Monopoly Power and Endogenous Variety in Dynamic Stochastic General Equilibrium: Distortions and Remedies*

Florin O. Bilbiie†
Oxford University, Nuffield College, and CEP-LSE

Fabio Ghironi‡
Boston College, EABCN, and NBER

Marc J. Melitz§
Harvard University, Princeton University, CEPR, and NBER

June 12, 2006
PRELIMINARY
Comments welcome

Abstract

We study the efficiency properties of a dynamic, stochastic, general equilibrium, macroeconomic model with monopolistic competition and firm entry subject to sunk costs, a time-to-build lag, and exogenous risk of firm destruction. Under inelastic labor supply and linearity of production in labor, the market economy is efficient if and only if symmetric, homothetic preferences are of the C.E.S. form studied by Dixit and Stiglitz (1977). Otherwise, efficiency is restored by properly designed sales, entry, or asset trade subsidies (or taxes) that induce markup synchronization across time and states, and align the consumer surplus and profit destruction effects of firm entry. 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. 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 the optimal split of lump-sum taxation between households and firms.

JEL Codes: D42; E20; H21; L10

Keywords: Decentralization; Efficiency; Entry; Monopoly power; Product creation; Variety

*For helpful comments, we thank seminar and conference participants at Catholic University Lisbon, Chicago GSB, Cleveland Fed, Cornell, CSEF-IGIER Symposium on Economics and Institutions, ECB, ESSIM 2005, London Business School, MIT, NBER Summer Institute 2005, New York Fed, Northeastern, Princeton, SED 2005, Swiss National Bank, and University of Connecticut. We are grateful to Nicholas Sim and Pinar Uysal for excellent research assistance. Remaining errors are our responsibility. Bilbiie thanks the ECB for hospitality in the summer of 2004. Ghironi and Melitz thank the NSF for financial support through a grant to the NBER.

†Nuffield College, New Road, Oxford, OX1 1NF, UK or florin.bilbiie@nuffield.oxford.ac.uk; URL: http://www.nuff.ox.ac.uk/Users/Bilbiie/index.htm

‡Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467-3859, U.S.A. or Fabio.Ghironi@bc.edu. URL: http://fmwww.bc.edu/ec/Ghironi.php.

§Department of Economics, Harvard University, Littauer Center 120, Cambridge, MA 02138, U.S.A. or mmelitz@harvard.edu. URL: http://post.economics.harvard.edu/faculty/melitz/melitz.html.

1 Introduction

What are the consequences of monopoly power for efficiency of business cycle fluctuations and new product creation? If market power results in inefficiency, 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 (2005 – henceforth, BGM). Specifically, we compare the competitive and planner equilibria, asking whether the market solution provides for efficient responses to exogenous shocks and the optimum amount of product variety when product creation is subject to sunk costs, a time-to-build lag, and exogenous risk of firm destruction. We then analyze 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.

In BGM, we argued that creation of new products is an important mechanism for business cycle propagation. Endogenous product creation subject to sunk entry costs provides a new mechanism of propagation and amplification of shocks (for instance, to technology) and makes it possible to reconcile theory with stylized facts on firm entry, product creation, and the cyclicality of profits and markups. By studying the efficiency properties of our DSGE 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), Judd (1985), and Grossman and Helpman (1991), among others.[1]

Under assumptions of inelastic labor supply and linearity of production in labor, our main result is that a monopolistically competitive market provides for socially efficient economic fluctuations and product entry (that is, the competitive and planner equilibria coincide) when consumers have homothetic preferences exhibiting love for variety if and only if preferences are such that: (i) markups are synchronized over time and across states; and (ii) the benefit of variety in elasticity form is functionally identical to the net markup in the pricing of goods. That is, efficiency requires that preferences be of the C.E.S. form originally studied by Dixit and Stiglitz (1977).

[1] See also Mankiw and Whinston (1986) and Benassy (1996). Kim (2004) also studies efficiency in his DSGE model with an endogenous number of firms. However, the entry decision is not fully endogenous in his model, and increasing returns can generate indeterminacy, whereas the equilibrium is always locally determinate in the log-linearized version of our model studied in BGM.

We identify two mechanisms that ensure this result. First, despite prices being above marginal cost, since price adjustment is frictionless and producers are symmetric, markups in the pricing of all goods that bring utility to the consumer are synchronized. While this is also true in a model with monopolistic competition and a fixed number of firms when labor supply is inelastic, our model with entry has an important additional implication. Namely, although we let one factor of aggregate output production (the number of firms) vary subject to a sunk entry cost, a time-to-build lag, and exogenous firm destruction, efficiency still holds. The resulting number of firms is socially optimal due to the key distinguishing feature of our framework – the entry mechanism based on C.E.S. preferences.

Moreover, efficiency also requires that markups be synchronized across goods, as we show by making the labor supply choice endogenous but assuming that, differently from the consumption good, leisure is not subject to a markup. Efficiency no longer holds when labor supply is endogenous. However, the relevant distortion is not the existence of a markup in the market for goods in and of itself, but heterogeneity in markups between the "goods" the consumer cares about: consumption goods and leisure (priced at "marginal cost" in a competitive labor market). It is this heterogeneity in markups that results in a wedge between marginal rates of substitution and transformation between consumption and leisure that distorts labor supply.

When the conditions above fail, and hence the market economy is inefficient, the policymaker can use a variety of distortionary fiscal instruments (in conjunction with lump-sum taxes or transfers) to ensure implementation of the first-best equilibrium. With inelastic labor supply, a properly designed sales subsidy can remove the effects of both intertemporal markup variation and non-synchronization of consumer surplus and profit destruction effects of firm entry. The same effect can be obtained with a proportional entry cost subsidy, a subsidy to net stock market trades, or a tax on gross trades.

When labor supply is elastic, 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 holds also 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] In fact, we show that inducing marginal cost pricing can

[2] We are implicitly assuming that the government is not contemporaneously subsidizing the entire amount of the

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 in Dixit and Stiglitz (1977).

Our results reiterate an argument made elsewhere in the literature that monopoly power in and of itself is not a distortion and show that, in the presence of entry subject to sunk costs, optimal monopoly profits should in fact be preserved. Indeed, while markup synchronization across time, states, and goods is still a necessary condition for efficiency, sufficiency requires that markups be aligned to the relatively higher level. Our findings thus caution against naive interpretations of statements in recent literature on the effects of a "monopolistic distortion" and the required remedies.

The structure of the paper is as follows. Section 2 describes the benchmark model with fixed labor supply and characterizes the competitive equilibrium. Section 3 studies the problem facing a social planner for our model economy. Section 4 states and proves our welfare theorem, and discusses the intuition for it. Section 5 extends the analysis to the case of endogenous labor supply. Section 6 studies optimal fiscal policies that implement the first-best allocation. Section 7 concludes.

2 Model: The Market Economy

Household Preferences

The economy is populated by a unit mass of atomistic households. All contracts and prices are written in nominal terms. Prices are flexible. Thus, we only solve for real variables in the model. However, as the composition of the consumption basket changes over time due to firm entry (affecting the definition of the consumption-based price index), we introduce money as a convenient unit of account for contracts. Money plays no other role in the economy. For this reason, we do not model the demand for cash currency, and resort to a cashless economy as in Woodford (2003).

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): $E_t\left(\sum_{s=t}^{\infty}\beta^{s-t}U\left(C_s\right)\right)$, where $\beta \in (0,1)$ is the subjective discount factor and $U\left(C\right)$ is a period utility function with the standard properties. At time t, the household consumes the basket of goods C_t, defined over a continuum of goods Ω. At any given time t, only

entry cost.

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 consumption index C_t and an associated welfare-based price 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.[3]

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

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

The benefit of additional product variety is described 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, $\zeta(N_t)$ and $\rho(N_t)$ completely characterize the effects of 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.

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.[4]

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 good. $f_{E,t}$ is exogenous and follows an $AR(1)$ process (in logarithms). There

[3] See the appendix for more details.

[4] Consistent with standard RBC theory, aggregate productivity Z_t affects all firms uniformly. We abstract from the more complex technology diffusion processes across firms of different vintages studied by Caballero and Hammour (1994) and Campbell (1998). We also do not address the growth effects of changes in product variety. Bils and Klenow (2001) document that these effects are empirically relevant for the U.S.

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.[5]

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 $\rho_t(\omega) \equiv p_t(\omega)/P_t = \mu_t w_t / Z_t$, where the markup is a function of the number of producers: $\mu_t = \mu(N_t) \equiv \zeta(N_t)/(\zeta(N_t) - 1)$. The firm's profit in units of consumption, returned to households as dividend, is $d_t(\omega) = d_t = \left(1 - \mu(N_t)^{-1}\right) Y_t^C / N_t$, where Y_t^C is total output of the consumption good and will in equilibrium be equal to total consumption demand C_t.

Preference Specifications and Markups

We consider three alternative preference specifications as special cases for illustrative purposes below. The first features a constant elasticity of substitution (C.E.S.) 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: $\mu(N_t) = \mu$, $\epsilon(N_t) = \epsilon$; and they are related by: $\epsilon = \mu - 1 = 1/(\theta - 1)$. The second specification is the C.E.S. variant introduced by Benassy (1996), which disentangles monopoly power (measured by the net markup $1/(\theta - 1)$) and consumer love for variety, captured by a parameter $\xi > 0$. With this specification, 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 specification uses the translog expenditure function proposed by Feenstra (2003), which introduces demand-side pricing complementarities. For this preference specification, the symmetric price elasticity of demand is $1 + \sigma N_t$, $\sigma > 0$: As N_t increases, goods become closer substitutes, and the elasticity of substitution increases. If goods are closer substitutes, then the markup $\mu(N_t)$ and the benefit of

[5] For 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.

additional varieties in elasticity form ($\epsilon(N_t)$) must decrease.[6] The change in $\epsilon(N_t)$ is only half the change in net markup generated by an increase in the number of producers. Table 1 contains the expressions for markup, relative price, and benefit of variety in elasticity form for each preference specification.

Table 1. Three frameworks

C.E.S.-DS	C.E.S.-Benassy	Translog
$\mu(N_t) = \mu = \frac{\theta}{\theta-1}$	$\mu_t = \mu = \frac{\theta}{\theta-1}$	$\mu(N_t) = 1 + \frac{1}{\sigma N_t}$
$\rho(N_t) = (N_t)^{\mu-1} = (N_t)^{\frac{1}{\theta-1}}$	$\rho(N_t) = (N_t)^{\xi}$	$\rho(N_t) = e^{-\frac{1}{2}\frac{\tilde{N}-N_t}{\sigma\tilde{N}N_t}}, \quad \tilde{N} \equiv Mass(\Omega)$
$\epsilon(N_t) = \mu - 1$	$\epsilon(N_t) = \xi$	$\epsilon(N_t) = \frac{1}{2\sigma N_t} = \frac{1}{2}(\mu(N_t) - 1)$

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

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

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. Finally, the timing of entry and production we have assumed implies 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

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

of producing firms represents the stock of capital of the economy. It is an endogenous state variable that behaves much like physical capital in the 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 implies necessarily that the relative price of each individual good increases $\rho'(N_t) > 0$. When there are more firms, households derive more welfare from spending a given nominal amount, *i.e.*, *ceteris paribus*, the price index decreases. It follows that the relative price of each individual good must rise. 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 cost and exit risk; *i.e.*, there are no option-value considerations pertaining to the entry decision.[7] This happens because all uncertainty in our model (including the "death" shock) is aggregate. 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 $\lim_{T\to\infty} \beta^T \Lambda_{t+T} = 0$ (*i.e.*, there is a zero value of waiting when reaching the terminal period), the only stable solution for the option value is $\Lambda_t = 0$.

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 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

[7]This is in contrast with models such as Caballero and Hammour (1995) and Campbell (1998).

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:

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

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 the equation for share holdings and absence of speculative bubbles yield the asset price solution in equation (1).[8]

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 of Campbell (1994) and many other studies, 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 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

[8] We omit the transversality condition that must be satisfied to ensure optimality.

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.[9]

The Competitive Equilibrium

The model with general homothetic preferences is summarized in Table 2.[10]

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 $f_{E,t}$

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

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

by combining free entry, pricing, variety, and markup equations:

$$v_t = f_{E,t} \frac{\rho(N_t)}{\mu(N_t)}. \tag{3}$$

By substitution of the equilibrium conditions in Table 2, the Euler equation for shares becomes:

$$f_{E,t}\rho(N_t) = \beta(1-\delta) E_t \left\{ \frac{U'(C_{t+1})}{U'(C_t)} \left[f_{E,t+1}\rho(N_{t+1}) \frac{\mu(N_t)}{\mu(N_{t+1})} + \frac{C_{t+1}}{N_{t+1}} \mu(N_t) \left(1 - \frac{1}{\mu(N_{t+1})}\right) \right] \right\}. \tag{4}$$

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:

$$N_{t+1} = (1-\delta)\left(N_t + \frac{Z_t L}{f_{E,t}} - \frac{C_t}{f_{E,t}\rho(N_t)}\right). \tag{5}$$

We are now in a position to define a competitive equilibrium of our economy.[11]

Definition 1: *A Competitive Equilibrium (CE) consists of a 2-tuple* $\{C_t, N_{t+1}\}$ *satisfying (4) and (5) for a given initial value* N_0 *and a transversality condition for investment in shares.*

The system of stochastic difference equations (4) and (5) has a unique stationary equilibrium under the following conditions. A steady state CE is defined by:

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

which, after eliminating C, leads to:

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

where $r \equiv (1-\beta)/\beta$.[12]

[11] It 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.

[12] Allowing 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.

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

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

this holds 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 planner equilibrium.

3 The Planning (Pareto) Optimum

Given the model of the previous section, we now study a hypothetical scenario in which a benevolent planner maximizes lifetime utility of the representative household by choosing quantities directly.

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:

$$\max_{\{L_s^C\}_{s=t}^{\infty}} E_t\left[\sum_{s=t}^{\infty}\beta^{s-t}U\left(Z_s\rho(N_s)L_s^C\right)\right],$$

$$s.t.\ N_{t+1}=(1-\delta)N_t+(1-\delta)\frac{\left(L-L_t^C\right)Z_t}{f_{E,t}},$$

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

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

The first-order condition for this problem 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

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

This equation, together with the dynamic constraint (5) (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 (5) and (6) for a given initial value N_0.*

The conditions for uniqueness of the stationary PE are similar to those for CE found in the previous section, where 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 (5) and (6) has a unique stationary equilibrium if and only if $\lim_{N\to 0} \epsilon(N) = \infty$, $\lim_{N\to\infty} \epsilon(N) = 0$, and $\epsilon'(N) \leq 0$.[13] The intuition for these uniqueness conditions is analogous to the one for the competitive equilibrium that more product variety lead to a "crowding in" of the product space and goods become closer substitutes

[13]Note 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 $\epsilon(N)$.

(with C.E.S. a limiting case). In the PE case, the condition is that there be decreasing returns to increased product variety (which is very similar to the condition that goods become closer substitutes). C.E.S. is again a limiting case where there are "constant returns" to increased product variety: Doubling product variety, holding spending constant, always increases welfare by the same percentage.

4 A Welfare Theorem

We are now in a position to state our main theorem, comparing the CE and PE equilibria obtained in the previous sections:

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 and markup functions are such that $\epsilon(x) = \mu(x) - 1$.

Proof. Sufficiency ('if') is directly verified by plugging conditions (i) and (ii) into (4) and (6).

Necessity ('only if') requires that if both (4) and (6) are satisfied, then necessarily (i) and (ii) hold. We prove this by contradiction. This proof is for the perfect-foresight case, where we can drop the expectations operator. The same procedure would apply 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, whereby $\mu(N_t) \neq \mu(N_{t+1})$ or $\epsilon(x) \neq \mu(x) - 1$ or both. We examine each case separately.

(A) $\mu(N_t) \neq \mu(N_{t+1})$ and $\epsilon(x) = \mu(x) - 1$ implies – substituting $\epsilon(N_{t+1})$ in the planner's Euler equation – that

$$U'(C_{t+1}) f_{E,t+1}\rho(N_{t+1}) \left[\frac{\mu(N_{t+1}) - \mu(N_t)}{\mu(N_{t+1})}\right] = U'(C_{t+1}) \frac{C_{t+1}}{N_{t+1}} (\mu(N_{t+1}) - \mu(N_t)) \left(\frac{1}{\mu(N_{t+1})} - 1\right). \tag{7}$$

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

$$(1 - \mu(N_{t+1})) = \frac{f_{E,t+1}\rho(N_{t+1}) N_{t+1}}{C_{t+1}} \leq 0, \text{ since } \mu(N_{t+1}) \geq 1. \tag{8}$$

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

For the stochastic case:

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

which is a contradiction since the term in square brackets is strictly greater than zero.

(B) $\mu\left(N_t\right)=\mu\left(N_{t+1}\right)=\mu$ and $\epsilon\left(x\right)\neq\mu\left(x\right)-1$ implies – using Theorem 1– that

$$U'\left(C_{t+1}\right)\frac{C_{t+1}}{N_{t+1}}\left[\epsilon\left(N_{t+1}\right)-\left(\mu-1\right)\right]=0. \tag{9}$$

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

(C) $\mu\left(N_t\right)\neq\mu\left(N_{t+1}\right)$ and $\epsilon\left(x\right)\neq\mu\left(x\right)-1$. In this case, a steady state is still defined by $N_t=N_{t+1}$, so $\mu\left(N_t\right)=\mu\left(N_{t+1}\right)=\mu\left(N\right)$ in steady state. Since the steady state ought to be the same under both CE and PE equilibria, we have (evaluating the Euler equations at the steady state) $\epsilon\left(N\right)=\mu\left(N\right)-1$, which contradicts the assumption $\epsilon\left(x\right)\neq\mu\left(x\right)-1$. This holds for the stochastic case too, and the same argument can be used for point (B). ■

Note that 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). We first discuss this special case (where the conditions of our welfare theorem hold) and then move to discuss the intuition for cases where efficiency fails.

Before we do so, we discuss some properties of the steady state. A sufficient condition for the number of firms in the CE, N^{CE}, to be lower (higher) than the number of firms in the PE, N^{PE}, is that the graph of $H^{CE}\left(N\right)$ lie below (above) the graph of $H^{PE}\left(N\right)$ for any N, or, equivalently, that:

$$\mu\left(N\right)-1<(>)\,\epsilon\left(N\right),\ \forall N.$$

This condition states that if, for a given number of producers, the profit incentives provided by the markup are, say, weaker than the variety effect on welfare, the CE will feature a suboptimally low number of firms (since externalities on consumer welfare of adding extra varieties are not internalized). Note that for translog preferences the opposite holds: The benefit of variety is only half the net markup for any N, so the competitive equilibrium features a suboptimally high number of firms.

Intuition: The C.E.S.-DS Case

Efficiency under C.E.S.-DS preferences stems from two features of our model economy: synchronization of markups and the entry mechanism under these preferences, the role of which we shall now explain in detail.[14] The first piece of intuition, which we will refer to as "the Lerner-Samuelson intuition," concerns the synchronization of markups. 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.[15]

Our model has the important, additional property that the market allocation is efficient even when a dynamic allocation problem is solved under free entry subject to a sunk cost, a time-to-build lag, and exogenous exit. This is important because it implies that the allocation of labor to the two sectors of our economy is efficient, and it contradicts Samuelson's further claim that "If we drop these highly special assumptions [that factors of production are fixed –...], we should not have an optimum situation" (op. cit., p. 240). We let one factor of (aggregate) production (the number of firms, or the stock of production lines) vary and show that the market equilibrium is still efficient since all the new firms charge the same markup.[16] This brings us to the second feature of our economy that ensures efficiency.

Despite synchronized markups, entry could lead to inefficiency due to two other possible distortions – if 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. Grossman and Helpman (1991) call these distortions the "consumer surplus effect" and the "profit destruction effect," respectively.

[14] Our analysis below echoes points made by Grossman and Helpman (1991).

[15] Notice, though, that the equilibrium of our model would be inefficient if, for some reason, the number of firms were fixed because agents are prevented from accessing the available technology for creation of new firms. Inefficiency would arise because the number of firms would be suboptimal.

[16] This result, however, does not hold if we relax the fixed-labor assumption, as shown in Section 5.

With C.E.S.-DS preferences, these two contrasting forces perfectly balance each other and the resulting equilibrium is efficient.[17] However, when preferences do not take the C.E.S.-DS form, inefficiency may arise.

Intuition: The General Case

As Theorem 1 emphasizes, efficiency of the CE requires:

(i) Markup synchronization over time/across states: Goods need to have the same markup at different points in time and in different states – not only markup synchronization across goods. 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 of 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.

(ii) Balancing of consumer surplus and profit destruction effects. This happens only for C.E.S.-DS preferences and is violated if the (net) markup function is different from the benefit of variety in elasticity form. In this case, even if markups were constant, the creation of a new product would have asymmetric effects on the profit incentives driving firm entry and on consumer welfare through the variety effect. For example, Benassy (1996) has proposed a C.E.S. preference specification which separates the degree of monopoly power from the consumer's taste for variety. The difference from the DS specification is that the benefit of variety, $\frac{\rho'(N)}{\rho(N)}N$, is captured separately by a parameter ξ. With these preferences, while the first condition holds (markups are synchronized), the second condition obviously fails since the benefit of variety ξ is generally different from the net markup $\mu - 1$. The economy ends up with a suboptimally low (high) number of producing firms if the parameter governing the taste for variety is lower (higher) than the degree of monopoly power (the net price markup). Nevertheless, this preference specification implies that the consumer derives utility from goods that (s)he never consumes, and similarly is worse off when a good disappears even if consumption of that good was zero. This unappealing feature clearly drives the welfare conclusions.

We have established that the competitive equilibrium of our benchmark model with fixed labor is

[17] See also Dixit and Stiglitz (1977) and Judd (1985) for a discussion of these issues.

efficient under C.E.S.-DS preferences and explained this result based on synchronization of markups and the entry mechanism. As should be intuitive by now, efficiency breaks down when there are differences in markups across firms or sectors of the economy, as is the case when firms are heterogenous and/or price adjustment is not frictionless.[18] Moreover, as we show below, efficiency fails even with C.E.S.-DS preferences when labor supply is endogenous. But we shall argue that this inefficiency is induced by the *absence* of a markup in the pricing of leisure, and not by monopoly power (generating a markup in the consumption production sector). Indeed, we will argue that monopoly power should not be removed (absent entry cost subsidies), since profit incentives are the driving force behind entry and production in our economy. Instead, a simple policy of subsidizing labor income can be designed that restores efficiency by effectively equalizing markups for all the goods the household cares about (including leisure).

5 Endogenous Labor Supply

In this section, we consider a model with endogenous labor supply. The only modification with respect to the model of Section 2 is that now households choose how much labor effort to supply for production of the consumption good and to set up new firms. Consequently, the period utility function features an additional term measuring the disutility of hours worked. For analytical convenience, we specify the utility function as additively separable in consumption and effort. Thus, we assume $U\left(C_t, L_t\right) = \ln C_t - \chi \left(L_t\right)^{1+1/\varphi} / \left(1 + 1/\varphi\right)$, where $\varphi \geq 0$ is the Frisch elasticity of labor supply to wages, and the intertemporal elasticity of substitution in labor supply.[19]

From inspection of Table 2, the only modification to the CE conditions is that L in the aggregate accounting identity now features a time index t. The new variable L_t is then determined in standard fashion by adding to the equilibrium conditions the intratemporal first-order condition of

[18] For instance, the welfare costs of inflation in modern monetary policy analysis relying on staggered price adjustment (*e.g.*, Woodford, 2003) can be easily explained in terms of the Lerner-Samuelson intuition. Staggered price adjustment implies that *ex post* markups are different across firms, and hence there is dispersion in relative prices. When nominal rigidity is introduced in the form of a cost of price adjustment that implies no relative price dispersion, it is time variation in the common markup that induces inefficiency. The policy prescription of price stability can then be explained in both cases in terms of satisfying the condition that markups be synchronized in order to maximize consumer welfare. We explore the implications of imperfect price adjustment in Bilbiie, Ghironi, and Melitz (in progress).

[19] Our choice of functional form for the utility function in this case is guided by results in King, Plosser, and Rebelo (1988): Given separable preferences, log utility from consumption ensures that income and substitution effects of real wage variation on effort cancel out in steady state; this guarantees constant steady-state effort and balanced growth – if there is productivity growth.

the household governing the choice of labor effort:

$$\chi \left(L_t\right)^{\frac{1}{\varphi}} = \frac{w_t}{C_t}. \tag{10}$$

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

$$\chi \left(L_t\right)^{\frac{1}{\varphi}} C_t = Z_t \rho\left(N_t\right)/\mu\left(N_t\right), \tag{11}$$

which 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:[20]

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

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, except for labor being now indexed by time. The additional intratemporal condition for the planning optimum is:

$$\chi \left(L_t\right)^{\frac{1}{\varphi}} C_t = Z_t \rho\left(N_t\right). \tag{12}$$

The only additional difference (with respect to the fixed-labor case) between the competitive market equilibrium and the planning optimum concerns the equations governing intratemporal substitution between consumption and leisure – equations (11) and (12). Comparing these two equations shows that the two equilibria differ as follows. At the Pareto optimum, the marginal rate of substitution between consumption and leisure ($\chi \left(L_t\right)^{\frac{1}{\varphi}} C_t$) is equal to the marginal rate at which hours and the consumption good can be transformed into each other ($Z_t \rho\left(N_t\right)$). In the competitive equilibrium this is no longer the case: There is a possibly time-varying wedge (equal to the reciprocal of the gross price markup, $\left(\mu\left(N_t\right)\right)^{-1}$) between these two objects that can be explained intuitively as follows. Since consumption goods are priced at a markup and leisure is not, the household is less willing than optimal to substitute from leisure into consumption. That is, a suboptimally high amount of leisure is purchased, since this is the relatively cheaper good (implying that hours worked

[20] It is possible to verify that the results on efficiency below hold for a general period utility function $U(C_t, L_t)$.

and consumption are suboptimally low). This result conforms 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." The absence of a markup ('social' degree of monopoly) for the leisure good induces non-synchronization of relative prices which leads to an inefficient allocation: Not enough resources (labor) are devoted to the production of the good with higher markup (consumption). Clearly, this distortion is independent of those emphasized in Theorem 1 (even if preferences were C.E.S.-DS, a wedge would still exist equal to $(\theta - 1)/\theta$, and the CE would be inefficient).

6 Optimal Fiscal Policy

We now study fiscal policies that can help implement the Pareto optimal PE as a competitive equilibrium (or alternatively, that decentralize the planning optimum) when the CE is otherwise inefficient. Our exercise is a 'first-best' one: We assume that lump-sum instruments are available to finance whatever taxation scheme ensures implementation of the optimum. Importantly, since the wedges between the PE and CE are state-contingent, optimal policies aimed at closing these wedges will also be state-contingent. Therefore, all the policies considered in this section can be thought of as feedback rules that specify the optimal, state-contingent responses of fiscal policy instruments to shocks. Since the 'elastic-labor' distortion is independent of those in Theorem 1, we treat it separately and start by looking at the inelastic-labor case; we turn to policies aimed at correcting for the elastic-labor distortion in the final subsection.

Optimal Policy 1: A Sales Subsidy

Suppose the planner subsidizes/taxes sales at rate τ_t and taxes/redistributes proceeds to the firms in a lump-sum amount T_t^f. The profit function becomes: $d_t = (1 + \tau_t)\,\rho_t y_t - w_t l_t - T_t^f$. Optimal pricing will imply $\rho_t = \frac{\mu(N_t)}{1+\tau_t}\frac{w_t}{Z_t}$, so the profit function becomes $d_t = (1 + \tau_t)\,\rho_t y_t - \frac{(1+\tau_t)\rho_t}{\mu(N_t)} y_t - T_t^f$. Assuming zero lump-sum household taxation, balanced budget implies: $T_t^f = \tau_t \rho_t y_t$, so profits are finally given by $d_t = \left(1 - \frac{1+\tau_t}{\mu(N_t)}\right)\rho_t y_t = \left(1 - \frac{1+\tau_t}{\mu(N_t)}\right)\frac{C_t}{N_t}$.

The value of a firm is given by $v_t = w_t \frac{f_{E,t}}{Z_t} = \frac{1+\tau_t}{\mu(N_t)}\rho\,(N_t)\,f_{E,t}$. Substituting these results in the

Euler equation for shares yields:

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

Comparing this with the planner's Euler equation (6), we find that the optimal path of the tax rate must obey:

$$
\begin{aligned}
\frac{1+\tau^*_{t+1}}{1+\tau^*_t} &= \frac{\mu\left(N_{t+1}\right)}{\mu\left(N_t\right)}, \\
\frac{\mu\left(N_t\right)}{1+\tau^*_t}\left(1-\frac{1+\tau^*_{t+1}}{\mu\left(N_{t+1}\right)}\right) &= \epsilon\left(N_{t+1}\right).
\end{aligned}
$$

Using both conditions, we find the optimal subsidy rate as:

$$
1+\tau^*_t = \frac{\mu\left(N_t\right)}{1+\epsilon\left(N_{t+1}\right)}. \tag{13}
$$

Note that this becomes a tax when net markup is less than $\epsilon\left(N_{t+1}\right)$. Also note that the tax rate today must be contingent on the number of firms producing tomorrow. This is a consequence of the time-to-build lag embedded in our model: Today's entrants start producing – and contributing to welfare via variety – tomorrow, and the optimal subsidy rate recognizes this lag in the entry-to-availability process. Importantly, using only *one instrument* – a sales subsidy/tax – is enough to restore efficiency in the general case, although there are *two distortions*: markup intertemporal dispersion and non-synchronization of consumer surplus and profit destruction, $\epsilon\left(x\right) \neq \mu\left(x\right) - 1$. The subsidy/tax rate is not overdetermined since the policy works along two dimensions: the functional form of the subsidy/tax rate at any given time, and the intertemporal path of the subsidy/tax rate.

For example, in the translog case, the optimal subsidy rate is:

$$
\tau_t^{*trans\log} = \frac{2N_{t+1}-N_t}{N_t\left(1+2\sigma N_{t+1}\right)},
$$

while for C.E.S.-Benassy preferences, it is: $1+\tau_t^{*\text{Benassy}} = 1+\tau^{*\,\text{Benassy}} = \frac{1+1/(\theta-1)}{1+\xi}$. Consistent with the implications of Theorem 1, equation (13) implies $\tau^*_t = 0$ in the C.E.S.-DS case.

Two Special Cases

For further illustration, consider the two separate cases where each condition of Theorem 1 holds/fails respectively.

(i) $\mu(N_t) = \mu(N_{t+1}) = \mu$ and $\epsilon(x) \neq \mu(x) - 1$. (For instance, with C.E.S.-Benassy preferences.)

(ii) $\mu(N_t) \neq \mu(N_{t+1})$ and $\epsilon(x) = \mu(x) - 1$. (For instance, this may happen with C.E.S.-DS preferences if we have time varying markups from industry structure – e.g. Cournot competition, implicit collusion, etc., rather than monopolistic competition across a continuum of producers as in our benchmark specification.)

We look at optimal policy in each case.

(i) From the expression for the optimal subsidy, we have $1 + \tau_t^* = \frac{\mu}{1+\epsilon(N_{t+1})}$. But the first condition on the optimal policy implies $\frac{1+\tau_{t+1}^*}{1+\tau_t^*} = 1$. Hence, optimal policy implies a flat tax/subsidy rate

$$\tau_t^* = \tau_0^* = \frac{\mu}{1+\epsilon(N_1)}, \forall t.$$

(ii) In this case, we have $1 + \tau_t^* = \frac{1+\epsilon(N_t)}{1+\epsilon(N_{t+1})}$. But the first condition on the optimal policy implies $\frac{1+\tau_{t+1}^*}{1+\tau_t^*} = \frac{1+\epsilon(N_{t+1})}{1+\epsilon(N_t)}$, so $\tau_{t+1}^* = 0$. The optimal path is to tax-subsidize in the first period and then do nothing:

$$\tau_0^* = \frac{1+\epsilon(N_0)}{1+\epsilon(N_1)}, \text{ and } \tau_t = 0 \, \forall t > 0$$

If the number of firms is expected to increase, producers are subsidized today in order to avoid over-saving, which would be harmful to consumers who would derive relatively less benefit from variety tomorrow.

Optimal Policy 2: An Entry Subsidy/Tax or (De)Regulation Policy

Assume now that the policymaker subsidizes entry at rate ϕ_t. Therefore, entrants pay only $(1-\phi_t) w_t f_{E,t}/Z_t$ entry cost in units of consumption. The only equation in Table 2 that is affected is the free entry condition, which becomes $v_t = w_t (1-\phi_t) f_{E,t}/Z_t$.[21] We can study the optimal value of ϕ_t that

[21] The labor market clearing condition should also be modified accordingly, but we have not listed this equation in Table 2. Note that we do not need to assume any taxation to finance this policy: The government merely legislates that, from date t on, entrants need to hire less labor to create a new product line. See Grossman and Helpman (1991) for an analogous treatment. Additionally, $\mu(N_t) - 1 > 0$ and $\epsilon(N_t) > 0$ will imply that the restriction $\phi_t < 1$ is never binding.

restores the planning optimum. The Euler equation for the competitive economy becomes

$$
\begin{aligned}
&(1-\phi_t) f_{E,t} \rho(N_t) U'(C_t) \qquad (14)\\
&= \beta(1-\delta) E_t \left\{ U'(C_{t+1}) \left[(1-\phi_{t+1}) f_{E,t+1} \rho(N_{t+1}) \frac{\mu(N_t)}{\mu(N_{t+1})} + \frac{C_{t+1}}{N_{t+1}} \mu(N_t) \left(1 - \frac{1}{\mu(N_{t+1})}\right) \right] \right\}.
\end{aligned}
$$

Comparing this with the planning optimum and using the conditions from Theorem 1, we find the path of entry subsidies that restores the optimum:

$$
\begin{aligned}
\frac{1-\phi^*_{t+1}}{1-\phi^*_t} &= \frac{\mu(N_{t+1})}{\mu(N_t)},\\
1-\phi^*_t &= \frac{\mu(N_t)}{\epsilon(N_{t+1})} \left(1 - \frac{1}{\mu(N_{t+1})}\right).
\end{aligned}
$$

Combining the two we obtain:

$$
\begin{aligned}
1-\phi^*_t &= \frac{\mu(N_t)-1}{\epsilon(N_t)},\\
1-\phi^*_{t+1} &= \frac{\mu(N_{t+1})}{\mu(N_t)} (1-\phi^*_t).
\end{aligned}
$$

Note:

1. This policy is procyclical: $N_{t+1} > N_t$ implies $\phi^*_{t+1} > \phi^*_t$ by the monotonicity of $\mu(.)$: More incentives for entry are provided in periods/states with low markups and a high number of firms.

2. Intuitively, a subsidy is used when the net markup is less than the benefit of variety today, since the market does not provide for enough entry incentives; this is in contrast with the sales subsidy for which the relevant object is the benefit of variety tomorrow.

3. The optimal entry subsidy is zero when the markup and benefit from variety are aligned, and markups are synchronized over time/across states. When only the first condition fails (for instance, Benassy preferences) we have $1-\phi^*_t = 1-\phi^*_{t+1} = (\mu-1)/\epsilon$ for any $t \geq 0$. When only the second condition fails ($\mu(N_{t+1}) \neq \mu(N_t)$), we have: $\phi^*_0 = 0$, $1-\phi^*_{t+1} = \frac{\mu(N_{t+1})}{\mu(N_t)}$ for $t \geq 0$.

4. In the translog case,

$$1-\phi_t^{*trans\log}=\frac{\frac{1}{\sigma N_t}}{\frac{1}{2\sigma N_t}}=2\Rightarrow\phi_t^{*trans\log}=-1,\text{ a tax.}$$

Since the benefit of variety is only half the net markup with translog preferences, optimal policy implies more "regulation" (doubling the entry cost).

Optimal Policy 3: Stock Market Taxes

In this subsection, we study the ability of distortionary fiscal instruments applied to the stock market to induce the efficient allocation of resources. We assume that all the taxes/subsidies we consider are rebated/financed through lump-sum transfers to/taxation of households.

To start with, note that a policy whereby only dividends are taxed at the household's level at rate τ_t^D cannot implement the optimal allocation. The CE Euler equation becomes:

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

The rate can be chosen such that markup/profit incentives and variety benefit are aligned:

$$1-\tau_{t+1}^D=\frac{\epsilon\left(N_{t+1}\right)}{\mu\left(N_t\right)\left(1-\frac{1}{\mu\left(N_{t+1}\right)}\right)}.$$

But this does not influence the time-variation in the markup, or any other intertemporal decision – the tax system cannot address both distortions. It can implement the optimum only when there is no endogenous variation in markups – for instance, with Benassy preferences.

Tax on Total Payoff

Suppose that both dividend income and proceeds from selling shares are taxed at the same rate τ_t^P. The household budget constraint is:

$$v_tN_{H,t}x_{t+1}+C_t+T_t=\left(1-\tau_t^P\right)\left(d_t+v_t\right)N_tx_t+w_tL,$$

and the CE Euler equation becomes:

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

The optimal path of the tax rate should satisfy:

$$
\begin{aligned}
\left(1-\tau_{t+1}^P\right)\frac{\mu\left(N_t\right)}{\mu\left(N_{t+1}\right)} &= 1, \\
\left(1-\tau_{t+1}^P\right)\mu\left(N_t\right)\left(1-\frac{1}{\mu\left(N_{t+1}\right)}\right) &= \epsilon\left(N_{t+1}\right),
\end{aligned}
$$

or:

$$
\begin{aligned}
1-\tau_{t+1}^P &= \frac{\mu\left(N_{t+1}\right)}{\mu\left(N_t\right)}, \\
1-\tau_{t+1}^P &= \frac{1+\epsilon\left(N_{t+1}\right)}{\mu\left(N_t\right)}.
\end{aligned}
$$

The system is overdetermined (unless preferences are such that $\epsilon\left(x\right)=\mu\left(x\right)-1$), and the initial tax rate τ_0^P is undetermined. The problem with this scheme is that it does not influence the investment (entry) decision, and hence it cannot correct for the misalignment of markup and variety effect.

Tax on Net Asset Trades

Suppose that all net changes in the asset position $N_{H,t}x_{t+1}-N_tx_t$ resulting from buying/selling shares at the price v_t are taxed at rate τ_t^{NA}. The household budget constraint is:

$$
v_tN_{H,t}x_{t+1}+\tau_t^{NA}v_t\left(N_{H,t}x_{t+1}-N_tx_t\right)+C_t+T_t=\left(d_t+v_t\right)N_tx_t+w_tL.
$$

The CE Euler equation for shares becomes:

$$
\begin{aligned}
& \left(1+\tau_t^{NA}\right)f_{E,t}\rho\left(N_t\right)U'\left(C_t\right) \qquad (15) \\
& = \beta\left(1-\delta\right)E_t\left\{U'\left(C_{t+1}\right)\left[\left(1+\tau_{t+1}^{NA}\right)f_{E,t+1}\rho\left(N_{t+1}\right)\frac{\mu\left(N_t\right)}{\mu\left(N_{t+1}\right)}+\frac{C_{t+1}}{N_{t+1}}\mu\left(N_t\right)\left(1-\frac{1}{\mu\left(N_{t+1}\right)}\right)\right]\right\}.
\end{aligned}
$$

Note that equation (15) is identical to (14): Subsidies on the net asset position $\left(-\tau_t^{NA}\right)$ are equivalent to entry subsidies ϕ_t. The intuition is straightforward. Net asset subsidies encourage households to hold shares/invest rather then directly encouraging firms to enter, but these two decisions mirror each other in general equilibrium. Therefore the optimal path of this tax is:

$$\tau_t^{NA*} = -\phi_t^*.$$

For example, under translog preferences, asset transactions are optimally taxed to discourage investment in firms that would provide 'too much' extra variety, in the sense that its benefit to the consumer would be less than the entry incentive to firms generated by the net markup.

Tax on Gross Asset Trades

Suppose that each time an asset trade is performed, the household pays a tax ψ_t proportional to the size of the gross trade. Since short sales never occur in equilibrium, the cost is always deducted from the proceeds of a share sale, and added to the cost of share purchases. The household budget constraint is:

$$v_t\left(1+\psi_t\right)N_{H,t}x_{t+1}+C_t+T_t=\left[d_t+\left(1-\psi_t\right)v_t\right]N_tx_t+w_tL.$$

The CE Euler equation becomes:

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

Optimal policy therefore obeys:

$$\begin{aligned}\frac{1-\psi_{t+1}^*}{1+\psi_t^*}\frac{\mu\left(N_t\right)}{\mu\left(N_{t+1}\right)}&=1,\\\frac{\mu\left(N_t\right)}{1+\psi_t^*}\left(1-\frac{1}{\mu\left(N_{t+1}\right)}\right)&=\epsilon\left(N_{t+1}\right),\end{aligned}$$

or:

$$\begin{aligned}1-\psi_t^*&=\frac{\mu\left(N_t\right)-1}{\epsilon\left(N_t\right)},\\1-\psi_{t+1}^*&=\frac{\mu\left(N_{t+1}\right)}{\mu\left(N_t\right)}\left(1+\psi_t^*\right).\end{aligned}$$

While the functional form of the optimal tax rate implied by the first of these equations is the same as for the entry subsidy ϕ_t^* or a tax on net asset trades, its dynamic path implied by the second equation is different. A high tax rate today implies, *ceteris paribus*, a lower tax rate tomorrow. Hence, the tax rate can be oscillatory. In the special case in which markups are constant over time/across states, the optimal policy is $\psi_{t+1}^* = -\psi_t^*$.

Optimal Policy 4: A Labor Subsidy

When labor supply is elastic, there is one more distortion to correct for. Efficiency can clearly be restored by subsidizing labor supply (or taxing leisure) at a rate equal to the net markup in the pricing of consumption goods and applying a lump-sum tax/transfer to the households. Suppose the government subsidizes labor at the rate τ_t^L. Lump-sum taxes are used to finance this policy. The first-order condition for the household's optimal choice of labor supply is the only equilibrium condition that is affected:

$$\chi \left(L_t\right)^{\frac{1}{\varphi}} C_t = \left(1 + \tau_t^L\right) w_t.$$

Combining this with the wage schedule $w_t = Z_t \rho\left(N_t\right)/\mu\left(N_t\right)$ yields:

$$\chi \left(L_t\right)^{\frac{1}{\varphi}} C_t = \left(1 + \tau_t^L\right) Z_t \rho\left(N_t\right)/\mu\left(N_t\right).$$

Comparing this equation to (12) shows that a rate of taxation equal to the net markup of price over marginal cost:

$$1 + \tau_t^{L*} = \mu\left(N_t\right) \tag{16}$$

restores efficiency of the market equilibrium. This policy ensures synchronization of markups, consistent with the Lerner-Samuelson intuition described above. Note that while the same policy would also induce efficiency in a model with a fixed number of firms, there is an important difference concerning optimal policy between that framework and our model. When N is exogenously fixed, this policy is equivalent to one that induces marginal cost pricing of consumption goods by subsidizing firm revenues (again synchronizing relative prices between consumption and leisure) and financing this subsidy with a lump-sum tax on firm profits.

As we verify formally below, this equivalence no longer holds in our framework with entry: Such a policy would remove the wedge from equation (11), 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. Therefore, while markup synchronization is *necessary* for efficiency, it is *not sufficient.* Absent an entry cost subsidy, the sufficient condition states that the planner needs to align markups to the *higher* (positive) level. Doing otherwise (inducing marginal cost pricing while driving equilibrium profits to zero) would make the economy stop producing altogether. 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 sunk costs that cannot be entirely subsidized. Indeed, note by direct comparison that the labor subsidy τ_t^{L*} in (16) is equal to the sales subsidy τ_t^* in (13) *if and only if* there is no benefit of product variety, *i.e.*, $\epsilon(x) = 0$ for any x. Note also that 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.

Implications for Some Common Policy Prescriptions

Many macroeconomic studies featuring monopolistic competition that look at issues related to welfare (for instance, in the context of optimal monetary policy) take for granted that monopoly power is a distortion having to do with price being above marginal cost. A common argument is that, before correcting for any other distortion (such as dispersion of relative prices due to staggered price setting), the "markup distortion" must be eliminated (for example, in order to make the steady state of the model efficient, before addressing stabilization around this steady state).[22] As we have argued, our results imply that the distortion has nothing to do with monopolistic competition and the presence of a markup in the market for goods in and of itself, but rather with the absence of a markup on leisure, or, more specifically, with the non-synchronization of markups on consumption and leisure.

Relatedly, in an influential paper, Erceg, Henderson, and Levin (2000) assume that there is a markup in the labor market too (to motivate the introduction of sticky wages) and, in order to restore steady-state efficiency, they argue that two subsidies are needed to eliminate the two monopoly distortions. Our insights imply that there is only one distortion, again related to *relative* markups, and only one subsidy should be used. Moreover, if entry is allowed for, markups should be aligned at the level ensuring the optimal level of entry.

In a series of papers, Schmitt-Grohé and Uribe (2004a,b) argue that "In our imperfectly competitive economy, profits represent the income to a fixed 'factor,' namely, monopoly rights. It is therefore optimal for the Ramsey planner to tax profits at a 100% rate."[23] Monopoly power is

[22] A prominent example is Khan, King, and Wolman (2003), who talk about the "markup distortion" separately and show how it can be eliminated, treating it independently from a series of other inefficiencies (relative price dispersion, shopping time, and monetary distortions).

[23] They go on to show that when this is not feasible, governments use the nominal interest rate in order to tax profits indirectly.

again viewed as a "bad" that optimal policy ought to eliminate. Finally, a large literature deals with optimal fiscal policy under imperfect competition in the public finance vein. In Auerbach and Hines (2002, 2003) the point that market power must be eliminated by a subsidy is made explicitly (though the setup features partial equilibrium and Cournot competition).[24] Moreover, in Auerbach and Hines (2002), they note that entry makes things worse, for subsidies provide incentives for inefficient entry by new firms.

In a general equilibrium model with entry, a policy targeted at inducing marginal cost pricing can have disastrous effects. For example, while in the C.E.S.-DS case with elastic labor a sales subsidy does restore the optimum when financed by lump-sum taxes on the consumer, this is quite a special case. When even a small fraction of the subsidy is financed by taxing the firm (as usually done in the literature quoted above), the optimum is no longer restored, as taxation of the firm affects the entry decision. When all the taxes are paid by firms, this policy leads to starvation as no firm will find it optimal to produce. Only if the split between the taxes paid by consumers and firms is determined optimally, marginal cost pricing can restore the optimum – and the optimal split features zero lump-sum taxation of firm profits in the C.E.S.-DS case. We demonstrate this point below by studying the effect of a policy inducing marginal cost pricing in the fully general case.

The Effect of Inducing Marginal Cost Pricing

Suppose the planner subsidizes/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 profit function becomes: $d_t = (1+\tau_t)\rho_t y_t - w_t l_t - T_t^F$. Optimal pricing will imply $\rho_t = \frac{\mu(N_t)}{1+\tau_t}\frac{w_t}{Z_t}$, so the profit function becomes $d_t = (1+\tau_t)\rho_t y_t - \frac{(1+\tau_t)}{\mu(N_t)}\rho_t y_t - T_t^F$. 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

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

In order 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

[24] They attribute the argument for subsidies in the presence of imperfect competition in order to induce marginal cost pricing, presumably restoring efficiency, to the early works of Joan Robinson, Cournot, and Musgrave.

$v_t = w_t \frac{f_{E,t}}{Z_t} = \rho(N_t) f_{E,t}$. Substituting these expression in the CE Euler equation for shares yields:

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

The PE Euler equation is:

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

Comparing the two, efficiency of the CE is restored if and only if the fraction of taxes paid for by the firm satisfies:

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

Recall that for C.E.S.-DS preferences (the case studies by all the literature reviewed above), $\epsilon = \mu - 1$. 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 the consumer, and none by the firm. 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 $\epsilon(x) = (\mu(x) - 1)/2$). So a policy inducing of marginal cost pricing can restore efficiency only if an optimal division of lump-sum taxes between consumers and firms is also ensured.

7 Conclusions

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. Under inelastic labor supply and linearity of production in labor, the market economy is efficient if and only if symmetric, homothetic preferences are of the C.E.S. form studied by Dixit and Stiglitz (1977). Otherwise, efficiency is restored by properly designed sales, entry, or asset trade subsidies (or taxes) that induce markup synchronization across time and states, and align the consumer surplus and profit destruction effects of firm entry. 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.

By studying efficiency and optimal policy in a DSGE environment, this paper contributes to the literature on the efficiency properties of models with monopolistic competition that dates back to at least Lerner (1934). The policy schemes that implement the planning optimum 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.

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 naive interpretations of statements in recent literature on the "distortionary" consequences of monopoly power and the required remedies.

References

[1] Auerbach, A. J., and J. R. Hines (2002): "Taxation and Economic Efficiency," in Auerbach, A., and M. Feldstein, eds., *Handbook of Public Economics*, vol. 3, 1347-1421, Elsevier, Amsterdam.

[2] Auerbach, A. J., and J. R. Hines (2003): "Perfect Taxation with Imperfect Competition," in Cnossen, S., and H.-W. Sinn, eds., *Public Finance and Public Policy in the New Century*, 127-153, MIT Press, Cambridge, MA.

[3] Benassy, J.-P. (1996): "Taste for Variety and Optimum Production Patterns in Monopolistic Competition," *Economics Letters* 52: 41-47.

[4] Bilbiie, F. O., F. Ghironi, and M. J. Melitz (2005): "Endogenous Entry, Product Variety and Business Cycles," manuscript, Oxford University, Boston College, and Harvard University.

[5] Bilbiie, F. O., F. Ghironi, and M. J. Melitz (in progress): "Monetary Policy, Inflation, and Firm Dynamics," manuscript, Oxford University, Boston College, and Harvard University.

[6] Caballero, R. J., and M. L. Hammour (1994): "The Cleansing Effect of Recessions," *American Economic Review* 84: 1350-1368.

[7] Campbell, J. R. (1998): "Entry, Exit, Embodied Technology, and Business Cycles," *Review of Economic Dynamics* 1: 371-408.

[8] Campbell, J. Y. (1994): "Inspecting the Mechanism: An Analytical Approach to the Stochastic Growth Model," *Journal of Monetary Economics* 33: 463-506.

[9] Dixit, A. K., and J. E. Stiglitz (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review* 67: 297-308.

[10] Erceg, C. J., D. W. Henderson, and A. T. Levin (2000): "Optimal Monetary Policy with Staggered Wage and Price Contracts," *Journal of Monetary Economics* 46: 281-313.

[11] Feenstra, R. C. (2003): "A Homothetic Utility Function for Monopolistic Competition Models, without Constant Price Elasticity," *Economics Letters* 78: 79-86.

[12] Grossman, G. M., and E. Helpman (1991): *Innovation and Growth in the Global Economy*, MIT Press, Cambridge, MA.

[13] Judd, K. L. (1985): "On the Performance of Patents," *Econometrica* 53: 567-585.

[14] Khan, A., R. G. King, and A. L. Wolman (2003): "Optimal Monetary Policy," *Review of Economic Studies* 70: 825-860.

[15] Kim, J. (2004): "What Determines Aggregate Returns to Scale?" *Journal of Economic Dynamics and Control* 28: 1577-1594.

[16] Lerner, A. (1934): "The Concept Monopoly and the Measurement of Monopoly Power," *Review of Economic Studies* 2: 157-175.

[17] Mankiw, N. G., and M. D. Whinston (1986): "Free Entry and Social Inefficiency," *RAND Journal of Economics* 17: 48-58.

[18] Samuelson, P. (1947): *Foundations of Economic Analysis*, Harvard University Press, Cambridge, MA.

[19] Schmitt-Grohé, S., and M. Uribe (2004*a*): "Optimal Fiscal and Monetary Policy under Sticky Prices," *Journal of Economic Theory* 114: 198-230.

[20] Schmitt-Grohé, S., and M. Uribe (2004*b*): "Optimal Fiscal and Monetary Policy under Imperfect Competition," *Journal of Macroeconomics* 26: 183-209.

[21] Spence, M. (1976): "Product Selection, Fixed Costs, and Monopolistic Competition," *Review of Economic Studies* 43: 217-235.

[22] Woodford, M. (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, Princeton.

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:

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