

NBER WORKING PAPER SERIES

MONOPOLY POWER AND ENDOGENOUS PRODUCT VARIETY:
DISTORTIONS AND REMEDIES

Florin O. Bilbiie
Fabio Ghironi
Marc J. Melitz

Working Paper 14383
<http://www.nber.org/papers/w14383>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2008

For helpful comments, we thank Sanjay Chugh, Ippei Fujiwara, Hugo Hopenhayn, Giammario Impullitti, Henning Weber, Mirko Wiederholt, and participants in numerous seminars and conferences. We are grateful to Nicholas Sim, Pinar Uysal, and Benjamin Chen for excellent research assistance, and to Pablo Winant for computational help. Bilbiie gratefully acknowledges without implicating the support of Banque de France via the eponymous Chair at PSE, and of Institut Universitaire de France. Ghironi and Melitz thank the NSF for financial support through a grant to the NBER. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2008 by Florin O. Bilbiie, Fabio Ghironi, and Marc J. Melitz. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Monopoly Power and Endogenous Product Variety: Distortions and Remedies
Florin O. Bilbiie, Fabio Ghironi, and Marc J. Melitz
NBER Working Paper No. 14383
October 2008, Revised June 2016
JEL No. D42,H32,L16

ABSTRACT

The inefficiencies related to endogenous product creation and variety under monopolistic competition are two-fold: one static—the misalignment between consumers and producers regarding the value of a new variety; and one dynamic—time variation in markups. Quantitatively, the welfare costs of the former are potentially very large relative to the latter. For a calibrated version of our model with these distortions, their total cost amounts to 2 percent of consumption. Appropriate taxation schemes can implement the optimum amount of entry and variety. Elastic labor introduces a further distortion that should be corrected by subsidizing labor at a rate equal to the markup for goods, in order to preserve profit margins and hence entry incentives.

Florin O. Bilbiie
PSE, Sorbonne, CEPR
florin.bilbiie@gmail.com

Fabio Ghironi
Department of Economics
University of Washington
Savery Hall, Box 353330
Seattle, WA 98195
and NBER
ghiro@uw.edu

Marc J. Melitz
Department of Economics
Harvard University
215 Littauer Center
Cambridge, MA 02138
and NBER
mmelitz@harvard.edu

1 Introduction

What are the consequences of monopoly power for the efficiency of business cycle fluctuations and new product creation? If market power results in inefficiency, how large are the welfare costs of inefficient entry and variety? How do they depend upon structural parameters, and what tools can the policymaker employ to maximize social welfare and restore efficiency? We address these questions in the context of the dynamic, stochastic, general equilibrium (DSGE) model with monopolistic competition and endogenous product creation developed in Bilbiie, Ghironi, and Melitz (2012 – henceforth, BGM). Specifically, we compare the competitive and planner equilibria, contrasting the market solution with the efficient responses to exogenous shocks yielding the optimum amount of product variety when product creation is subject to sunk costs, a time-to-build lag, and an obsolescence risk. We then quantify the welfare cost of those distortions in a calibrated version of the model that has been used by BGM to study U.S. business cycle data on entry, product creation, and the cyclicalities of profits and markups. Lastly, we outline some fiscal policies that ensure implementation of the Pareto optimum as a competitive equilibrium when efficiency of the market solution fails. The policy schemes that implement efficiency in our model fully specify the optimal path of the relevant distortionary instruments over the business cycles triggered by unexpected shocks to productivity and entry costs.¹

Our main theorem identifies two distortions as the sources of inefficient entry and product variety in this dynamic model with general preferences over consumption varieties. The first distortion, which we label “static,” pertains to the intratemporal misalignment between the benefit on an extra variety to the consumer, and the profit incentive for an entrant to produce that extra variety. The second distortion, which we label “dynamic,” is associated with the intertemporal variation of markups. Both distortions disappear if and only if preferences are of the C.E.S. form originally studied by Dixit and Stiglitz (1977)—case in which our dynamic market equilibrium is also efficient. We quantify the welfare cost of these inefficiencies in a calibrated version of our model and find that they can be sizable, in particular those due to the static distortion. For a calibration used by BGM (2012) to replicate US business cycle moments regarding entry, profits, and markups, the cost is very large: around 2 percentage points of consumption.

The policymaker can use a variety of fiscal instruments (in conjunction with lump-sum taxes or

¹By studying the efficiency properties of our model, this paper contributes to the literature on the efficiency properties of monopolistic competition started by the original work of Lerner (1934) and developed by Samuelson (1947), Spence (1976), Dixit and Stiglitz (1977), and Grossman and Helpman (1991), among others. See also Mankiw and Whinston (1986), Benassy (1996), Kim (2004), and Opp, Parlour and Walden (2014).

transfers) to alleviate these distortions and ensure implementation of the first-best equilibrium. We study an example consisting of a combination of appropriately designed VAT and dividend/profit taxes that can implement the first-best equilibrium. The dividend/profit tax aligns firms' entry incentives with consumers' love for variety, while the VAT tax corrects for the intertemporally inefficient allocation of resources that is due to the intertemporal misalignment of markups.

Efficiency also requires that markups be synchronized across *all* items that bring utility (or disutility) to consumers. We show this by considering the case of endogenous labor supply, thereby introducing a leisure good that is not subject to a markup; this opens a wedge between marginal rates of substitution and transformation between consumption and leisure that distorts labor supply. Efficiency is restored if the government taxes leisure (or subsidizes labor supply) at a rate equal to the net markup in consumption goods prices, even if goods remain priced above marginal cost. While this result also holds in a model with a fixed number of firms, an equivalent optimal policy in that setup would have the markup removed by a proportional revenue subsidy. In our model, such a policy of inducing marginal cost pricing – if financed with lump-sum taxation of firm profits – would eliminate entry incentives, since the sunk entry cost could not be covered in the absence of profits.² Our results show that monopoly profits should in fact be *preserved* whenever product variety is endogenously determined by firm entry, for they play a crucial role in generating the welfare-maximizing level of product variety in equilibrium.³

The framework and results developed herein provide the foundation for a number of applications and extensions that have appeared in subsequent normative analyses of different policies in macroeconomic models. To give some examples, Bilbiie, Ghironi, and Melitz (2008) rely on results in this paper when discussing optimal monetary policy in a sticky-price version of the model in which policy can deliver the first-best outcome. Bilbiie, Fujiwara, and Ghironi (2014) consider the case of Ramsey-optimal monetary policy in a second-best environment. Bergin and Corsetti (2008, 2014), Cooke (2015), Etro and Rossi (2015), Faia (2009, 2012), and Lewis (2013) also build on the framework and insights developed herein—or use related frameworks—to provide results on optimal monetary policy in a variety of scenarios.⁴ In fiscal policy analysis, Chugh and Ghironi

²We are implicitly assuming that the government is not contemporaneously subsidizing the entire amount of the entry cost. In Appendix E, we show that inducing marginal cost pricing can implement the efficient equilibrium in our model only when the lump-sum taxation that finances the necessary sales subsidy is optimally split between households and firms, and that this requires zero lump-sum taxation of firm profits when preferences are of the form studied by Dixit and Stiglitz (1977).

³Our results thus stand in sharp contrast to the common policy prescription of *eliminating* monopoly profits, found in a large body of literature studying optimal monetary and fiscal policy in the presence of monopolistic competition.

⁴Cacciatore, Fiori, and Ghironi (2016) and Cacciatore and Ghironi (2012) integrate results and intuitions in this paper in their analyses of the interaction of optimal monetary policy with market reforms, such as reductions in entry

(2015) use our model to study Ramsey-optimal fiscal policy. Lewis and Wikler (2015) focus on the consequences of oligopolistic competition in a model that is simplified by removing dynamic features of our analysis, while Colciago (2015) introduces oligopolistic competition in our dynamic framework. Bertolotti and Etro (2015, 2016) and Etro (2016) use a setup with more general (and not necessarily homothetic) preferences, and also study normative implications. Epifani and Gancia (2011) study misallocation resulting from heterogeneous markups, and the welfare effects of trade liberalization. Dhingra and Morrow (2013) look at optimum product variety with heterogeneous firms and variable elasticity of substitution. The results in the aforementioned (and other) studies build upon or otherwise use the normative insights developed in this paper.

The structure of the paper is as follows. Section 2 describes the benchmark model with fixed labor supply and characterizes the competitive equilibrium and (in subsection 2.6) the Pareto-optimal allocation of the social planner. Section 3 states and proves our welfare theorem, and discusses the intuition for it; the same section computes numerically the welfare costs of inefficient variety in our model and studies optimal fiscal policies that implement the first-best allocation. Section 4 extends the analysis to the case of endogenous labor supply, and section 5 concludes.

2 A Model of Endogenous Entry and Product Variety

This section outlines the model and solves for the monopolistically competitive market equilibrium and for the Pareto-optimal planner equilibrium, respectively.

2.1 Household Preferences

The economy is populated by a unit mass of atomistic households. We begin by assuming that the representative household supplies L units of labor inelastically in each period at the nominal wage rate W_t . The household maximizes expected intertemporal utility from consumption (C): $E_t \sum_{s=t}^{\infty} \beta^{s-t} U(C_s)$, where $\beta \in (0, 1)$ is the subjective discount factor and $U(C)$ is a period utility function with the standard properties. At time t , the household consumes the basket of goods C_t , defined as a homothetic aggregate over a continuum of goods Ω . At any given time t , only a subset of goods $\Omega_t \subset \Omega$ is available. Let $p_t(\omega)$ denote the nominal price of a good $\omega \in \Omega_t$. Our model can be solved for any parametrization of symmetric homothetic preferences. For any such preferences, there exists a well defined homothetic consumption index C_t and an associated welfare-based price

barriers in a monetary union or trade integration.

index P_t . The demand for an individual variety, $c_t(\omega)$, is then obtained as $c_t(\omega)d\omega = C_t\partial P_t/\partial p_t(\omega)$, where we use the conventional notation for quantities with a continuum of goods as flow values.⁵

Given the demand for an individual variety, $c_t(\omega)$, the symmetric price elasticity of demand is in general a function of the number N_t of goods/producers (where N_t is the mass of Ω_t , and θ measures the elasticity of substitution):

$$\theta(N_t) \equiv -\frac{\partial c_t(\omega)}{\partial p_t(\omega)} \frac{p_t(\omega)}{c_t(\omega)}, \quad \text{for any symmetric variety } \omega.$$

The benefit of additional product variety is captured by the relative price ρ :

$$\rho_t(\omega) = \rho(N_t) \equiv \frac{p_t(\omega)}{P_t}, \quad \text{for any symmetric variety } \omega,$$

or, in elasticity form:

$$\epsilon(N_t) \equiv \frac{\rho'(N_t)}{\rho(N_t)} N_t.$$

Together, $\theta(N_t)$ and $\rho(N_t)$ completely characterize the choice of symmetric homothetic preferences in our model; explicit expressions can be obtained for these objects upon specifying functional forms for preferences, as will become clear in the discussion below.

2.2 Firms

There is a continuum of monopolistically competitive firms, each producing a different variety $\omega \in \Omega$. Production requires only one factor, labor. Aggregate labor productivity is indexed by Z_t , which represents the effectiveness of one unit of labor. Z_t is exogenous and follows an $AR(1)$ process (in logarithms). Output supplied by firm ω is $y_t(\omega) = Z_t l_t(\omega)$, where $l_t(\omega)$ is the firm's labor demand for productive purposes. The unit cost of production, in units of the consumption good C_t , is w_t/Z_t , where $w_t \equiv W_t/P_t$ is the real wage.⁶

Prior to entry, firms face a sunk entry cost of $f_{E,t}$ effective labor units, equal to $w_t f_{E,t}/Z_t$ units of the consumption basket. $f_{E,t}$ is exogenous and follows an $AR(1)$ process (in logarithms). There are no fixed production costs. Hence, all firms that enter the economy produce in every period, until they are hit with a “death” shock, which occurs with probability $\delta \in (0, 1)$ in every period.⁷

⁵See the appendix for more details. Since this is a real model, the price of the final good is not determined; we use the final good as the numeraire.

⁶Consistent with standard *real business cycle* theory, aggregate productivity Z_t affects all firms uniformly.

⁷For simplicity, we do not consider endogenous exit. As we show in BGM, appropriate calibration of δ makes it possible for our model to match several important features of the data.

Given our modeling assumption relating each firm to an individual variety, we think of a firm as a production line for that variety, and the entry cost as the development and setup cost associated with the latter (potentially influenced by market regulation). The exogenous “death” shock also takes place at the individual variety level. Empirically, a firm may comprise more than one of these production lines, but – for simplicity – our model does not address the determination of product variety within firms.

Firms set prices in a flexible fashion as markups over marginal costs. In units of consumption, firm ω 's price is $\rho_t(\omega) = \mu_t w_t / Z_t$, where the markup μ_t is in general a function of the number of producers: $\mu_t = \mu(N_t) \equiv \theta(N_t) / (\theta(N_t) - 1)$. The firm's profit in units of consumption, returned to households as dividend, is $d_t(\omega) = \left(1 - \mu(N_t)^{-1}\right) Y_t^C / N_t$, where Y_t^C is total output of the consumption basket and will in equilibrium be equal to total consumption demand C_t .

Preference Specifications and Markups

We consider four alternative preference specifications as special cases for illustrative purposes below. The first preference specification features a constant elasticity of substitution between goods as in Dixit and Stiglitz (1977). For these C.E.S. preferences (henceforth, C.E.S.-DS), the consumption aggregator is $C_t = \left(\int_{\omega \in \Omega} c_t(\omega)^{(\theta-1)/\theta} d\omega\right)^{\theta/(\theta-1)}$, where $\theta > 1$ is the symmetric elasticity of substitution across goods. The consumption-based price index is then $P_t = \left(\int_{\omega \in \Omega_t} p_t(\omega)^{1-\theta} d\omega\right)^{1/(1-\theta)}$, and the household's demand for each individual good ω is $c_t(\omega) = (p_t(\omega) / P_t)^{-\theta} C_t$. It follows that the markup and the benefit of variety are independent of the number of goods: $\mu(N_t) - 1 = \epsilon(N_t) = \epsilon = 1 / (\theta - 1)$. The second specification a variant of C.E.S. with *generalized love of variety* introduced by the working paper version of Dixit and Stiglitz (1977) and used also by Benassy (1996). This variant disentangles monopoly power (measured by the net markup $1 / (\theta - 1)$) and consumer love for variety, captured by a constant parameter $\xi > 0$. With this specification (labelled “general C.E.S.” henceforth), the consumption basket is $C_t = (N_t)^{\xi - \frac{1}{\theta-1}} \left(\int_{\omega \in \Omega} c_t(\omega)^{(\theta-1)/\theta} d\omega\right)^{\theta/(\theta-1)}$. The third preference specification uses the *translog expenditure function* proposed by Feenstra (2003). For this specification, the symmetric price elasticity of demand is $\theta(N_t) = 1 + \sigma N_t$, where $\sigma > 0$ is a free parameter. The expression for relative price $\rho(N_t)$ is given in Table 1, where \tilde{N} is the measure of all possible varieties, $\tilde{N} \equiv \text{Mass}(\Omega)$. As N_t increases, goods become closer substitutes, and the elasticity of substitution increases. If goods are closer substitutes, then both the markup $\mu(N_t)$

and the benefit of additional varieties in elasticity form ($\epsilon(N_t)$) must decrease;⁸ for this specific functional form, the change in $\epsilon(N_t)$ is only *half* the change in the net markup generated by an increase in the number of producers. Finally, the fourth preference specification features *exponential love-of-variety* (we call it “exponential” for short) and is in some sense just the *opposite* of the General C.E.S. specification: the elasticity of substitution is not constant (because of demand-side pricing complementarities), but the benefit of variety is equal to the net markup.⁹ Specifically, the symmetric elasticity of substitution is of the same form as under translog $\theta(N_t) = 1 + \alpha N_t$, where $\alpha > 0$ is a free parameter. However, differently from translog, the relative price is given by $\rho(N_t) = e^{-\frac{1}{\alpha N_t}}$: it follows that the benefit of variety is equal to the markup (and profit incentive for entry): $\epsilon(N_t) = \mu(N_t) - 1 = 1/\alpha N_t$. Table 1 summarizes the expressions for markup, relative price, and benefit of variety in elasticity form for each preference specification.

Table 1. Four Preference Specifications: Markup, relative price and benefit of variety

C.E.S.-D.S.	General C.E.S.	Translog	Exponential
$\mu(N_t) = \mu = \frac{\theta}{\theta-1}$	$\mu(N_t) = \mu = \frac{\theta}{\theta-1}$	$\mu(N_t) = \frac{\theta(N_t)}{\theta(N_t)-1} = 1 + \frac{1}{\sigma N_t}$	$\mu(N_t) = \frac{\theta(N_t)}{\theta(N_t)-1} = 1 + \frac{1}{\alpha N_t}$
$\rho(N_t) = N_t^{\frac{1}{\theta-1}}$	$\rho(N_t) = N_t^\xi$	$\rho(N_t) = e^{-\frac{1}{2} \frac{N-N_t}{\sigma N N_t}}$	$\rho(N_t) = e^{-\frac{1}{\alpha N_t}}$
$\epsilon(N_t) = \mu - 1$	$\epsilon(N_t) = \xi$	$\epsilon(N_t) = \frac{1}{2\sigma N_t} = \frac{\mu(N_t)-1}{2}$	$\epsilon(N_t) = \frac{1}{\alpha N_t} = \mu(N_t) - 1$

Firm Entry and Exit

In every period, there is a mass N_t of firms producing in the economy and an unbounded mass of prospective entrants. These entrants are forward looking, and correctly anticipate their expected future profits $d_s(\omega)$ in every period $s \geq t + 1$ as well as the probability δ (in every period) of incurring the exit-inducing shock. Entrants at time t only start producing at time $t + 1$, which introduces a one-period time-to-build lag in the model. The exogenous exit shock occurs at the very end of the time period (after production and entry). A proportion δ of new entrants will therefore never produce. Prospective entrants in period t compute their expected post-entry value

⁸This property for the markup occurs whenever the price elasticity of residual demand decreases with quantity consumed along the residual demand curve.

⁹The exponential specification eliminates entry inefficiency but introduces markup misalignment over time, whereas the general C.E.S. specification features inefficient entry but with constant markups; in this sense, we refer to them as being “opposite”. Both distortions operate under translog preferences. It should be noted that the utility representation for both general C.E.S. and exponential preferences would involve the number of goods available, even if not consumed.

($v_t(\omega)$) given by the present discounted value of their expected stream of profits $\{d_s(\omega)\}_{s=t+1}^{\infty}$:

$$v_t(\omega) = E_t \sum_{s=t+1}^{\infty} [\beta(1-\delta)]^{s-t} \frac{U'(C_s)}{U'(C_t)} d_s(\omega). \quad (1)$$

This also represents the value of incumbent firms *after* production has occurred (since both new entrants and incumbents then face the same probability $1 - \delta$ of survival and production in the subsequent period). Entry occurs until firm value is equalized with the entry cost, leading to the free entry condition $v_t(\omega) = w_t f_{E,t}/Z_t$. This condition holds so long as the mass $N_{E,t}$ of entrants is positive. We assume that macroeconomic shocks are small enough for this condition to hold in every period.¹⁰ Finally, the timing of entry and production is such that the number of producing firms during period t is given by $N_t = (1 - \delta)(N_{t-1} + N_{E,t-1})$. The number of producing firms represents the capital stock of the economy. It is an endogenous state variable that behaves much like physical capital in a benchmark real business cycle (RBC) model.

Symmetric Firm Equilibrium

All firms face the same marginal cost. Hence, equilibrium prices, quantities, and firm values are identical across firms: $p_t(\omega) = p_t$, $\rho_t(\omega) = \rho_t$, $l_t(\omega) = l_t$, $y_t(\omega) = y_t$, $d_t(\omega) = d_t$, $v_t(\omega) = v_t$. In turn, equality of prices across firms implies that the consumption-based price index P_t and the firm-level price p_t are such that $p_t/P_t \equiv \rho_t = \rho(N_t)$. An increase in the number of firms is associated with an increase in this relative price: $\rho'(N_t) > 0$, capturing the love of variety utility gain. The aggregate consumption output of the economy is $Y_t^C = N_t \rho_t y_t = C_t$.

Importantly, in the symmetric firm equilibrium, the value of waiting to enter is zero, despite the entry decision being subject to sunk costs and exit risk; i.e., there are no option-value considerations pertaining to the entry decision. This happens because all uncertainty in our model (including the “death” shock) is aggregate.¹¹

2.3 Household Budget Constraint and Intertemporal Decisions

We assume without loss of generality that households hold only shares in a mutual fund of firms. Let x_t be the share in the mutual fund of firms held by the representative household *entering* period

¹⁰Periods with zero entry ($N_{E,t} = 0$) may occur as a consequence of large enough (adverse) exogenous shocks. In these periods, the free entry condition would hold as a strict inequality: $v_t(\omega) < w_t f_{E,t}/Z_t$.

¹¹See the appendix for the proof. In this paper, we assume that the exogenous shocks are small enough to rule out this possibility.

t . The mutual fund pays a total profit in each period (in units of currency) equal to the total profit of all firms that produce in that period, $P_t N_t d_t$. During period t , the representative household buys x_{t+1} shares in a mutual fund of $N_{H,t} \equiv N_t + N_{E,t}$ firms (those already operating at time t and the new entrants). Only $N_{t+1} = (1 - \delta) N_{H,t}$ firms will produce and pay dividends at time $t + 1$. Since the household does not know which firms will be hit by the exogenous exit shock δ at the *very end* of period t , it finances the continuing operation of all pre-existing firms and all new entrants during period t . The date t price (in units of currency) of a claim to the future profit stream of the mutual fund of $N_{H,t}$ firms is equal to the nominal price of claims to future firm profits, $P_t v_t$.

The household enters period t with mutual fund share holdings x_t and receives dividend income and the value of selling its initial share position, and labor income. The household allocates these resources between purchases of shares to be carried into next period, consumption, and lump-sum taxes T_t levied by the government. The period budget constraint (in units of consumption) is:

$$v_t N_{H,t} x_{t+1} + C_t + T_t = (d_t + v_t) N_t x_t + w_t L. \quad (2)$$

The household maximizes its expected intertemporal utility subject to (2). The Euler equation for share holdings is:

$$v_t = \beta (1 - \delta) E_t \left[\frac{U'(C_{t+1})}{U'(C_t)} (v_{t+1} + d_{t+1}) \right].$$

As expected, forward iteration of this equation and absence of speculative bubbles yield the asset price solution in equation (1).¹²

2.4 Aggregate Accounting and Equilibrium

Aggregating the budget constraint (2) across households and imposing the equilibrium condition $x_{t+1} = x_t = 1 \forall t$ yields the aggregate accounting identity $C_t + N_{E,t} v_t = w_t L + N_t d_t$: Total consumption plus investment (in new firms) must be equal to total income (labor income plus dividend income).

Different from the benchmark, one-sector, RBC model, our model economy is a two-sector economy in which one sector employs part of the labor endowment to produce consumption and the other sector employs the rest of the labor endowment to produce new firms. The economy's GDP, Y_t , is equal to total income, $w_t L + N_t d_t$. In turn, Y_t is also the total output of the economy, given by consumption output, $Y_t^C (= C_t)$, plus investment output, $N_{E,t} v_t$. With this in mind, v_t is

¹²We omit the transversality condition that must be satisfied to ensure optimality.

the relative price of the investment “good” in terms of consumption.

Labor market equilibrium requires that the total amount of labor used in production and to set up the new entrants’ plants must equal aggregate labor supply: $L_t^C + L_t^E = L$, where $L_t^C = N_t l_t$ is the total amount of labor used in production of consumption, and $L_t^E = N_{E,t} f_{E,t} / Z_t$ is labor used to build new firms. In the benchmark RBC model, physical capital is accumulated by using as investment part of the output of the same good used for consumption. In other words, all labor is allocated to the only productive sector of the economy. When labor supply is fixed, there are no labor market dynamics in the model, other than the determination of the equilibrium wage along a vertical supply curve. In our model, even when labor supply is fixed, labor market dynamics arise in the allocation of labor between production of consumption and creation of new plants. The allocation is determined jointly by the entry decision of prospective entrants and the portfolio decision of households who finance that entry. The value of firms, or the relative price of investment in terms of consumption v_t , plays a crucial role in determining this allocation.¹³

2.5 The Competitive Equilibrium

The model with general homothetic preferences is summarized in Table 2 in the Appendix C; as shown there, the model can be reduced to a system of two equations in two variables, N_t and C_t . In particular, the reduced-form Euler equation linking consumption and the number of goods is:

$$f_{E,t}\rho(N_t) = \beta(1 - \delta) E_t \left\{ \frac{U'(C_{t+1})}{U'(C_t)} \left[f_{E,t+1}\rho(N_{t+1}) \frac{\mu(N_t)}{\mu(N_{t+1})} + \frac{C_{t+1}}{N_{t+1}} \mu(N_t) \left(1 - \frac{1}{\mu(N_{t+1})} \right) \right] \right\}. \quad (3)$$

The number of new entrants as a function of consumption and number of firms is $N_{E,t} = Z_t L / f_{E,t} - C_t / (f_{E,t}\rho(N_t))$. Substituting this into the law of motion for N_t (scrolled forward one period) yields:

$$N_{t+1} = (1 - \delta) \left(N_t + \frac{Z_t L}{f_{E,t}} - \frac{C_t}{f_{E,t}\rho(N_t)} \right). \quad (4)$$

We are now in a position to define a *competitive equilibrium* of our economy.¹⁴

Definition 1: A *Competitive Equilibrium (CE)* consists of a 2-tuple $\{C_t, N_{t+1}\}$ satisfying (3) and (4) for a given initial value N_0 and a transversality condition for investment in shares.

¹³When labor supply is elastic, labor market dynamics operate along two margins as the interaction of household and firm decisions determines jointly the total amount of labor and its allocation to the two sectors of the economy.

¹⁴It is understood that we use ‘competitive equilibrium’ to refer to the equilibrium of the market economy in which firms compete in the assumed monopolistically competitive fashion with no intervention of the policymaker in the economy. Thus, the use of the word ‘competitive’ implies no reference to perfect competition.

The system of stochastic difference equations (3) and (4) has a unique stationary equilibrium under the following conditions. A steady-state CE satisfies:

$$f_E \rho(N) = \beta(1 - \delta) \left[f_E \rho(N) + \frac{C}{N} (\mu(N) - 1) \right],$$

$$C = Z \rho(N) L - \rho(N) f_E \frac{\delta}{1 - \delta} N.$$

After eliminating C , this system reduces to:

$$H^{CE}(N) \equiv \frac{ZL(1 - \delta)}{f_E \left(\frac{r + \delta}{\mu(N) - 1} + \delta \right)} = N,$$

where $r \equiv (1 - \beta) / \beta$.¹⁵

The steady-state number of firms in the CE, N^{CE} , is a fixed point of $H^{CE}(N)$. We assume that $\lim_{N \rightarrow 0} \mu(N) = \infty$ and $\lim_{N \rightarrow \infty} \mu(N) = 1$. Since $H^{CE}(N)$ is continuous, $\lim_{N \rightarrow 0} H^{CE}(N) = \infty$, and $\lim_{N \rightarrow \infty} H^{CE}(N) = 0$, $H^{CE}(N)$ has a unique fixed point if and only if $[H^{CE}(N)]' \leq 0$. Given

$$[H^{CE}(N)]' = \mu'(N) \frac{(1 - \delta)(r + \delta)ZL}{(r + \delta\mu(N))^2 f_E},$$

this will hold if and only if $\mu'(N) \leq 0$.

The intuition for the uniqueness condition is that more product variety leads to a “crowding in” of the product space and goods becoming closer substitutes (with C.E.S. a limiting case). This is a very reasonable condition: If goods were to become more differentiated as product variety increases, then the motivation for multiple equilibria would be apparent: There could be one equilibrium with many firms charging high markups and producing little, and another with few firms charging low markups and producing relatively more.

In BGM, we study the business cycle properties of the competitive equilibrium. In the present paper, we compare this with the planning optimum.

¹⁵Allowing households to hold bonds in our model would simply pin down the real interest rate as a function of the expected path of consumption determined by the system in Table 2. In steady state, the real interest rate would be such that $\beta(1 + r) = 1$. For notational convenience, we thus replace the expression $(1 - \beta) / \beta$ with r when the equations in Table 2 imply the presence of such term.

2.6 The Planning (Pareto) Optimum

Given the model just outlined, we now study a hypothetical scenario in which a benevolent planner maximizes lifetime utility of the representative household by choosing quantities directly (including the number of goods produced via the number of entrants).

The “production function” for aggregate consumption output is $C_t = Z_t \rho(N_t) L_t^C$. Hence, the problem solved by the planner can be written as:

$$\begin{aligned} \max_{\{L_s^C\}_{s=t}^{\infty}} E_t \sum_{s=t}^{\infty} \beta^{s-t} U(Z_s \rho(N_s) L_s^C), \\ \text{s.t. } N_{t+1} = (1 - \delta) N_t + (1 - \delta) \frac{(L - L_t^C) Z_t}{f_{E,t}}, \end{aligned}$$

or, substituting the constraint into the utility function and treating next period’s state as the choice variable:

$$\max_{\{N_{s+1}\}_{s=t}^{\infty}} E_t \sum_{s=t}^{\infty} \beta^{s-t} U \left[Z_s \rho(N_s) \left(L - \frac{1}{(1 - \delta)} \frac{f_{E,s}}{Z_s} N_{s+1} + \frac{f_{E,s}}{Z_s} N_s \right) \right]. \quad (5)$$

As we show in Appendix C, the first-order condition for this problem can be written as:

$$U'(C_t) \rho(N_t) f_{E,t} = \beta (1 - \delta) E_t \left\{ U'(C_{t+1}) \left[f_{E,t+1} \rho(N_{t+1}) + \frac{C_{t+1}}{N_{t+1}} \epsilon(N_{t+1}) \right] \right\}. \quad (6)$$

This equation, together with the dynamic constraint (4) (which is the same under the competitive and planner equilibria) leads to the following definition.

Definition 2: A *Planning Equilibrium (PE)* consists of a 2-tuple $\{C_t, N_{t+1}\}$ satisfying (4) and (6) for a given initial value N_0 .

The conditions for uniqueness of the stationary PE are similar to those for the CE found in the previous section. The steady-state number of firms N^{PE} is the fixed point of a function similar to $H^{CE}(N)$, where the variety effect $\epsilon(N)$ replaces the net markup:

$$H^{PE}(N) \equiv \frac{ZL(1 - \delta)}{f_E \left(\frac{r + \delta}{\epsilon(N)} + \delta \right)}.$$

Therefore, the system of stochastic difference equations (4) and (6) has a unique stationary equilibrium if and only if $\lim_{N \rightarrow 0} \epsilon(N) = \infty$, $\lim_{N \rightarrow \infty} \epsilon(N) = 0$, and $\epsilon'(N) \leq 0$.¹⁶ The intuition for

¹⁶Note that the solution for the stationary PE can be obtained by replacing the net markup function $\mu(N)$ in the stationary CE solution with the benefit of variety function $\epsilon(N)$.

these uniqueness conditions is analogous to the one for the competitive equilibrium: more product variety leads to a “crowding in” of product space and goods become closer substitutes (with C.E.S. a limiting case). In the PE case, this requires decreasing returns to increased product variety (very similar to the condition that goods become closer substitutes). C.E.S. is again a limiting case where there are “constant elasticity returns” to increased product variety: Doubling product variety, holding spending constant, always increases welfare by the same percentage.

3 A Welfare Theorem

We now state our main *theorem*, which provides the conditions under which the competitive (CE) and planner (PE) equilibria coincide with strictly positive entry costs.¹⁷

Theorem 1 *The Competitive and Planner equilibria are equivalent – i.e., $CE \Leftrightarrow PE$ – if and only if the following two conditions are jointly satisfied:*

- (i) $\mu(N_t) = \mu(N_{t+1}) = \mu$ and
- (ii) *the elasticity of product variety and the markup functions are such that $\epsilon(x) = \mu(x) - 1$.*

Proof. See Appendix. ■

The conditions of Theorem 1 basically imply that, for efficiency to obtain, *preferences must be of the C.E.S. form* studied by Dixit and Stiglitz (1977) – a special, knife-edge case of the general homothetic preferences for variety that we consider.

3.1 Sources of Inefficiency in Entry and Product Variety

Inefficiency occurs in our dynamic model of endogenous entry and variety through two distortions, associated with the failure of the conditions outlined in Theorem 1.

Static Distortion: When the welfare benefit of variety $\epsilon(N_t)$ and the net markup $\mu(N_t) - 1$ (which measures the profit incentive for firms to enter the market) are not aligned within a given period, entry is inefficient from a social standpoint. When, for instance, the benefit of variety is low compared to the desired markup ($\epsilon(N_t) < \mu(N_t) - 1$), the consumer surplus of creating a new variety is lower than the profit signal received by a potential entrant; equilibrium entry is therefore too high (with the size of the distortion being governed by the difference between the two objects). The opposite holds when $\epsilon(N_t) > \mu(N_t) - 1$. Inefficiency occurs, through this channel, if

¹⁷We focus on situations where a strictly positive sunk cost (related to technology or regulation) is associated with creating new firms.

new entrants ignore on the one hand the positive effect of a new variety on consumer surplus and on the other the negative effect on other firms' profits. We refer to this distortion as the “static entry distortion”, to highlight that it still operates in a static model, or in the steady state of our dynamic stochastic model.¹⁸ With C.E.S.-DS preferences, these two contrasting forces perfectly balance each other and the resulting equilibrium is efficient.¹⁹

Dynamic Distortion: Variations in desired markups over time (induced by changes in N_t) introduce an additional discrepancy—equal to the ratio $\mu(N_t)/\mu(N_{t+1})$ —between the “private” (competitive equilibrium) and “social” (Pareto optimum) return to a new variety. When there is entry, the future markup is lower than the current one, and this ratio increases, generating an additional inefficient reallocation of resources to entry in the current period. Just like differences in markups across goods imply inefficiencies (more resources should be allocated to the production of the high markup goods – a point we illustrate below in the case of endogenous labor supply), differences in markups over time/across states also imply inefficiencies: More resources should be allocated to production in periods/states with high markups. For example, if the social planner knew that productivity would be lower in the future (resulting in less entry and a higher markup), the optimal plan would be to develop additional varieties now, so that more labor can be used for production during low productivity periods. We label this the “dynamic entry distortion” below, making explicit that it operates only with preferences that allow for time-varying desired markups, such as the translog and exponential preferences we introduced.²⁰ Finally, notice that both distortions are operative for translog preferences.

¹⁸Under general C.E.S. preferences (the second column of Table 1) the static distortion is the only one operating. A feature of this preference specification that is important for its welfare implications is that consumers derive utility from goods that they never consume, and they are worse off when a good disappears even if consumption of that good was zero.

¹⁹See also Dixit and Stiglitz (1977), Judd (1985a), and Grossman and Helpman (1991) for a discussion of these issues.

²⁰The point that efficiency occurs with synchronized markups can be traced back to Lerner (1934) and Samuelson (1947). Lerner (1934, p. 172) first noted that the allocation of resources is efficient when markups are equal in the pricing of all goods: “The conditions for that optimum distribution of resources between different commodities that we designate the absence of monopoly are satisfied if prices are all proportional to marginal cost.” Samuelson (1947, p. 239-240) also makes this point clearly: “If all factors of production were indifferent between different uses and completely fixed in amount – the pure Austrian case –, then [...] proportionality of prices and marginal cost would be sufficient.” This makes it clear that *equality* of prices to marginal cost is *not necessary* for achieving an optimal allocation, contrary to an argument often found in the macroeconomic policy literature. This point is equally true in a model with a fixed number of firms N , where the planner merely solves a static allocation problem, allocating labor to the symmetric individual goods evenly.

3.2 The Welfare Costs of Inefficient Entry and Variety

What are the welfare costs of the distortions associated with entry and variety identified above? We answer this question in a calibrated version of our model, using the same calibration used by BGM to reproduce business cycle facts. In particular, a discount factor $\beta = 0.99$ (implying that the steady-state interest rate is $r = 0.01$), an exogenous destruction rate $\delta = 0.025$, a labor elasticity set to 4 ($\varphi = 0.25$) and an elasticity of substitution between goods $\theta = 3.8$. The sunk entry cost parameter f_E is normalized to 1, and productivity follows an $AR(1)$ process with persistence 0.979 and standard deviation of innovations of 0.0072.

We solve the dynamic stochastic model using nonlinear methods and evaluate welfare under the planner solution and under the competitive equilibrium.²¹ Each panel of Figure 1 plots the “compensating variation” in the tradition of Lucas (1983), namely the percentage points of consumption required to make the representative household indifferent between the Pareto optimum and the competitive equilibrium. This is how much a household living in the CE world would be willing to pay in order to have a benevolent planner make the entry decisions. For each of the three preference specifications that imply inefficiency, this measure of inefficiency is plotted as a function of the relevant parameter: ξ for general C.E.S., α/f_E for exponential, and σ/f_E for translog.

In the *general C.E.S.* case, in which only the “static” distortion operates, the welfare loss is reassuringly zero in the case corresponding to C.E.S.-DS, namely $\xi = 1/(\theta - 1) = 0.357$. Otherwise, the welfare loss is sizable. To take two rather extreme examples, when the benefit of variety ξ is equal to half the net markup, the loss is about 3.5 percentage points of consumption, while when the benefit of variety is twice as large as the net markup, the loss is around 8 percentage points of consumption.

Under *translog preferences*, the relevant parameter governing *both* the steady-state desired markup and the benefit of variety is σ (recall that for translog preferences for instance, $\mu(N) = 1 + (\sigma N)^{-1}$ and $\epsilon(N) = 1/2\sigma N$); similarly, under exponential preferences the parameter α plays the same role. Furthermore, because both the steady-state markup and the benefit of variety depend on the number of firms, the value of the sunk entry cost f_E now matters. To understand the role of these parameters in shaping the welfare properties, notice that the steady-state number of firms

²¹The codes were developed by Pablo Winant, to whom we are grateful, and are available upon request.

under translog is (see also BGM, Appendix A):

$$N^{\text{translog}} = \frac{-\delta + \sqrt{\delta^2 + 4\frac{\sigma}{f_E}L(r+\delta)(1-\delta)}}{2\sigma(r+\delta)}.$$

Intuitively, the steady-state number of firms is decreasing with the level of regulation, i.e., with the sunk entry cost f_E . It follows that the *elasticity of substitution* between goods is:

$$1 + \sigma N^{\text{translog}} = 1 + \frac{-\delta + \sqrt{\delta^2 + 4\frac{\sigma}{f_E}L(r+\delta)(1-\delta)}}{2(r+\delta)}. \quad (7)$$

Evidence on the elasticity of substitution between goods can therefore only be used to pin down the ratio σ/f_E (given the values of L, r and δ), but not the individual values of σ and f_E ; in other words, σ and f_E individually affect the scale of the economy (the steady-state number of firms), but only their ratio affects the elasticity of substitution and the steady-state markup. Therefore, in the remainder of the paper, we treat σ/f_E as the relevant parameter under translog (by the same reasoning, the relevant parameter under exponential preferences is α/f_E).²²

The second panel of Figure 1 plots the case of *exponential preferences*, where only the dynamic distortion operates. The welfare loss is rather small: at most 0.07 percentage points when the distortion is at its largest (α/f_E close to 0). This illustrates that the dynamic distortion, by itself, is likely to be quantitatively insignificant.²³

The third panel plots the *translog preference specification* shown by BGM to fit business cycle facts pertaining to entry, markup, and profit dynamics. For these preferences, both distortions combine to generate significant welfare losses. For the value of $\sigma/f_E = 0.35$ calibrated by BGM to fit data moments with this model, the welfare cost associated with inefficient entry and variety is about 2 percentage points—most of it being due to the static entry distortion. The size of the distortion is decreasing in σ/f_E , because the elasticity of substitution is increasing in that parameter. It follows that the gap between the net markup and the benefit of variety, which

²²Note also that steady-state ratios that could be conceived as allowing to calibrate the sunk cost independently are also a function of the steady-state markup, hence of the elasticity of substitution and finally only of the ratio σ/f_E . Thus, the share of labor used to pay for the sunk cost into total labor is $L_E/L = f_E N_E/L = \delta f_E N / [(1-\delta)L]$, which, using the expression for N , is only a function of σ/f_E ; under the baseline calibration with $\sigma/f_E = 0.354$ and $L = 1/3$, we have $L_E/L = 0.325$. The output value of resources used for the sunk cost as a share of GDP is $wL_E/Y = Q/(1+Q)$, where $Q = \delta(\mu-1)/[\mu(r+\delta)]$, which is again only a function of σ/f_E . Under the baseline calibration, $wL_E/Y = 0.16$.

²³The results of Bilbiie, Fujiwara, and Ghironi (2014) further illustrate this finding: they show that a Ramsey planner does not use a costly, distortionary instrument (inflation) over the cycle in order to correct this dynamic distortion: in other words, the distortion itself is “small.”

governs the static distortion, is decreasing in σ/f_E —thus it is increasing with the regulation cost f_E . Intuitively, more regulation (higher f_E) leads *ceteris paribus* to a lower number of firms in steady-state, and hence higher desired markups. Since the benefit of variety is half the desired (net) markup, it also increases proportionally.

Evidence on entry costs (see Ebell and Haefke, 2009) points to large heterogeneity across countries: while it “costs” 8.6 days or 1 percent of annual per capita GDP to start a firm in the United States (with similar numbers for Australia, the UK, and Scandinavian countries), the costs are an order of magnitude higher in most continental European countries (at the extreme, a whopping 84.5 days in Spain and 48 percent of annual per capita GDP for Greece). The preference parameter σ is less likely to vary as much across countries. Thus, the model identifies the degree of entry regulation as a key determinant of the inefficiencies pertaining to entry and product creation.²⁴

Our model also has stark implications regarding the optimality of deregulation. If preferences take the general C.E.S. form, deregulation, by promoting entry, is only optimal if the CE does not feature enough entry, that is if the benefit of variety is higher than the steady-state markup. In the opposite case, there is too much entry and more regulation is in fact optimal.

3.3 Optimal Fiscal Policy

Fiscal policies can implement the Pareto optimal PE as a competitive equilibrium (or alternatively, can decentralize the planning optimum) when the CE is otherwise inefficient. We assume that lump-sum instruments are available to finance whatever taxation scheme ensures implementation of the optimum, and give one example of such a taxation scheme here. Several recent studies use our model to study optimal taxation in second-best environments (Bilbiie, Fujiwara and Ghironi, 2014; Chugh and Ghironi, 2012; Lewis and Winkler, 2015).

Since in the competitive equilibrium there are two distortions generating inefficiencies, it is natural to look at an implementation scheme that uses two tax instruments. One intuitive example consists of a combination of consumption taxes (or VAT) and profit or dividend taxes. In particular, assume that τ_t^C is a proportional tax on the consumption good, and τ_t^D the rate of dividend (profits) proportional taxation. It is immediate to show that the Euler equation in the competitive

²⁴In a model with nominal rigidities, this further implies that the degree of regulation is a key determinant of the optimal inflation rate, as noted by Bilbiie, Fujiwara, and Ghironi (2014),

equilibrium becomes, under this taxation scheme,

$$\begin{aligned}
& f_{E,t\rho}(N_t) U'(C_t) \\
&= \beta (1 - \delta) E_t \left\{ U'(C_{t+1}) \frac{1 + \tau_t^C}{1 + \tau_{t+1}^C} \frac{\mu(N_t)}{\mu(N_{t+1})} \left[f_{E,t+1\rho}(N_{t+1}) + \frac{C_{t+1}}{N_{t+1}} (1 - \tau_{t+1}^D) (\mu(N_{t+1}) - 1) \right] \right\}.
\end{aligned} \tag{8}$$

Direct comparison with the Euler equation under the Pareto optimum delivers the state-contingent paths for the optimal taxes:

$$\begin{aligned}
1 - \tau_t^{D*} &= \frac{\epsilon(N_t)}{\mu(N_t) - 1} \\
\frac{1 + \tau_t^{C*}}{1 + \tau_{t+1}^{C*}} &= \frac{\mu(N_{t+1})}{\mu(N_t)}
\end{aligned}$$

The dividend tax corrects the static distortion, bringing the entry incentives in line with the benefit of variety, within the period. Intuitively, when the benefit of variety is lower than the net markup, $\epsilon(N_t) < \mu(N_t) - 1$, it is optimal to tax profits because the competitive equilibrium features too much entry (the market provides “too much” incentive to enter).

The VAT tax corrects the dynamic distortion by providing the “right” intertemporal price for consumption: intuitively, it is optimal to increase future VAT relative to present ($\tau_{t+1}^{C*} > \tau_t^{C*}$) when entry is “too low” today, inducing higher markups today than tomorrow ($N_t < N_{t+1} \rightarrow \mu(N_t) > \mu(N_{t+1})$). This makes the consumption good relatively more expensive today; optimal policy corrects this intertemporal markup misalignment by making today’s consumption relatively less expensive.

Finally, we note that the implementation of the Pareto optimum with a single tax instrument is generally not possible, because there are two distortions to address. In particular, focusing on the taxes considered above, a dividend tax by itself does not affect the dynamic distortion and hence cannot provide the right intertemporal price; whereas a VAT tax does not affect the static distortion, and cannot provide the right within-period entry incentives. This “impossibility” result generalizes to a large menu of taxes, such as sales or entry subsidies—even though appropriate combinations of such instruments can also lead to implementation of the social optimum.

4 Endogenous Labor Supply and the Importance of Monopoly Profits

As an example of the inefficiency associated with the lack of markup synchronization across goods, we now introduce an endogenous labor/leisure choice. With a perfectly competitive labor market, this provides a specific example of the inefficiencies implied by differences in markups across the items that bring utility to households. The only modification with respect to the model of Section 2 is that households now choose how much labor effort to supply in every period. Consequently, the period utility function features an additional term measuring the disutility of hours worked. We specify a general, non-separable utility function over consumption and effort: $U(C_t, L_t)$ and employ standard assumptions on its partial derivatives ensuring that the marginal utility of consumption is positive, $U_C > 0$, the marginal utility of effort is negative $U_L < 0$, and utility is concave: $U_{CC} \leq 0$; $U_{LL} \leq 0$ and $U_{CC}U_{LL} - (U_{CL})^2 \geq 0$.²⁵

As we show in the Appendix, optimal labor supply in the CE and PE is determined by the equations that govern intratemporal substitution between consumption and leisure. These are, respectively:

$$-U_L(C_t, L_t)/U_C(C_t, L_t) = Z_t \rho(N_t) / \mu(N_t), \quad (9)$$

in the CE, and

$$-U_L(C_t, L_t)/U_C(C_t, L_t) = Z_t \rho(N_t), \quad (10)$$

in the PE.

Except for the change in notation for the marginal utility of consumption and the fact that L is now time-varying, the only difference (with respect to the fixed-labor case) between the CE and PE is captured in equations (9) and (10). At the Pareto optimum, the marginal rate of substitution between consumption and leisure ($-U_L(C_t, L_t)/U_C(C_t, L_t)$) is equal to the marginal rate at which hours and consumption can be transformed into each other ($Z_t \rho(N_t)$). In the competitive equilibrium this is no longer the case. There is a wedge between these two objects equal to the reciprocal of the gross price markup, $(\mu(N_t))^{-1}$. Since consumption goods are priced at a markup while leisure is not, demand for the latter is sub-optimally high (hence, hours worked and consumption are sub-optimally low). Clearly, this distortion is independent of those emphasized in Theorem 1 (even if preferences were C.E.S.-DS, a wedge equal to $(\theta - 1)/\theta$ would still exist, and the CE would be inefficient). As we shall see below, taxing leisure at a rate equal to the net markup in the pricing

²⁵Note that an utility function that is separable in consumption and effort occurs as a special case when $U_{CL} (= U_{LC}) = 0$.

of goods removes this distortion by ensuring effective markup synchronization across arguments of the utility function.²⁶

4.1 A Labor Subsidy versus a Revenue Subsidy

Suppose the government subsidizes labor at the rate τ_t^L , financing this policy with lump-sum taxes on household income. Combining the first-order condition for the household's optimal choice of labor supply with the wage schedule $w_t = Z_t \rho(N_t) / \mu(N_t)$ now yields:

$$-U_L(C_t, L_t) / U_C(C_t, L_t) = (1 + \tau_t^L) Z_t \rho(N_t) / \mu(N_t).$$

Comparing this equation to (10) shows that a rate of taxation of leisure equal to the net markup of price over marginal cost,

$$1 + \tau_t^{L*} = \mu(N_t), \tag{11}$$

restores efficiency of the market equilibrium. This policy ensures synchronization of markups, consistent with intuitions that can be traced back to Lerner (1934) and Samuelson (1947). The optimal labor subsidy is countercyclical, since markups in this model are countercyclical ($\mu'(x) \leq 0$): Stronger incentives to work are used in periods/states with a low number of producers.

When product variety is exogenously fixed, this optimal labor subsidy is equivalent to a revenue subsidy that induces marginal cost pricing of consumption goods (again synchronizing relative prices between consumption and leisure) and financing this subsidy with a lump-sum tax on firm profits. This is another option to restore efficiency studied by virtually every paper addressing the possible distortions associated with monopoly ever since Robinson (1933, pp. 163-165).

However, this equivalence no longer holds in our framework with producer entry: A revenue subsidy financed with lump-sum taxation of firm profits would remove the wedge from equation (9), but no firm would find it profitable to enter (in the absence of an additional entry subsidy) since there would be no profit with which to cover the entry cost. While in the C.E.S.-DS case with elastic labor a sales subsidy restores the optimum when financed by lump-sum taxes on the consumer, this is a special case. When even a small fraction of the subsidy is financed by taxing the firm (as is implicitly or explicitly assumed in much of the literature), the optimum is no longer restored, as taxation of the firm affects the entry decision. In fact, in the C.E.S.-DS case, the

²⁶Thus, our results conform with the argument in Lerner (1934, p. 172) that "If the 'social' degree of monopoly is the same for all final products [including leisure] there is no monopolistic alteration from the optimum at all."

optimal split of financing a revenue subsidy between lump-sum taxation of consumer income versus firm profits requires *exactly zero* taxation of firm profits. We demonstrate this point formally by studying the effect of a policy inducing marginal cost pricing in the fully general case in Appendix E. This highlights once more that monopoly power *in itself* is not a distortion and should in fact be preserved if firm entry is subject to costs that cannot be entirely subsidized.

5 Conclusions

This paper contributes to the literature on the efficiency properties of models with monopolistic competition that can be traced back to at least Robinson (1933) and Lerner (1934). We studied the efficiency properties of a DSGE macroeconomic model with monopolistic competition and firm entry subject to sunk costs, a time-to-build lag, and exogenous risk of firm destruction.

Our main theoretical result is a theorem stating that, unless preferences for variety follow the knife-edge C.E.S. form studied by Dixit and Stiglitz (1977), the market equilibrium is inefficient because of two distortions: A static one, pertaining to the difference between the consumer surplus of a new variety, and the market incentive to create that variety; And a dynamic one, coming from the inefficiency of markup variations over time. A quantification of the welfare costs associated with these distortions in a calibrated version of the model reveals that the former is likely to be more relevant quantitatively: for a calibration with translog preferences that can match U.S. business cycle facts regarding markups, profits and entry, the welfare costs are very large: around 2 percentage points of consumption. Properly designed taxes can eliminate these costs by inducing markup synchronization across time and states, and aligning the consumer surplus and profit destruction effects of firm entry; one example we provide consists of a combination of VAT and dividend/profit taxes.

When labor supply is elastic, heterogeneity in markups across consumption and leisure introduces an additional distortion. Efficiency is then restored by subsidizing labor at a rate equal to the markup in the market for goods, thus removing the effect of markup heterogeneity on the competitive equilibrium. Our results highlight the importance of preserving the optimal amount of monopoly profits in economies in which firm entry is costly. Inducing marginal cost pricing restores efficiency only when the required sales subsidies are financed with an optimal split of lump-sum taxation between households and firms. With the Dixit-Stiglitz preferences that are popular in the literature, this requires zero lump-sum taxation of firm profits. Our findings thus caution against interpretations of statements in recent literature on the “distortionary” consequences of monopoly

power and the required remedies.

References

- [1] Adão, B., I. Correia, and P. Teles (2003): “Gaps and Triangles,” *Review of Economic Studies* 70: 699-713.
- [2] Benassy, J.-P. (1996): “Taste for Variety and Optimum Production Patterns in Monopolistic Competition,” *Economics Letters* 52: 41-47.
- [3] Bergin, P. R., and G. Corsetti (2008): “The Extensive Margin and Monetary Policy,” *Journal of Monetary Economics* 55: 1222-1237.
- [4] Bergin, P. R., and G. Corsetti (2014): “International Competitiveness and Monetary Policy,” *mimeo*, Cambridge University and University of California, Davis.
- [5] Bertolotti, P. and F. Etro (2015): “Preferences, Entry and Market Structure”, forthcoming *RAND Journal of Economics*
- [6] Bertolotti, P. and F. Etro (2016): “Monopolistic Competition when Income Matters”, forthcoming *The Economic Journal*
- [7] Bilbiie, F. O., I. Fujiwara, and F. Ghironi (2014): “Optimal Monetary Policy with Endogenous Entry and Product Variety,” *Journal of Monetary Economics* 64: 1-20.
- [8] Bilbiie, F. O., F. Ghironi, and M. J. Melitz (2008a): “Monetary Policy and Business Cycles with Endogenous Entry and Product Variety,” in Acemoglu, D., K. S. Rogoff, and M. Woodford, eds., *NBER Macroeconomics Annual 2007*, University of Chicago Press, Chicago, IL, pp. 299-353.
- [9] Bilbiie, F. O., F. Ghironi, and M. J. Melitz (2012): “Endogenous Entry, Product Variety, and Business Cycles,” *Journal of Political Economy*.
- [10] Blanchard, O. J., and N. Kiyotaki (1987): “Monopolistic Competition and the Effects of Aggregate Demand,” *American Economic Review* 77: 647-666.
- [11] Cacciatore, M., G. Fiori, and F. Ghironi (2016): “Market Deregulation and Optimal Monetary Policy in a Monetary Union,” *Journal of International Economics* 99: 120-137.

- [12] Cacciatore, M., and F. Ghironi (2012): “Trade, Unemployment, and Monetary Policy,” *mimeo*, HEC Montréal and University of Washington.
- [13] Chugh, S. K., and F. Ghironi (2015): “Optimal Fiscal Policy with Endogenous Product Variety,” CEPR DP 10674.
- [14] Colciago, A. and Etro, F. (2010): “Endogenous Market Structures and Business Cycles”, *The Economic Journal*
- [15] Colciago, A. (2015): “Endogenous Market Structures and Optimal Taxation,” *The Economic Journal*, forthcoming.
- [16] Cooke, D. (2015): “Optimal Monetary Policy with Endogenous Export Participation,” *Review of Economic Dynamics*, forthcoming.
- [17] Dixit, A. K., and J. E. Stiglitz (1977): “Monopolistic Competition and Optimum Product Diversity,” *American Economic Review* 67: 297-308.
- [18] Dhingra, S. and J. Morrow (2013), “Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity,” *Mimeo London School of Economics*
- [19] Epifani, P. and G. Gancia, (2011), “Trade, Markup Heterogeneity and Misallocations,” *Journal of International Economics*
- [20] Etro, F., 2016: “The Ramsey Model with Imperfect Competition and General Preferences,” *Economic Letters*, Forthcoming
- [21] Etro, F., and L. Rossi (2015): “New-Keynesian Phillips Curve with Bertrand Competition and Endogenous Entry,” *Journal of Economic Dynamics and Control* 51: 318-340.
- [22] Faia, E. (2009): “Ramsey Monetary Policy with Labor Market Frictions,” *Journal of Monetary Economics* 56: 570-581.
- [23] Faia, E. (2012): “Oligopolistic Competition and Optimal Monetary Policy,” *Journal of Economic Dynamics and Control* 36: 1760-1774.
- [24] Feenstra, R. C. (2003): “A Homothetic Utility Function for Monopolistic Competition Models, without Constant Price Elasticity,” *Economics Letters* 78: 79-86.

- [25] Grossman, G. M., and E. Helpman (1991): *Innovation and Growth in the Global Economy*, MIT Press, Cambridge, MA.
- [26] Hopenhayn, H. (1992): “Entry, Exit, and Firm Dynamics in Long Run Equilibrium,” *Econometrica* 60: 1127–1150.
- [27] Khan, A., R. G. King, and A. L. Wolman (2003): “Optimal Monetary Policy,” *Review of Economic Studies* 70: 825-860.
- [28] Kim, J. (2004): “What Determines Aggregate Returns to Scale?” *Journal of Economic Dynamics and Control* 28: 1577-1594.
- [29] Lerner, A. (1934): “The Concept Monopoly and the Measurement of Monopoly Power,” *Review of Economic Studies* 2: 157-175.
- [30] Lewis, V. (2013): “Optimal Monetary Policy and Firm Entry,” *Macroeconomic Dynamics* 17: 1687-1710.
- [31] Lewis, V., and R. Winkler (2015): “Product Diversity, Market Structure, and Optimal Taxation,” *Economic Inquiry* 53: 979-1003.
- [32] Mankiw, N. G., and M. D. Whinston (1986): “Free Entry and Social Inefficiency,” *RAND Journal of Economics* 17: 48-58.
- [33] Melitz, M. J. (2003): “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica* 71: 1695-1725.
- [34] Opp, M., C. Parlour and J. Walden, (2014) “Markup cycles, dynamic misallocation, and amplification,” *Journal of Economic Theory*, 154, 126-161.
- [35] Robinson, J. (1933): *The Economics of Imperfect Competition*, Cambridge University Press, Cambridge, UK.
- [36] Samuelson, P. (1947): *Foundations of Economic Analysis*, Harvard University Press, Cambridge, MA.
- [37] Spence, M. (1976): “Product Selection, Fixed Costs, and Monopolistic Competition,” *Review of Economic Studies* 43: 217-235.

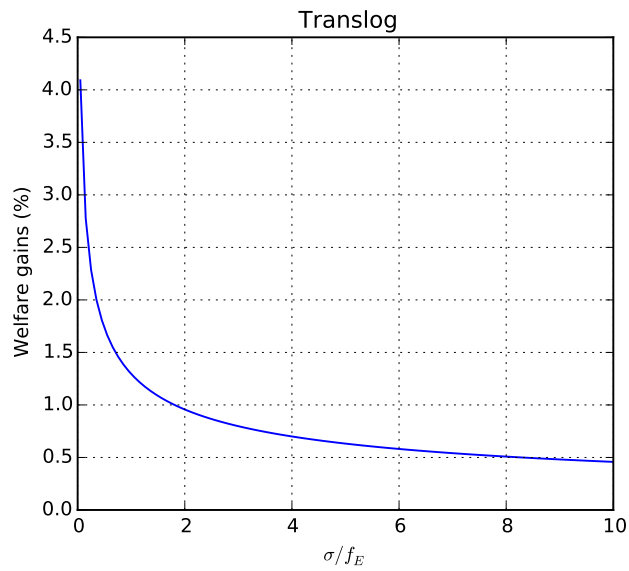
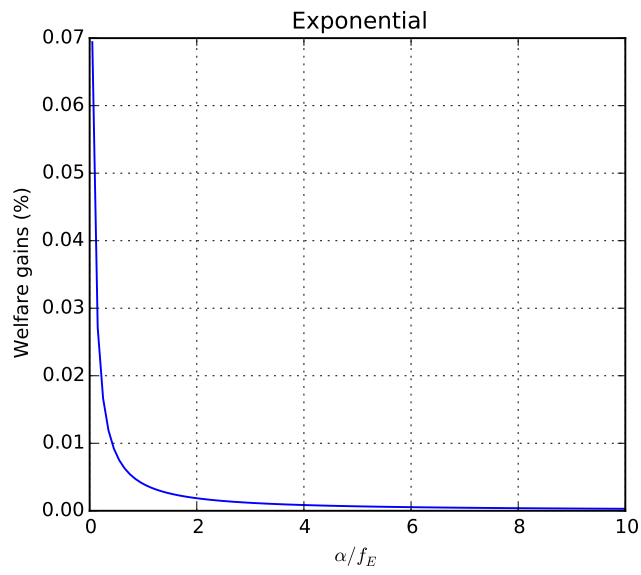
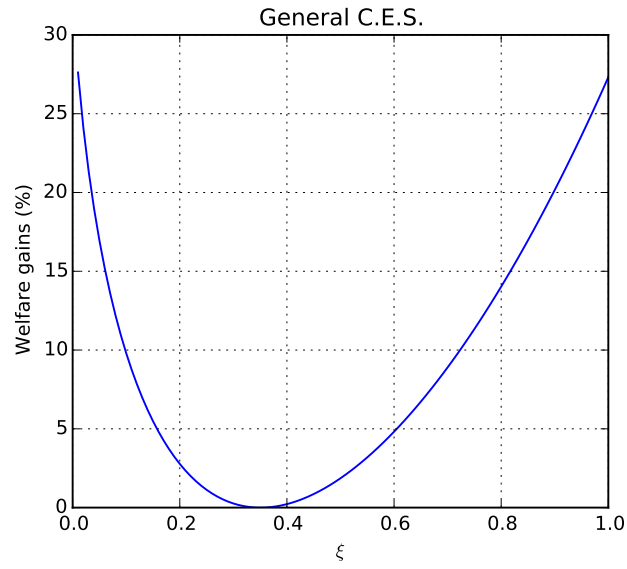


Figure 1: Efficiency Gains Relative to Competitive Equilibrium

Appendix

A Homothetic Consumption Preferences

Consider an arbitrary set of homothetic preferences over a continuum of goods Ω . Let $p(\omega)$ and $c(\omega)$ denote the prices and consumption level (quantity) of an individual good $\omega \in \Omega$. These preferences are uniquely represented by a price index function $P \equiv h(\mathbf{p})$, $\mathbf{p} \equiv [p(\omega)]_{\omega \in \Omega}$, such that the optimal expenditure function is given by PC , where C is the consumption index (the utility level attained for a monotonic transformation of the utility function that is homogeneous of degree 1). Any function $h(\mathbf{p})$ that is non-negative, non-decreasing, homogeneous of degree 1, and concave, uniquely represents a set of homothetic preferences. Using the conventional notation for quantities with a continuum of goods as flow values, the derived Marshallian demand for any variety ω is then given by:

$$c(\omega)d\omega = C \frac{\partial P}{\partial p(\omega)}.$$

B No Option Value of Waiting to Enter

Let the option value of waiting to enter for firm ω be $\Lambda_t(\omega) \geq 0$. In all periods t , $\Lambda_t(\omega) = \max[v_t(\omega) - w_t f_{E,t}/Z_t, \beta \Lambda_{t+1}(\omega)]$, where the first term is the payoff of undertaking the investment and the second term is the discounted payoff of waiting. If firms are identical (there is no idiosyncratic uncertainty) and exit is exogenous (uncertainty related to firm death is also aggregate), this becomes: $\Lambda_t = \max[v_t - w_t f_{E,t}/Z_t, \beta \Lambda_{t+1}]$. Because of free entry, the first term is always zero, so the option value obeys: $\Lambda_t = \beta \Lambda_{t+1}$. This is a contraction mapping because of discounting, and by forward iteration, under the assumption $\lim_{T \rightarrow \infty} \beta^T \Lambda_{t+T} = 0$ (i.e., there is a zero value of waiting when reaching the terminal period), the only stable solution for the option value is $\Lambda_t = 0$.

C Derivations for Competitive Equilibrium and Planner Problem

The **competitive equilibrium** is summarized by the following table:²⁷

²⁷The labor market equilibrium condition is redundant once the variety effect equation is included in the system in Table 2.

Table 2. Model Summary

Pricing	$\rho_t = \mu_t \frac{w_t}{Z_t}$
Variety effect	$\rho_t = \rho(N_t)$
Markup	$\mu_t = \mu(N_t)$
Profits	$d_t = \left(1 - \frac{1}{\mu_t}\right) \frac{C_t}{N_t}$
Free entry	$v_t = w_t \frac{f_{E,t}}{Z_t}$
Number of firms	$N_t = (1 - \delta)(N_{t-1} + N_{E,t-1})$
Euler equation	$v_t = \beta(1 - \delta) E_t \left[\frac{U'(C_{t+1})}{U'(C_t)} (v_{t+1} + d_{t+1}) \right]$
Aggregate accounting	$C_t + N_{E,t} v_t = w_t L + N_t d_t$

We can reduce the system in Table 2 to a system of two equations in two variables, N_t and C_t . To see this, write firm value as a function of the endogenous state N_t and the exogenous state $f_{E,t}$ by combining free entry, pricing, variety, and markup equations:

$$v_t = f_{E,t} \frac{\rho(N_t)}{\mu(N_t)}. \quad (12)$$

Substituting this, together with the profits' definition, in the Euler equation, we obtain (3) in text.

The first-order condition for the **planner's problem** (5) is:

$$\begin{aligned} & U'(C_t) Z_t \rho(N_t) \frac{1}{1 - \delta} \frac{f_{E,t}}{Z_t} \\ & = \beta E_t \left\{ U'(C_{t+1}) Z_{t+1} \rho'(N_{t+1}) \left[L - \frac{1}{(1 - \delta)} \frac{f_{E,t+1}}{Z_{t+1}} N_{t+2} + \frac{f_{E,t+1}}{Z_{t+1}} N_{t+1} + \frac{f_{E,t+1}}{Z_{t+1}} \frac{\rho(N_{t+1})}{\rho'(N_{t+1})} \right] \right\}. \end{aligned}$$

The term in square brackets in the right-hand side of this equation is:

$$L - \frac{1}{(1 - \delta)} \frac{f_{E,t+1}}{Z_{t+1}} N_{t+2} + \frac{f_{E,t+1}}{Z_{t+1}} N_{t+1} + \frac{f_{E,t+1}}{Z_{t+1}} \frac{\rho(N_{t+1})}{\rho'(N_{t+1})} = L_{t+1}^C + \frac{f_{E,t+1}}{Z_{t+1}} \frac{\rho(N_{t+1})}{\rho'(N_{t+1})}.$$

Hence, the first-order condition becomes:

$$U'(C_t) \rho(N_t) f_{E,t} = \beta(1 - \delta) E_t \left\{ U'(C_{t+1}) Z_{t+1} \rho'(N_{t+1}) \left[L_{t+1}^C + \frac{f_{E,t+1}}{Z_{t+1}} \frac{\rho(N_{t+1})}{\rho'(N_{t+1})} \right] \right\},$$

leading to (6).

D Proof of Theorem 1

Sufficiency ('if') is directly verified by plugging conditions (i) and (ii) into (3) and (6).

Necessity ('only if') requires that, whenever both (3) and (6) are satisfied, then (i) and (ii) hold. We prove this by contradiction. We first look at the simpler perfect-foresight case (where we can drop the expectations operator) and then extend our proof to the stochastic case.

Suppose by reductio ad absurdum that there exists a 2-tuple $\{C_t, N_{t+1}\}$ that is both a CE and a PE, with $\mu(N_t) \neq \mu(N_{t+1})$ or $\epsilon(x) \neq \mu(x) - 1$ or both. We examine each case separately.

(A) $\mu(N_t) \neq \mu(N_{t+1})$ and $\epsilon(x) = \mu(x) - 1$:

Substituting $\epsilon(N_{t+1})$ in the planner's Euler equation, $\mu(N_t) \neq \mu(N_{t+1})$ and $\epsilon(x) = \mu(x) - 1$ imply that

$$U'(C_{t+1}) f_{E,t+1} \rho(N_{t+1}) \left[\frac{\mu(N_{t+1}) - \mu(N_t)}{\mu(N_{t+1})} \right] = U'(C_{t+1}) \frac{C_{t+1}}{N_{t+1}} (\mu(N_{t+1}) - \mu(N_t)) \left(\frac{1}{\mu(N_{t+1})} - 1 \right). \quad (13)$$

After further simplification, using $\mu(N_t) \neq \mu(N_{t+1})$ and $U'(C_{t+1}) \neq 0$, this yields:

$$1 - \mu(N_{t+1}) = \frac{f_{E,t+1} \rho(N_{t+1}) N_{t+1}}{C_{t+1}} \leq 0, \text{ since } \mu(N_{t+1}) \geq 1. \quad (14)$$

But this is a contradiction, since all terms on the right-hand side are strictly positive.

For the stochastic case:

$$E_t \left\{ U'(C_{t+1}) \frac{\mu(N_{t+1}) - \mu(N_t)}{\mu(N_{t+1})} \left[f_{E,t+1} \rho(N_{t+1}) + \frac{C_{t+1}}{N_{t+1}} (\mu(N_{t+1}) - 1) \right] \right\} = 0,$$

which is a contradiction since $\mu(N_t) \neq \mu(N_{t+1})$, $U'(C_{t+1}) \neq 0$, and the term in square brackets is strictly greater than zero ($\mu(N_{t+1}) \geq 1$).

(B) $\mu(N_t) = \mu(N_{t+1}) = \mu$ and $\epsilon(x) \neq \mu(x) - 1$:

Using Theorem 1, $\mu(N_t) = \mu(N_{t+1}) = \mu$ and $\epsilon(x) \neq \mu(x) - 1$ imply that

$$U'(C_{t+1}) \frac{C_{t+1}}{N_{t+1}} [\epsilon(N_{t+1}) - (\mu - 1)] = 0. \quad (15)$$

This would further imply that either $U'(C_{t+1}) = 0$ or $C_{t+1} = 0$ or $\epsilon(N_{t+1}) = \mu - 1$, which are all contradictions.

(C) $\mu(N_t) \neq \mu(N_{t+1})$ and $\epsilon(x) \neq \mu(x) - 1$:

In this case, a steady state is still defined by $N_t = N_{t+1}$, so $\mu(N_t) = \mu(N_{t+1}) = \mu(N)$ in steady

state. If the CE and PE equilibria are identical, then (evaluating the Euler equations at the steady state) $\epsilon(N) = \mu(N) - 1$, which contradicts the assumption $\epsilon(x) \neq \mu(x) - 1$. This holds for the stochastic case too (note that the same argument can be used in part (B), including the stochastic case).

E Endogenous Labor Supply

From inspection of Table 2, the two modifications to the CE conditions implied by endogeneity of labor supply are that L in the aggregate accounting identity now features a time index t , and the marginal utility of consumption, now denoted by $U_C(C_t, L_t)$, depends on hours worked. The new variable L_t is then determined in standard fashion by adding to the equilibrium conditions the intratemporal first-order condition of the household governing the choice of labor effort:

$$-U_L(C_t, L_t) = w_t U_C(C_t, L_t).$$

Combining this with the wage schedule $w_t = Z_t \rho(N_t) / \mu(N_t)$, which holds also with endogenous labor supply, yields the condition:

$$-U_L(C_t, L_t) / U_C(C_t, L_t) = Z_t \rho(N_t) / \mu(N_t).$$

This, in turn, can be solved to obtain hours worked as a function of consumption, the number of firms, and productivity.

The PE when labor supply is endogenous is found by solving:

$$\max_{\{L_s, N_{s+1}\}_{s=t}^{\infty}} E_t \sum_{s=t}^{\infty} \beta^{s-t} U \left[Z_s \rho(N_s) \left(L_s - \frac{1}{(1-\delta)} \frac{f_{E,s}}{Z_s} N_{s+1} + \frac{f_{E,s}}{Z_s} N_s \right), L_s \right].$$

The Euler equation for the planner's optimal choice of N_{t+1} and the law of motion for the number of firms are identical to the case of fixed labor supply, up to the addition of a time index for labor and to recognizing the dependence of the marginal utility of consumption upon the level of effort. The additional intratemporal condition for the planning optimum is:

$$-U_L(C_t, L_t) / U_C(C_t, L_t) = Z_t \rho(N_t).$$

E.1 Labor or Revenue Subsidy

This Appendix contains the details of the analysis underlying our discussion of optimal policy under elastic labor. Suppose the planner subsidizes or taxes sales at rate τ_t and each firm is taxed lump-sum T_t^F for a possibly time-varying fraction γ_t of this expenditure. The following proposition holds.

Proposition 1 *A sales subsidy that induces marginal cost pricing, financed by lump-sum taxes on both firms and consumers, restores efficiency of the competitive equilibrium if and only if the fraction of taxes paid by the firm, γ_t , satisfies:*

$$\gamma_t^* = 1 - \frac{\epsilon(N_t)}{\mu(N_t) - 1}.$$

Proof. The profit function becomes: $d_t = (1 + \tau_t) \rho_t y_t - w_t l_t - T_t^F$ or (under optimal pricing $\rho_t = \frac{\mu(N_t) w_t}{1 + \tau_t Z_t}$), $d_t = (1 + \tau_t) \rho_t y_t - \frac{(1 + \tau_t)}{\mu(N_t)} \rho_t y_t - T_t^F$. Balanced budget implies that total taxes are $\tau_t \rho_t N_t y_t$, so the fraction of taxes paid by a firm is $T_t^F = \gamma_t \tau_t \rho_t y_t$. It follows that profits are finally given by

$$d_t = \left[1 + (1 - \gamma_t) \tau_t - \frac{1 + \tau_t}{\mu(N_t)} \right] \rho_t y_t = \left[1 + (1 - \gamma_t) \tau_t - \frac{1 + \tau_t}{\mu(N_t)} \right] \frac{C_t}{N_t}.$$

To eliminate the wedge between the marginal rate of substitution and the marginal rate of transformation between consumption and leisure, we know that the optimal value of τ_t is such that $1 + \tau_t = \mu(N_t)$, implying $d_t = (1 - \gamma_t) (\mu(N_t) - 1) \frac{C_t}{N_t}$. The value of a firm is given by $v_t = w_t \frac{f_{E,t}}{Z_t} = \rho(N_t) f_{E,t}$. Substituting these expressions in the CE Euler equation for shares yields:

$$U_C(C_t, L_t) \rho(N_t) f_{E,t} = \beta (1 - \delta) E_t \left\{ U_C(C_{t+1}, L_{t+1}) \left[f_{E,t+1} \rho(N_{t+1}) + (1 - \gamma_{t+1}) (\mu(N_{t+1}) - 1) \frac{C_{t+1}}{N_{t+1}} \right] \right\}.$$

Comparing this with the planner's Euler equation (6) written for the case of endogenous labor (and hence replacing $U'(C)$ with $U_C(C, L)$), we obtain the optimal fraction of taxes paid by the firm, γ_t^* , as in the Proposition. ■

A policy inducing marginal cost pricing can restore efficiency only if an optimal division of lump-sum taxes between consumers and firms is also ensured. Recall that for C.E.S.-DS preferences (the most common case in the literature) $\epsilon = \mu - 1$. It follows that efficiency is restored by inducing marginal cost pricing if and only if $\gamma_t = 0$, i.e., if all the subsidy for firm sales is paid for by consumers, and none by firms. Otherwise, taxation of firms affects the relationship between firm

profits and total sales, and therefore affects the entry decision. In the extreme case where all of the subsidy is financed by lump-sum taxes on firms, $\gamma_t = 1$, it is clear that equilibrium firm profits become zero, and no firm will have incentives to enter. Clearly, γ_t^* is non-zero only when the markup and benefit from variety are not aligned, $\epsilon(x) \neq \mu(x) - 1$, as for Benassy or translog preferences. Note that, for the latter, the optimal division of taxes between consumers and firms is an equal split (since $\epsilon(x) = (\mu(x) - 1)/2$).