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 Hugo Hopenhayn, Gianmario Impullitti, Jinill Kim, and seminar and conference participants at Catholic University Lisbon, Chicago GSB, Cleveland Fed, Cornell, CSEF-IGIER Symposium on Economics and Institutions, ECB, EUI, ESEM 2007, ESSIM 2005, EUI, Fed Board, IZA Workshop on Firm Dynamics, Leuven, London Business School, MIT, NBER EFCE Summer Institute 2005, New York Fed, Northeastern, Princeton, SED 2007, Swiss National Bank, Universite de Toulouse, 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, the CEP at LSE, and the NBER for hospitality in 2004, 2005, and 2006-2007, respectively, and Nuffield College at Oxford for financial support during 2004-2007. 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.

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
JEL No. D42,H32,L16

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.¸

Florin O. Bilbiie
Finance and Economics Department
HEC Paris Business School
1 rue de la Libration
78352 Jouy en Josas cedex
France
bilbiie@hec.fr

Marc J. Melitz
Dept of Economics & Woodrow Wilson School
Princeton University
308 Fisher Hall
Princeton, NJ 08544
and NBER
mmelitz@princeton.edu

Fabio Ghironi
Boston College
Department of Economics
140 Commonwealth Avenue
Chestnut Hill, MA 02467-3859
and NBER
fabio.ghironi@bc.edu

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 (2007 – henceforth, BGM). Specifically, we compare the competitive and planner equilibria, asking whether the market solution generates 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 an obsolescence risk. 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.[1]

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), Chamberlin (1950), Spence (1976), Dixit and Stiglitz (1977), Judd (1985a), and Grossman and Helpman (1991), among others.[2]

Under assumptions of inelastic labor supply and linearity of production in labor, our main result is that a monopolistically competitive market generates socially efficient economic fluctuations and product entry (that is, the competitive and planner equilibria coincide) when consumers have symmetric, 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 (the profit signal driving entry). This requires that preferences be of the C.E.S. form originally studied by Dixit and Stiglitz (1977).

[1] In BGM, we focus on productivity as the sole source of business cycle shocks. Here, we allow also for random fluctuations in entry costs to highlight the generality of the properties of the policy schemes we consider.

[2] See also Mankiw and Whinston (1986), Benassy (1996), and Kim (2004).

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. The competitive equilibrium is then no longer efficient. However, the relevant distortion is not the existence of a markup in the market for goods, but heterogeneity in markups between all "goods" consumed: 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 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 the gains from variety with the profit incentive for 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 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.[3] In fact, 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).

Our results reiterate an argument made elsewhere in the literature that monopoly power in and of itself is not a distortion, and show that monopoly profits should in fact be preserved whenever entry is endogenously determined. Indeed, while markup synchronization across time, states, and goods is still a necessary condition for efficiency, sufficiency requires that markups also be aligned with the benefit of additional product variety. Our findings are also closely related to a large body

[3] We are implicitly assuming that the government is not contemporaneously subsidizing the entire amount of the entry cost.

of literature studying optimal monetary and fiscal policy in the presence of monopolistic competition; however, our results stand in sharp contrast to the common policy prescription eliminating monopoly profits. When product variety is endogenously determined by firm entry, monopoly profits play a crucial role in generating the welfare maximizing level of product variety.

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 and is followed by a discussion in Section 7 relating our results to the existing literature. Section 8 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): $E_t \sum_{s=t}^{\infty} \beta^{s-t} U(C_s)$, 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 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 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.[4]

Given the demand for an individual variety, $c_t(\omega)$, the symmetric price elasticity of demand ζ

[4] See the appendix for more details.

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 over consumption goods 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.[5]

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.[6]

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

[5] Consistent with standard real business cycle theory, aggregate productivity Z_t affects all firms uniformly.

[6] 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.

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: $\mu_t = \mu(N_t) \equiv \zeta(N_t)/(\zeta(N_t)+1)$. 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 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 additional varieties in elasticity form ($\epsilon(N_t)$) must decrease.[7] The change in $\epsilon(N_t)$ is only half the change in the 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.

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

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 $\{d_s(\omega)\}_{s=t+1}^{\infty}$:

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

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.[8] 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 of producing firms represents the capital stock of the economy. It is an endogenous state variable that behaves much like physical capital in the benchmark real business cycle (RBC) model.

[8] Periods 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: $v_t(\omega) < w_t f_{E,t}/Z_t$.

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 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.[9]

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 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

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

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 this equation and absence of speculative bubbles yield the asset price solution in equation (1).[10]

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 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

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

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.[11]

The Competitive Equilibrium

The model with general homothetic preferences is summarized in Table 2.[12]

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:

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

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

[11] When 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.

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

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.[13]

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 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 $r \equiv (1-\beta)/\beta$.[14]

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, $\lim_{N\to 0}H^{CE}(N) = \infty$, and $\lim_{N\to\infty}H^{CE}(N) = 0$, $\left[H^{CE}(N)\right]' \leq 0$ is a sufficient and locally necessary condition for $H^{CE}(N)$ to have a unique fixed point. Since

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

the condition $\left[H^{CE}(N)\right]' \leq 0$ is equivalent to $\mu'(N) \leq 0$.

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

[13] 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.

[14] 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.

very reasonable condition: If goods were to become more differentiated as product variety increases, then the possibility of 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.[15] In the present paper, we compare the competitive equilibrium with the planning optimum.

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 (including the number of goods produced).

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:

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

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

$$\max_{\{N_{s+1}\}_{s=t}^{\infty}} E_t \sum_{s=t}^{\infty} \beta^{s-t} U\left[Z_s \rho(N_s) \left(L - \frac{1}{(1-\delta)} \frac{f_{E,s}}{Z_s} N_{s+1} + \frac{f_{E,s}}{Z_s} N_s\right)\right]. \quad (6)$$

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

$$U'(C_t) \rho(N_t) f_{E,t} = \beta (1-\delta) E_t \left\{ U'(C_{t+1}) \left[f_{E,t+1} \rho(N_{t+1}) + \frac{C_{t+1}}{N_{t+1}} \epsilon(N_{t+1}) \right] \right\}. \quad (7)$$

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 (7) 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

[15] We show there that the competitive equilibrium is always locally determinate in the log-linearized version of our model.

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 conditions $\lim_{N\to 0}\epsilon(N)=\infty$, $\lim_{N\to\infty}\epsilon(N)=0$, and $\epsilon'(N)\leq 0$ ensure that the system of stochastic difference equations (5) and (7) has a unique stationary equilibrium.[16] The intuition for 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.

4 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. Sufficiency ('if') is directly verified by plugging conditions (i) and (ii) into (4) and (7).

Necessity ('only if') requires that, whenever both (4) and (7) 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 $\epsilon(x)=\mu(x)-1$:

[16] 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 $\epsilon(N)$.

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

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

$$U'(C_{t+1}) f_{E,t+1} \rho(N_{t+1}) \left[\frac{\mu(N_{t+1}) - \mu(N_t)}{\mu(N_{t+1})} \right] = U'(C_{t+1}) \frac{C_{t+1}}{N_{t+1}} (\mu(N_{t+1}) - \mu(N_t)) \left(\frac{1}{\mu(N_{t+1})} - 1 \right). \tag{8}$$

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

$$1 - \mu(N_{t+1}) = \frac{f_{E,t+1} \rho(N_{t+1}) N_{t+1}}{C_{t+1}} \leq 0, \text{ since } \mu(N_{t+1}) \geq 1. \tag{9}$$

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

$$U'(C_{t+1}) \frac{C_{t+1}}{N_{t+1}} [\epsilon(N_{t+1}) - (\mu - 1)] = 0. \tag{10}$$

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). ■

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 describe alternative 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 $H^{CE}(N)$ lie below (above) the graph of $H^{PE}(N)$ for any N; or equivalently that $\mu(N) - 1 < (>)\, \epsilon(N)$, $\forall N$. This condition states that if, for a given number of producers, the profit incentives provided by the markup are weaker (stronger) than the variety effect on welfare, then the CE will feature a suboptimally low (high) number of firms. Note that for translog preferences, the benefit of variety is only half the net markup for any . The competitive equilibrium thus features a suboptimally high number of firms.

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

C.E.S.-DS preferences induce two special features in our model economy that jointly deliver efficiency: synchronization of markups (across firms and time) and the alignment of profit incentives with the benefits of product variety.[18] 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.[19]

Our model generalizes this efficiency property to the case of a dynamic allocation problem solved under free entry subject to 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).

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

[19] 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.

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.[20] 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". With C.E.S.-DS preferences, these two contrasting forces perfectly balance each other and the resulting equilibrium is efficient.[21] However, when preferences do not take the C.E.S.-DS form, inefficiency arises.

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 – so markup synchronization only across goods is not sufficient. 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.

(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 that separates the degree of monopoly power from the consumer's taste for variety.[22] The difference from the benchmark DS specification is that the benefit of variety, $\rho'(N)\,N/\rho(N)$, is captured separately by a parameter ξ. With these preferences, while the first condition for efficiency of the CE holds

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

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

[22] This separation was present also in Dixit and Stiglitz (1975) – a working paper version of Dixit and Stiglitz (1977) – and Ethier (1982).

(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 higher (lower) than the degree of monopoly power (the net price markup). 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.[23]

We have established that the competitive equilibrium of our benchmark model with fixed labor is 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.[24] Inefficiency also occurs when labor supply is endogenous since leisure is not subjected to the markup applied to consumption goods. However, we shall argue that this does not imply that monopoly power should be removed (absent entry cost subsidies), since profit incentives are the driving force behind entry (which delivers the welfare benefit of product variety). 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

To address the consequences of lack of markup synchronization across goods, we now introduce endogenous labor supply. 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,

[23] More recently, Kim (2004) studies the role of taste for variety in the determination of aggregate returns to scale with Benassy preferences. Koeniger and Licandro (2006) develop a model that disentangles the elasticity of substitution from the degree of market power based on industrial structure rather than preferences.

[24] 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 (2008).

nonseparable utility function over consumption and effort, $U(C_t, L_t)$, under standard assumptions 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]

From inspection of Table 2, the two modifications to the CE conditions 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:

$$-U_L(C_t, L_t) = w_t U_C(C_t, L_t). \tag{11}$$

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:

$$-U_L(C_t, L_t)/U_C(C_t, L_t) = Z_t \rho(N_t)/\mu(N_t). \tag{12}$$

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:

$$-U_L(C_t, L_t)/U_C(C_t, L_t) = Z_t \rho(N_t). \tag{13}$$

The only 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 (12) and (13). Comparing these two equations shows

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

that the two equilibria differ as follows: At the Pareto optimum, the marginal rate of substitution between consumption and leisure $\left(-U_L\left(C_t, L_t\right)/U_C\left(C_t, L_t\right)\right)$ is equal to the marginal rate at which hours and the consumption good can be transformed into each other $\left(Z_t \rho\left(N_t\right)\right)$. 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, $\left(\mu\left(N_t\right)\right)^{-1}$. Since consumption goods are priced at a markup while leisure is not, demand for the latter is suboptimally high (hence, hours worked and consumption are suboptimally 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 of goods removes this distortion by ensuring effective markup synchronization.[26]

6 Optimal Fiscal Policy

We now study fiscal policies that can implement the Pareto optimal PE as a competitive equilibrium (or alternatively, that 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. 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: An Entry Subsidy/Tax or (De)Regulation Policy

We start by studying a policy that subsidizes firm entry at rate ϕ_t. Therefore, entrants pay only $\left(1-\phi_t\right) w_t f_{E,t}/Z_t$ entry cost in units of consumption, and this subsidy is fully financed by lump-sum taxes on consumers T_t. The only equations in Table 2 that are affected are the free entry condition, which becomes $v_t = w_t\left(1-\phi_t\right) f_{E,t}/Z_t$, and the aggregate resource constraint, which becomes $C_t + N_{E,t} v_t/\left(1-\phi_t\right) = w_t L + N_t d_t$ after substituting the government balanced budget

[26] Thus, 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."

requirement $T_t = \phi_t N_{E,t} w_t f_{E,t}/Z_t$.[27] Proposition 1 studies the optimal value of ϕ_t that restores the planning optimum.

Proposition 1 *A subsidy to firm entry restores efficiency of the competitive equilibrium if:*

$$
\begin{aligned}
1-\phi_t^* &= \frac{\mu\left(N_t\right)-1}{\epsilon\left(N_t\right)}, \\
1-\phi_{t+1}^* &= \frac{\mu\left(N_{t+1}\right)}{\mu\left(N_t\right)}\left(1-\phi_t^*\right).
\end{aligned} \tag{14}
$$

Proof. The Proposition is readily proven once one observes that under the entry subsidy ϕ_t the state equation for the number of firms (5) is unaffected and the Euler equation for the competitive economy becomes

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

Comparing this with the planning optimum and using the conditions from Theorem 1, we find the optimal path of entry subsidies ϕ_t^* as in Proposition 1.[28] ■

Consistent with our intuition, an entry subsidy will work to align (i) the profit incentive $\mu\left(N_t\right)-1$ and the benefit of variety $\epsilon\left(N_t\right)$ (the first equation in (14)) and (ii) markups over time and across states (the second equation in (14)). Specifically, a subsidy is in place in any period in which the benefit of variety is higher than the profit incentive, $\epsilon\left(N_t\right) > \mu\left(N_t\right)-1$, for in that case policy must provide extra incentives for entry/investment that are not provided by the market. Moreover, the subsidy rate is procyclical in order to counteract the time variation in markups: Higher subsidies are needed in periods/states with a higher number of firms, since it is precisely in those situations of relatively low markups that extra incentives for entry are needed (recall that the markup is a decreasing function of the number of firms). Importantly, using only one instrument (an entry subsidy/tax) is enough to restore efficiency in the general case even if there are two distortions (different markups over time and non-synchronization of consumer surplus and profit destruction,

[27] Note that an increase in the subsidy ϕ_t is fundamentally different from a fall in the exogenous component of the sunk entry cost $f_{E,t}$ which does not affect the aggregate resource constraint. Indeed, the latter is equivalent to a deregulation policy that would simply legislate a lower labor requirement for opening up new firms without any financing needs. Such a policy, while implying a similar free-entry condition, would leave the resource constraint unaffected but would instead change the labor market clearing condition. This explains why the optimal entry subsidy obtained below is not $\phi_t = 1$, although $f_{E,t} = 0 \forall t$ implies optimality.

[28] Additionally, note that $\mu\left(N_t\right) - 1 > 0$ and $\epsilon\left(N_t\right) > 0$ imply that the restriction $\phi_t < 1$ is never binding.

$\epsilon(x) \neq \mu(x) - 1$). 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.

To substantiate this intuition, we briefly review a few examples. As one would expect from Theorem 1, the optimal entry subsidy is zero when the markup and benefit from variety are aligned, and markups are synchronized over time/across states – i.e., for C.E.S.-DS preferences. When only the first condition fails (for instance, for Benassy preferences) the optimal subsidy is constant and simply re-balances the markup and the benefit of variety: $1 - \phi_t^* = 1 - \phi_{t+1}^* = (\mu - 1)/\epsilon$ for any $t \geq 0$. When only the second condition fails ($\mu(N_{t+1}) \neq \mu(N_t)$), we have: $\phi_0^* = 0$, $1 - \phi_{t+1}^* = \mu(N_{t+1})/\mu(N_t)$ for $t \geq 1$. In this case, the subsidy is used only to synchronize markups over time and across states. In the translog case, where both distortions are at work, the first expression in (14) ensures that benefit of variety and markup are aligned and provides an initial condition for the subsidy rate:

$$\phi_0^{*trans\log} = 1 - \frac{(\sigma N_0)^{-1}}{(2\sigma N_0)^{-1}} = -1,$$

hence 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) in the initial period. From period $t = 1$, the second condition in (14) fully determines the dynamics of the subsidy rate given this initial condition, and corrects for the distortion associated with markup non-synchronization over time and across states.

Optimal Policy 2: A Subsidy to Net Asset Trades

Decentralization by means of an entry subsidy is not the unique way to implement the PE as an outcome of market behavior in our model. The first alternative that we review exploits the general equilibrium structure of the model by noting that, since the decision of firms to enter is mirrored by a decision of households to invest in new firms, an equivalent policy will try to influence the latter. Specifically, a subsidy/tax to net asset trades can achieve efficiency as follows. Suppose that all net changes in the asset position of the household $N_{H,t}x_{t+1} - N_t x_t$ resulting from buying/selling shares at the price v_t are subsidized at rate τ_t^A. The household budget constraint is:

$$v_t N_{H,t} x_{t+1} - \tau_t^A v_t (N_{H,t} x_{t+1} - N_t x_t) + C_t + T_t = (d_t + v_t) N_t x_t + w_t L.$$

Comparing the Euler equation in the CE to its PE counterpart shows that the optimal path of this subsidy is:

$$\tau_t^{A*} = \phi_t^*.$$

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.[29]

Optimal Policy 3: A Sales Subsidy

Another option to restore efficiency studied by virtually every paper addressing the possible distortions associated with monopoly since Robinson (1933) is a subsidy to firm sales that affects the pricing decisions of firms. We study this option also in order to clearly relate our results to those of the literature we discuss in the Conclusions.

Specifically, 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 following Proposition finds the path of this sales subsidy that makes the competitive equilibrium efficient:

Proposition 2 *A subsidy to firm sales financed by lump-sum taxes on firm profits restores efficiency of the competitive equilibrium if:*

$$1 + \tau_t^* = \frac{\mu(N_t)}{1 + \epsilon(N_{t+1})}. \tag{15}$$

Proof. To prove this statement, note that the profit function becomes: $d_t = (1+\tau_t)\rho_t y_t - w_t l_t - T_t^f$. Optimal pricing implies $\rho_t = \frac{\mu(N_t)}{1+\tau_t}\frac{w_t}{Z_t}$, 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:

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

[29] We refer the interested reader to the appendix for a more complete study of stock market taxes as a way to decentralize the planning optimum. In particular, we explain why some forms of taxation (such as a tax on total payoff or dividends) do not achieve the decentralization, while others (such as a tax on gross asset trades) do.

Comparing this with the planner's Euler equation (7), we find that the optimal path of τ_t must obey:

$$\frac{1+\tau_{t+1}^*}{1+\tau_t^*} = \frac{\mu\left(N_{t+1}\right)}{\mu\left(N_t\right)} \quad \text{and} \quad \frac{\mu\left(N_t\right)}{1+\tau_t^*}\left(1-\frac{1+\tau_{t+1}^*}{\mu\left(N_{t+1}\right)}\right) = \epsilon\left(N_{t+1}\right). \tag{16}$$

Combining these conditions yields the optimal subsidy rate as in Proposition 2. ■

Similarly to the entry subsidy, this subsidy works by aligning the benefit of variety with the incentive for entry. However, note that the subsidy becomes a tax when the incentive for entry in the current period ($\mu\left(N_t\right)-1$) is less than the benefit of variety in the next ($\epsilon\left(N_{t+1}\right)$). The current tax rate must be contingent on the number of firms producing in the next period due to the time-to-build lag embedded in our model: Entrants at time t start producing – and contributing to welfare via variety – at time $t+1$, and the optimal subsidy rate recognizes this lag in the entry-to-availability process. This is fundamentally different from the entry subsidy studied above, since that policy instrument works directly *ex ante* on the decision to enter, and not *ex post* on sales of producing firms who already entered. The translog example studied below delves more into the intuition behind this argument.

We now discuss a few special cases for the entry subsidy. Consistent with Theorem 1, equation (15) implies $\tau_t^* = 0$ in the C.E.S.-DS case. If markups are synchronized over time/across states ($\mu\left(N_t\right) = \mu\left(N_{t+1}\right) = \mu$), but the benefit of variety is different from the profit incentive ($\epsilon\left(x\right) \neq \mu\left(x\right)-1$), the optimal subsidy is $1+\tau_t^* = \frac{\mu}{1+\epsilon(N_{t+1})}$. But the first condition on the optimal policy implies $\frac{1+\tau_{t+1}^*}{1+\tau_t^*} = 1$. Hence, the optimal policy implies a flat tax/subsidy rate

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

For instance, with C.E.S.-Benassy preferences, this is: $1+\tau_t^{*\text{Benassy}} = 1+\tau^{*\,\text{Benassy}} = \theta/\left[\left(\theta-1\right)\left(1+\xi\right)\right].$

If the benefit of variety and the entry incentive are aligned ($\epsilon\left(x\right) = \mu\left(x\right)-1$), but markups are not synchronized ($\mu\left(N_t\right) \neq \mu\left(N_{t+1}\right)$), the optimal sales subsidy is $1+\tau_t^* = \left(1+\epsilon\left(N_t\right)\right)/\left(1+\epsilon\left(N_{t+1}\right)\right)$.[30] But the first condition for optimal policy in (16) implies $\tau_{t+1}^* = 0$. Therefore, the optimal policy is to tax (or subsidize) in the first period and then do nothing:

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

[30] For instance, this happens with C.E.S.-DS preferences if time-varying markups arise either from industry structure (Cournot competition, implicit collusion, etc., rather than monopolistic competition across a continuum of producers as in our specification) or as a consequence of an exogenously time-varying elasticity of substitution θ_t.

If the number of firms is expected to increase, producers are subsidized in the current period in order to avoid over-saving, which would reduce consumer welfare by causing less benefit from variety in the next period. This result is reminiscent of the Judd (1985b) and Chamley (1986) result that capital income should be taxed heavily in the initial period and not taxed at all thereafter. Doing otherwise would distort the intertemporal entry/investment decision in a way that is similar to the distortion of physical investment decisions introduced by positive taxes on physical capital in Judd's and Chamley's models.

Finally, in the translog case, where both conditions of Theorem 1 fail, the optimal subsidy rate is:

$$1+\tau_t^{*trans\log}=\frac{1+\frac{1}{\sigma N_t}}{1+\frac{1}{2\sigma N_{t+1}}}.$$

The numerator of this expression corrects for time-variation of markups under translog preferences, and the denominator aligns the benefit of variety with entry incentives. A tax is optimal ($\tau_t^{*trans\log}<0$) whenever the number of firms tomorrow compared to today is such that the benefit of variety tomorrow $(2\sigma N_{t+1})^{-1}$ is higher than the market profit signal for entry provided by the net markup today, $(\sigma N_t)^{-1}$. In the absence of the tax, there would therefore be an incentive to under-invest today, implying too little variety tomorrow. This is optimally corrected by taxing sales today, therefore increasing the incentives to allocate resources to investing in new varieties (the extensive margin), rather than producing and selling more of the existing varieties (the intensive margin).

In the remainder of this Section, we address two questions that are related to our analysis of sales subsidies. First, we show that a labor subsidy can restore efficiency when labor supply is elastic, and we argue that a labor subsidy is equivalent to a sales subsidy only when there is no endogenous product variety. Second, we study the conditions under which a policy that induces marginal cost pricing (as originally suggested by Robinson, 1933) restores efficiency.

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, financing this policy with lump-sum taxes on household income. The first-order condition for the household's optimal choice of labor supply is the only equilibrium condition that is affected:

$$-U_L(C_t, L_t)/U_C(C_t, L_t) = \left(1 + \tau_t^L\right) w_t.$$

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

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

Comparing this equation to (13) shows that a rate of taxation of leisure equal to the net markup of price over marginal cost,

$$1 + \tau_t^{L*} = \mu(N_t), \tag{17}$$

restores efficiency of the market equilibrium. This policy ensures synchronization of markups, consistent with the Lerner-Samuelson intuition described above. 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.

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 below, this equivalence no longer holds in our framework with producer entry: Such a policy would remove the wedge from equation (12), 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 (17) is equal to the sales subsidy τ_t^* in (15) *if and only if* there is no benefit of product variety, i.e., $\epsilon(x) = 0$ for any x.

Optimal Policy 5: The Effect of Inducing Marginal Cost Pricing

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 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. When all the taxes are paid by firms, this policy would induce a disastrous outcome since no firm would find it optimal to enter. Marginal cost pricing would restore the optimum. In the C.E.S.-DS case, the optimal split features zero lump-sum taxation of firm profits in the C.E.S.-DS case. We demonstrate this point by studying the effect of a policy inducing marginal cost pricing in the fully general case. Specifically, 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.

Proposition 3 *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: $d_t = (1+\tau_t)\rho_t y_t - w_t l_t - T_t^F$. Optimal pricing implies $\rho_t = \frac{\mu(N_t)}{1+\tau_t}\frac{w_t}{Z_t}$, 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

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

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 (7) 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 Proposition 3. ■

Proposition 3 shows that 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 $\epsilon(x) = (\mu(x) - 1)/2$).

7 Discussion

Our results are related to a large literature studying optimal monetary and fiscal policy in the presence of monopolistic competition. An incomplete list includes Blanchard and Kiyotaki (1987), Adão, Correia and Teles (2003), Khan, King, and Wolman (2003), Schmitt-Grohé and Uribe (2004a,b), Erceg, Henderson, and Levin (2000), and Woodford (2003) for monetary policy, and Auerbach and Hines (2002, 2003) and references therein for fiscal policy. All of these studies take for granted that monopoly power is a source of distortion due to prices 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, in the case of optimal monetary policy), 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). As we have argued, the distortion is not due to the presence of a markup in the market for goods in and of itself, but rather to the non-synchronization of markups on consumption and leisure. When there is a markup in the labor market too (as in Blanchard and Kiyotaki, 1987, and Erceg, Henderson, and Levin, 2000), our results imply that there is only one distortion, related to relative markups, and that only one, appropriately designed subsidy is sufficient to correct it. Moreover, if producer entry is allowed for, markups should be aligned at the level that ensures the optimal level of entry.

In a model with entry such as ours, eliminating profits is not optimal since profits are needed to provide the right incentives for product creation. This is in contrast with policy prescriptions that suggest the elimination of monopoly profits either directly by inducing marginal cost pricing or indirectly by creating inflation (Schmitt-Grohé and Uribe, 2004a,b). The prescription that profits should be removed by the equalization of prices to marginal costs (which can be traced back to Robinson, 1933, pp. 163-165) is omnipresent in the public finance literature studying optimal taxation in the presence of imperfect competition (see e.g. Auerbach and Hines, 2002, 2003). Since the stock of available products/firms acts as a physical capital stock and investment in firm entry is similar to investment in physical capital, our results are also related to the literature studying optimal taxation of capital income, e.g. Judd (1985b) and Chamley (1986). In particular, the optimality of subsidizing firm entry or investment in our framework echoes the result of Judd (1997, 2002) that the optimal tax rate on capital income is negative, obtained in a second-best environment (i.e., with distortionary taxes only). Indeed, our results can be viewed as providing conditions for the desirability of a "negative capital tax" (i.e., an entry subsidy), when capital accumulation takes the form of investment in the creation of new goods, even in an environment in which access to lump-sum instruments makes it possible to implement the first-best outcome.[31] Finally, our model can be viewed as a generalization of Judd (1985a), who first studied the optimal length of patents in a model with product innovation and differentiation.

8 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

[31] Our results are related also to findings in Abel (2007). The combination of consumption tax and leisure tax that he considers is reminiscent of our markup plus leisure tax scenario if one thinks of the markup as a tax on consumption. Under certain circumstances, this tax does not result in inefficiency. When labor supply is elastic, consumption and leisure taxes must be properly aligned.

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 Robinson (1933) and Lerner (1934). The policy schemes that implement the planning optimum in our model fully specify the optimal path of the relevant distortionary instruments over 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 interpretations of statements in recent literature on the "distortionary" consequences of monopoly power and the required remedies.

References

[1] Abel, A. B. (2006): "Optimal Capital Income Taxation," NBER Working Paper 13354.

[2] Adão, B., I. Correia, and P. Teles (2003): "Gaps and Triangles," *Review of Economic Studies* 70: 699-713.

[3] 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.

[4] 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.

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

[6] Bilbiie, F. O., F. Ghironi, and M. J. Melitz (2007): "Endogenous Entry, Product Variety, and Business Cycles," NBER Working Paper 13646.

[7] Bilbiie, F. O., F. Ghironi, and M. J. Melitz (2008): "Monetary Policy, Inflation, 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 (forthcoming).

[8] Blanchard, O. J., and N. Kiyotaki (1987): "Monopolistic Competition and the Effects of Aggregate Demand," *American Economic Review* 77: 647-666.

[9] Chamberlin, E. (1950): "Product Heterogeneity and Public Policy," *American Economic Review* 40: 85-92.

[10] Chamley, C. (1986): "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica* 54: 607-622.

[11] Dixit, A. K., and J. E. Stiglitz (1975): "Monopolistic Competition and Optimum Product Diversity," University of Warwick Working Paper 64.

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

[13] 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.

[14] Ethier, W. J. (1982): "National and International Returns to Scale in the Modern Theory of International Trade," *American Economic Review* 72: 389-405.

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

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

[17] Hopenhayn, H. (1992): "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica* 60: 1127–1150.

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

[19] Judd, K. (1985b): "Redistributive Taxation in a Simple Perfect Foresight Model," *Journal of Public Economics* 28: 59-83.

[20] Judd, K. (1997): "The Optimal Tax on Capital Income is Negative," NBER Working Paper 6004.

[21] Judd, K. (2002): "Capital Income Taxation with Imperfect Competition," *American Economic Review* 92: 417-421.

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

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

[24] Koeniger, W., and O. Licandro (2006): "On the Use of Substitutability as a Measure of Competition," *B.E. Press Topics in Macroeconomics* 6(1), Article 2.

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

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

[27] Melitz, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica* 71: 1695-1725.

[28] Robinson, J. (1933): *The Economics of Imperfect Competition*, Cambridge University Press, Cambridge, UK.

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

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

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

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

[33] 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:

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

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, $\Lambda_t(\omega) = \max\left[v_t(\omega) - w_t f_{E,t}/Z_t, \beta\Lambda_{t+1}(\omega)\right]$,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: $\Lambda_t = \max\left[v_t - w_t f_{E,t}/Z_t, \beta\Lambda_{t+1}\right]$. 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 $\Lambda_t = 0$.

C The First-Order Condition for the Planning Problem

The first-order condition for problem (6) 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:

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

Hence, the first-order condition becomes:

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

leading to (7).

D Optimal Stock Market Taxes

In this appendix, we delve deeper into the question of what distortionary fiscal instruments applied to the stock market can induce the efficient allocation of resources. More precisely, we study two taxation schemes that cannot restore efficiency and one that can (in addition to the subsidy to net asset trades explored in the main text). The purpose of studying the policies that fail to restore efficiency is to gain a better understanding of the interaction of fiscal policy, the asset market, and the entry decision. We assume that all the taxes/subsidies we consider are rebated/financed through lump-sum transfers to/taxation of households. To start with, we consider **taxation of profit income** (dividends) at the household's level at rate τ_t^D. The Euler equation in the CE 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 τ_t^D can be chosen such that markup/profit incentives and variety benefit are aligned:

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

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

Next, consider a **tax on total payoff**, i.e., suppose that both dividend income and proceeds from selling shares are taxed at the same rate τ_t^P. The household budget constraint is:

$$v_t N_{H,t} x_{t+1} + C_t + T_t = \left(1 - \tau_t^P\right) (d_t + v_t) N_t x_t + w_t L,$$

and the CE Euler equation becomes:

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

Comparing this with the planner's Euler equation (7), we find that the optimal path of the tax rate should satisfy:

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

The system is overdetermined (unless preferences are such that $\epsilon (x) = \mu (x) - 1$), 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.

Finally, for comparison with the subsidy to net asset trades study in the main text consider a **tax on gross asset trades**: 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:

$$v_t (1 + \psi_t) N_{H,t} x_{t+1} + C_t + T_t = [d_t + (1 - \psi_t) v_t] N_t x_t + w_t L.$$

The CE Euler equation becomes:

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

Optimal policy therefore obeys (by comparing the above equation with the planner's Euler equation

(7)):

$$1-\psi_t^* = \frac{\mu(N_t)-1}{\epsilon(N_t)},$$
$$1-\psi_{t+1}^* = \frac{\mu(N_{t+1})}{\mu(N_t)}(1+\psi_t^*).$$

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 $\psi_{t+1}^* = -\psi_t^*$.