilde{\mu}(\sigma;p)}{\partial\sigma}$$

When $\xi = \rho$ the latter two terms in the above expression are 0. Note that $\partial\widetilde{\mu}/\partial\sigma = 0$ when $\xi \in \{-\infty, 0\}$, otherwise

$$\frac{\partial\widetilde{\mu}}{\partial\sigma} = \frac{\xi(1-\alpha)[(1/\alpha-1)\zeta]^{\frac{\xi}{1-\xi}}e^{\frac{2\sigma\xi-\sigma}{1-\xi}}}{(1-\xi)\big(\alpha p^{\frac{\xi}{\xi-1}} + (1-\alpha)[(1/\alpha-1)\zeta e^{\sigma}]^{\frac{\xi}{1-\xi}}\big)}$$

which varies in σ and p non-trivially. □