NBER WORKING PAPER SERIES

FIRM HETEROGENEITY, ENDOGENOUS ENTRY, AND THE BUSINESS CYCLE

Gianmarco I.P. Ottaviano

Working Paper 17433
http://www.nber.org/papers/w17433

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2011

I am grateful to Bocconi University, Bruegel, FEEM, and the European Commission for financial support. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Firm Heterogeneity, Endogenous Entry, and the Business Cycle
Gianmarco I.P. Ottaviano
NBER Working Paper No. 17433
September 2011
JEL No. E20,E32,L11,L16

ABSTRACT

This paper investigates the role that the entry and exit of heterogeneous firms plays in shaping aggregate fluctuations in economic activity. In so doing, it develops a dynamic stochastic general equilibrium model in which procyclical entry and countercyclical exit along a real business cycle lead to endogenous cyclical movements in average firm productivity. These movements stem from a composition effect due to the reallocation of market shares among firms with different levels of efficiency and affect the propagation of exogenous technological shocks. Numerical analysis suggests that existing models with representative firms may overstate the actual role of procyclical entry and exit in imperfectly competitive markets as a propagation mechanism of exogenous technology shocks. The reason is that procyclical entry and countercyclical exit disproportionately involve less efficiency firms whose impact on aggregate economic activity is hampered by their smaller size.

Gianmarco I.P. Ottaviano
Department of Economics
Bocconi University
Via Roentgen 1
20136 Milan
Italy
and CEPR, and also FEEM
gianmarco.ottaviano@unibocconi.it

1 Introduction

This paper investigates the role that the entry and exit of heterogeneous firms play in shaping the aggregate fluctuations of economic activity. In so doing, it develops a dynamic stochastic general equilibrium model in which procyclical entry and countercyclical exit along a real business cycle lead to endogenous cyclical movements in average firm productivity that affect the propagation of exogenous technological shocks.

These endogenous cyclical movements stem from a composition effect due to the reallocation of market shares among firms with different levels of efficiency. Reallocation happens at the extensive margins due to entry and exit that disproportionately involve less efficient firms. It also happens at the intensive margin as incumbents with different efficiency levels perceive different changes in demand elasticity even if faced with the same aggregate shock.

The model features endogenous markups. Whether these move on average pro- or counter-cyclically depends on the balance between two opposing effects as the elasticity of demand varies along the business cycle for two reasons. First, it increases with the number of competitors. As entry is procyclical and exit is countercyclical, the resulting fluctuations in net business creation would lead to countercyclical markups. Second, the elasticity of demand also increases with the average efficiency of competitors. As entry and exit disproportionately involve less efficient firms, the resulting countercyclical fluctuations in average firm efficiency would generate pro-cyclical markups. Accordingly, which effect eventually dominates depends on the chosen parametrization of the distribution of firm heterogeneity and the speed at which the two effects materialize in the model.

The present paper is related to three main traditions in the macroeconomic literature. The first tradition concerns the models of creative destruction and, in particular, their implications in terms of "cleansing" during recessions. As discussed by Caballero and Hammour (1994), the key premise of these models is that the continuous process of creation and destruction of production units that results from product and process innovation is essential for understanding not only growth, but also business cycles. In their vintage model with heterogeneous technologies, Caballero and Hammour (1994) show that industries undergoing continuous creative destruction accommodate demand fluctuations during the business cycle by varying the rates at which more efficient production units are created and less efficient ones are destroyed. As their model is able to generate greater cyclicality of destruction than creation, recessions can be interpreted as times of "cleansing", when less efficient techniques and products are expelled from the market. Caballero and Hammour (1994) argue that this differential cyclicality is consistent with the empirical patterns of job creation and job destruction reported, for instance, by Davis and Haltiwanger (1990, 1992).

In the second tradition, search models of unemployment with endogenous job separation à la Mortensen and Pissarides (1994) generate an analogous cleansing effect in the labor market when matches are heterogeneous. This is due to the presence of countercyclical selectivity in the decision whether to continue a match, implying that labor productivity in the endogenous separations model responds less than one-to-one to aggregate productivity shocks. These models shed light on the apparent weak cyclicality of real wages by stressing a composition effect for workers analogous to the one for techniques highlighted by Caballero and Hammour (1994): as less-skilled workers are more vulnerable to layoffs, they account for a smaller share of employment at business cycle troughs than at business cycle peaks. These compositional changes induce a countercyclical bias in the aggregate real wage despite procyclical labor productivity (Abraham and Haltiwanger, 1995). The magnitude of new hires and job separations during the business cycle draws attention to questions of search and matching, providing strong motivation for theories of frictional unemployment (Davis, Faberman and Haltiwanger, 2006).

The third and closest tradition to the methodological approach of the present paper highlights the importance of imperfect competition for understanding the amplitude of the business cycle. Rotemberg and Woodford (1991, 1992, 1995) investigate the implications of oligopoly arguing that countercyclical movements in markups due to implicit collusive behaviour magnify the fluctuations of aggregate economic activity. Gali (1994) shows that, when firms face demands from different sources, also variations in demand composition can generate cyclical movements in markups. These works take the number of competitors as fixed. More recently, Jaimovich and Floetotto (2008) argue that cyclical variations in the number of competitors leading to countercyclical markups can act as a powerful propagation mechanism of technology shocks. In Bilbiie, Ghironi and Melitz (2007) countercyclical markups coexist with procyclical profits, a feature of the data that previous models had a hard time to explain, as discussed, for instance, by Rotemberg and Woodford (1999). All these works assume representative firms.

Differently from these contributions the present paper introduces firm heterogeneity and shows that entry and exit affect the elasticity of demand not only because of the implied variations in the number of competitors but also because of the resulting changes in the average efficiency of competitors. This way it generates a composition effect that is reminiscent of the one highlighted by the aforementioned flow approach to the labor market. It also yields a cleansing effect that is analogous to the one stressed by the literature on creative destruction.

This is achieved by formulating a stochastic dynamic general equilibrium growth model in the spirit of Bilbiie, Ghironi and Melitz (2007). These authors propose a discrete-time, stochastic, general equilibrium version of a two-sector variety-based growth model with representative firms à la Grossman and Helpman (1991). This is used to focus on the business cycle very much like the textbook exogenous growth model is used in the real business cycle literature. Ghironi and Melitz (2005) introduce firm heterogeneity in a similar framework but abstract from the business cycle implications of entry and exit, focusing instead on how selection into export status affects real exchange rate adjustment. They also assume a constant elasticity of demand and, therefore, constant markups.

In the two-sector model à la Grossman and Helpman (1991) proposed in the present paper a sector is devoted to capital accumulation and employs labor under constant returns to scale and perfect competition. The other sector supplies an array of horizontally differentiated products under increasing returns to scale and monopolistic competition. Each product is offered by a firm employing a fixed amount of capital and a variable amount of labor. Firms are heterogeneous with respect to this variable amount. Heterogeneity is itself endogenous as in Melitz (2003). To enter the market firms have to hire the required fixed amount of capital. After paying the corresponding rental price, they draw their variable unit labor requirements (the inverse of their labor productivity or "efficiency") from some common probability distribution. Then, knowing their own labor productivity and the productivity of their potential competitors, they decide whether to start producing or to exit. The exit decision obeys a cutoff rule of survival: only entrants with high enough labor productivity become producers; the other entrants leave the market without even starting production.

On the demand side, the proposed model borrows its instantaneous utility function from the static setup of Melitz and Ottaviano (2008). However, by removing the linear component of their quasi-linear quadratic function, it crucially introduces income effects and variable marginal utility of income as in Neary (2007). The resulting demand system exhibits variable elasticity with less productive firms facing higher demand elasticity than more productive ones. Accordingly, they charge lower markups. This is not enough to offset their inefficiency, so they quote higher prices and are smaller in terms of output, revenues, and profits. All these implications comply with the empirical evidence recently collected, for example, in the trade literature (Tybout, 2003; Bernard, Jensen, Redding and Schott, 2007).

The focus of the analysis is on how the proposed model reacts to exogenous technology shocks that change the number of efficiency units per worker. The main result is that on impact more efficiency units per worker make survival easier for a larger number of less efficient firms. Accordingly, the number of producers increases but their average efficiency falls, leading to higher average price and average markup as well as lower industry concentration and average output per firm. After impact, as capital begins to accumulate, the number of entrants starts growing. Survival becomes tougher, triggering firm exit until the number of producers, their average efficiency, price, markup and size together with industry concentration go back to their long run levels. The initial wave of entry just after capital starts accumulating is strong enough to impose a J-shaped adjustment to average firm efficiency, markup and price: after impact average efficiency rises above its long run level, hence approaching it from above; average markup and price fall below their long run levels before approaching them from below.

As new producers and quitters are less efficient than incumbents, during business cycle upswings on impact there is more entry, more survival after entry, and surviving firms are on average less efficient and smaller. The opposite is true during downswings. Hence, the impact of changing the number of efficiency units per worker on aggregate output per worker and welfare is reduced by the pro-cyclical entry and the countercyclical exit of less efficient firms. Due to variable demand elasticity, such a dampening effect of firm selection on aggregate productivity and welfare works also through a second channel. Holding the number of incumbents constant, in an upswing industry concentration decreases as market shares are reallocated towards less efficient firms due to the fact that the elasticity of demand falls more for high-price firms than for low-price ones. In a downswing the opposite happens. This second channel would be muted if demand exhibited constant elasticity as in Ghironi and Melitz (2005).

The dampening effect of firm selection depends on the degree of firm heterogeneity: the impact of changing the number of efficiency units per worker on aggregate output per worker and welfare is stronger the less heterogeneous firms are. This reveals a way through which microeconomic heterogeneity may crucially affect the propagation of shocks at the aggregate level. Specifically, it implies that existing models with representative firms (such as the ones by Jaimovich and Floetotto, 2008, and Bilbiie, Ghironi and Melitz, 2007) may overstate the actual role of procyclical entry and exit in imperfectly competitive markets as a propagation mechanism of technology shocks. The reason is that procyclical entry and countercyclical exit disproportionately involve less efficiency firms whose impact on aggregates is hampered by their smaller size.

As inefficient firms are typically small, it may be argued that variations in their number are potentially of limited practical relevance for aggregate fluctuations. This is not what the present findings imply. Indeed, as argued by Jaimovich and Floetotto (2008), variations in the number of firms are only one of the channels that generate actual changes in the number of competitors. In particular, a new establishment or franchise by an existing firm increases the number of competitors without affecting the number of active firms. Based on quarterly data from 1992 to 2005 for the US, Jaimovich and Floetotto (2008) find that the average fraction of quarterly gross job-gains (job-losses) that can be explained by the opening (closing) of establishments is about 20%. Similarly, around a third of the cyclical volatility of the job-gains (job-losses) comes from opening (closing) establishments. In the same vein, Broda and Weinstein (2007) find that net product creation is strongly procyclical. This provides evidence for a sizable variation in the number of competitors at the business cycle frequency as long as one adopts a loose interpretation of entry and exit that includes not only firms but also their portfolios of establishments, franchises, and products. Additional supportive evidence along these lines is gathered by Bilbiie, Ghironi and Melitz (2007). On the other hand, it is worthwhile stressing that entry and exit do not represent the only way through which competitive pressures vary in the present paper: even if entry and exit were blocked, market shares would still be reallocated across heterogeneous incumbents due to variable demand elasticity, thus leading to compositional changes at the aggregate level. What the findings in the present paper do imply is, instead, that the role of entry and exit as a propagation mechanism of technology shocks cannot be fully assessed without simultaneously considering firm heterogeneity.

The rest of the paper is organized in five additional sections. Section 2 lays down the model. Section 3 characterizes its equilibrium. Section 4 solves for its steady state and discusses its comparative statics properties. Section 4 studies the transitionary dynamics around the steady state in the presence of technology shocks. Section 5 concludes.

2 The Model

2.1 Endowments

There are L identical workers each supplying z_s efficiency units of labor inelastically every period. Accordingly, $L_s = Lz_s$ is the number of efficiency units of labor available each period and z_s can be interpreted as an aggregate labor productivity shock. At any time s, there are also K_s units of capital owned by workers. Whereas the labor stock is exogenously given, the capital stock is endogenously accumulated.

2.2 Preferences

Workers' individual preferences are captured by the following intertemporal expected utility function

$$E_t\left\{U_t\right\} = E_t\left\{\sum_{s=t}^{\infty}\beta^{s-t}u(q_s^c(\omega)),\ \omega\in[1,N_s])\right\} \tag{1}$$

where $\beta \in (0,1)$ denotes the rate of time preference and instantaneous utility is defined over a continuum of horizontally differentiated products

$$u(q_s^c(\omega),\ \omega\in[0,N_s]) = \alpha\int_0^{N_s} q_s^c(\omega)d\omega - \frac{1}{2}\gamma\int_0^{N_s}\left(q_s^c(\omega)\right)^2 d\omega - \frac{1}{2}\eta\left(\int_0^{N_s} q_s^c(\omega)d\omega\right)^2 \tag{2}$$

with N_s and $q_s^c(\omega)$ respectively denoting the measure ("number") of available products and the individual consumption level of product ω. Parameters are all positive with γ measuring product differentiation.

There is free borrowing and lending on a perfect financial market where bonds and capital are freely traded. Intertemporal utility (1) is maximized subject to a standard dynamic budget constraint defined in nominal terms

$$B_{s+1}^c - B_s^c + I_s^c + R_s^c = i_s B_s^c + Y_s^c \tag{3}$$

where B_s^c is bond holdings, I_s^c is investment in capital accumulation, Y_s^c is income and $R_s^c = \int_0^{N_s} p_s(\omega)q_s^c(\omega)d\omega$ is expenditures on the consumption of the differentiated products with $p_s(\omega)$ denoting the price of product ω. Iterating the dynamic budget constraint (3) gives the corresponding intertemporal budget constraint

$$(1+i_t)B_t^c + \sum_{s=t}^{\infty} F_{t,s} Y_s^c = \sum_{s=t}^{\infty} F_{t,s}\left(I_s^c + R_s^c\right) \tag{4}$$

where $F_{t,s}$ is the discount factor defined as

$$F_{t,s} = \begin{cases} 1 & s = t \\ \frac{1}{\prod_{v=t+1}^{s}(1+i_v)} & s = t+1, ... \end{cases} \tag{5}$$

and the transversality condition

$$\lim_{T\to\infty} F_{t,t+T} B_{t+T+1}^c = 0$$

has been imposed.

2.3 Technology

There are two sectors, one supplying the differentiated products and the other supplying additional units of capital. The differentiated products are supplied by monopolistically competitive firms employing both capital and labor. In particular, the supply of any product requires a fixed input requirement in terms of f units of capital and a variable input requirement in terms of efficiency units of labor. Firms enter and exit the market freely so that at any time the expected profit from entry is capitalized in the value V_s of each of the f units of capital a firm needs to start production.

The capital stock evolves through time driven by depreciation and investment in capital accumulation. The supply of new capital takes place under perfect competition. A new unit of capital is produced by employing f_I efficiency units of labor and becomes available for production with a one-period time-to-build lag. In every period all units of capital face the same probability $\delta \in (0,1)$ of being destroyed. This implies that a fraction δ of the capital stock is destroyed every period or, equivalently, the capital stock depreciates at rate δ. The exogenous destruction shock occurs after production and investment have taken place at the very end of the time period. Therefore, a fraction δ of new units of capital never becomes available for goods production.

While all firms face the same fixed capital requirement f, their labor requirements per unit of output varies depending on their individual productivity, which they get to know only after entering the market by hiring capital. Firm productivity is determined as follows. At the beginning of period s there are: $(1-\delta)K_{s-1}$ "old" units of capital that were already avaliable at time $s-1$, plus $(1-\delta)I_{s-1}/V_{s-1}$ "new" units of capital accumulated through investment I_{s-1} at time $s-1$ by paying the corresponding price V_{s-1}. In order to enter the market in period s, potential firms competitively bid for the available units of capital $K_s = (1-\delta)\left(K_{s-1} + I_{s-1}/V_{s-1}\right)$.

Due to the fixed capital requirement f, only K_s/f firms are eventually able to enter. Once capital has been allocated to the winning bidders, entrants are assigned their unit labor requirement c (in efficiency units) as a random draw from a common time invariant continuous differentiable distribution with c.d.f. $G(c)$ over the support $[0, c_M]$. Based on their draws, entrants then decide whether to produce or not. Letting N_s and ρ_s respectively denote the mass ("number") and the share of entrants that decide to produce, the former equals $N_s = \rho_s K_s / f$.

Given this set of assumptions, individual investment, bond holdings and income can be respectively written as $I_s^c = V_s \left(K_{s+1} x_{s+1}^c / (1-\delta) - K_s x_s^c \right)$, $B_s^c = B_s y_s^c$ and $Y_s^c = D_s x_s^c + W_s z$, where x_s^c is the individual share of the capital stock, y_s^c is the individual share of bonds, D_s is the aggregate dividend paid by entrants that decide to produce, V_s is the (ex-dividend) value of a unit of capital, and W_s is the wage per efficiency unit. The intertemporal budget constraint (4) then becomes

$$(1+i_t) B_t y_t^c + \sum_{s=t}^{\infty} F_{t,s} \left(D_s x_s^c + W_s z \right) = \sum_{s=t}^{\infty} F_{t,s} \left[V_s \left(\frac{K_{s+1} x_{s+1}^c}{1-\delta} - K_s x_s^c \right) + \int_0^{N_s} p_s(\omega) q_s^c(\omega) d\omega \right] \quad (6)$$

3 Equilibrium

3.1 Consumption and Investment

The utility maximization problem can be solved by the Lagrangian method. Using (1) and (6), the Lagrangian can be written as

$$E_t \left\{ \sum_{s=t}^{\infty} \beta^{s-t} \left[\alpha \int_0^{N_s} q_s^c(\omega) d\omega - \frac{1}{2} \gamma \int_0^{N_s} \left(q_s^c(\omega) \right)^2 d\omega - \frac{1}{2} \eta \left(\int_0^{N_s} q_s^c(\omega) d\omega \right)^2 \right] \right\}$$
$$-\lambda \left[\sum_{s=t}^{\infty} F_{t,s} \left(\frac{V_s K_{s+1} x_{s+1}^c}{1-\delta} - V_s K_s x_s^c + \int_0^{N_s} p_s(\omega) q_s^c(\omega) d\omega \right) \right.$$
$$\left. -(1+i_s) B_t y_t - \sum_{s=t}^{\infty} F_{t,s} \left(D_s x_s^c + W_s z \right) \right]$$

3.1.1 Consumption Decision

The FOC with respect to $q_s^c(\omega)$ requires

$$\beta^{s-t} \left(\alpha - \gamma q_s^c(\omega) - \eta Q_s^c \right) = \lambda F_{t,s} p_s(\omega)$$

with

$$Q_s^c = \int_0^{N_s} q_s^c(\omega) d\omega$$

Note that $s = t$ implies that λ equals the initial marginal utility of consumption

$$\lambda P_t = N_t \alpha - (\gamma + \eta N_t) Q_t^c$$

with $P_s = \int_0^{N_s} p_s(\omega) d\omega$.[1] Given (5), if one defines

$$\lambda_s \equiv \frac{\lambda F_{t,s}}{\beta^{s-t}}$$

the instantaneous inverse demand for product ω can be written as

$$\alpha - \gamma q_s^c(\omega) - \eta Q_s^c = \lambda_s p_s(\omega) \tag{7}$$

with Euler condition

$$\beta E_s \{(1 + i_{s+1})\lambda_{s+1}\} = \lambda_s \tag{8}$$

Individual consumption can then be obtained by integrating (7) across products and solving for

$$Q_s^c = \frac{N_s \alpha - \lambda_s P_s}{\gamma + \eta N_s} \tag{9}$$

which shows that $\alpha > \lambda_s \widetilde{p}_s$ with $\widetilde{p}_s = P_s / N_s$ has to hold if any consumption has to take place at all ($Q_s^c > 0$).

Substituting (9) in (7) gives

$$q_s^c(\omega) = \frac{\lambda_s}{\gamma} \left(\frac{\frac{\alpha}{\lambda_s}\gamma + \eta P_s}{\gamma + \eta N_s} - p_s(\omega) \right)$$

Hence, products priced above the choke price

$$p_s = \frac{\frac{\alpha}{\lambda_s}\gamma + \eta P_s}{\gamma + \eta N_s} \tag{10}$$

are not bought ($q_s^c(\omega) = 0$). Individual inverse demand for product ω can then be written as

$$p_s(\omega) = p_s - \frac{\gamma}{\lambda_s} q_s^c(\omega)$$

with corresponding total demand and total inverse demand respectively equal to

$$q_s(\omega) = q_s^c(\omega) L = \frac{\lambda_s L}{\gamma} (p_s - p_s(\omega))$$

$$p_s(\omega) = p_s - \frac{\gamma}{\lambda_s L} q_s(\omega) \tag{11}$$

The associated elasticity of demand is an increasing function of the own price $p_s(\omega)$ and a decreasing function of the choke price p_s:

$$\left| \frac{dq_s(\omega)}{dp_s(\omega)} \frac{p_s(\omega)}{q_s(\omega)} \right| = \left(\frac{p_s}{p_s(\omega)} - 1 \right)^{-1} = \left(\frac{p_s}{p_s - \frac{\gamma}{\lambda_s L} q_s(\omega)} - 1 \right)^{-1} \tag{12}$$

[1] In the static quasi-linear case of Melitz and Ottaviano (2008), the marginal utility of income is $\lambda = 1$.

It is also an increasing function of the number of consumers L and the marginal utility of income as well as a decreasing function of the quantity demanded $q_s(\omega)$ and the extent of product differentiation γ. Note that the impact of changing p_s is stronger for higher $p_s(\omega)$. In turn, going back to (10), the choke price p_s is a decreasing function of the marginal utility of income λ_s and the number of producers N_s as well as a decreasing function of their average price $\widetilde{p}_s$. Hence, any increase (decrease) in the marginal utility of income and the number of producers as well as any decrease (increase) in their average price leads to a rise (fall) in the elasticity of demand. This makes competition tougher (softer) for all firms but disproportionately so for high price firms.

3.1.2 Investment Decision

The FOC with respect to x_{s+1}^c requires

$$-\frac{F_{t,s}V_sK_{s+1}}{1-\delta}+F_{t,s+1}V_{s+1}K_{s+1}+F_{t,s+1}D_{s+1}=0$$

which, by (5), can be rewritten as the no-arbitrage condition

$$1+i_{s+1}=(1-\delta)\left(\frac{V_{s+1}}{V_s}+\frac{D_{s+1}}{V_sK_{s+1}}\right) \tag{13}$$

which states that there are no profits to be made by arbitraging between bonds and capital.

3.2 Goods Production and Dividends

Profit maximization in goods production requires marginal revenue to match marginal cost. Given total inverse demand (11), the FOC for profit maximization in period s by a firm with unit labor requirement c implies output

$$q_s(c)=\frac{\lambda_sL}{2\gamma}\left(p_s-W_sc\right)$$

This uniquely identifies a cutoff unit labor requirement in efficiency units

$$c_s=\frac{p_s}{W_s} \tag{14}$$

such that $q_s(c_s)=0$ and only firms whose unit labor requirement satisfies $c\leq c_s$ end up producing. The share of entrants that decide to produce therefore equals $\rho_s=G(c_s)$ so that the number of producers is

$$N_s=G(c_s)\frac{K_s}{f} \tag{15}$$

Expression (14) can be used to rewrite firm output as

$$q_s(c)=\frac{\lambda_sW_sL}{2\gamma}\left(c_s-c\right)$$

which can be plugged into total inverse demand (11) to obtain the corresponding price, markup, revenue and profit:

$$\begin{array}{cc} p_s(c) = \frac{W_s}{2}\left(c_s + c\right) & \mu_s(c) = \frac{W_s}{2}\left(c_s - c\right) \\ r_s(c) = \frac{\lambda_s L (W_s)^2}{4\gamma}\left((c_s)^2 - (c)^2\right) & \pi_s(c) = \frac{\lambda_s L (W_s)^2}{4\gamma}\left(c_s - c\right)^2 \end{array} \qquad (16)$$

Profit is equally shared as dividends among the f units of capital hired by the firm. More productive firms have lower value of c. They are, therefore, bigger in terms of both output and revenues. They quote lower prices but have higher markups. As higher markups are associated with larger output, more productive firms also generate more profits. A lower cutoff c_s reduces the price, the output, the revenues and the profits of all firms. As it increases the elasticity of demand, it also reduces the markup, which makes c_s an inverse measure of the toughness of competition.

Based on (16), average price, average markup and average output evaluate to

$$\begin{aligned} \frac{P_s}{N_s} &= \int_o^{c_s} p_s(c)dG_s(c) = \frac{W_s}{2}\left(c_s + \widetilde{c}_s\right) \\ \frac{M_s}{N_s} &= \int_o^{c_s} \mu_s(c)dG_s(c) = \frac{W_s}{2}\left(c_s - \widetilde{c}_s\right) \\ \frac{Q_s}{N_s} &= \int_o^{c_s} q_s(c)dG_s(c) = \frac{\lambda_s W_s L}{2\gamma}\left(c_s - \widetilde{c}_s\right) \end{aligned} \qquad (17)$$

where $\widetilde{c}_s$ labels the average unit labor requirement of goods producers, i.e. the mean unit labor requirement calculated for the conditional distribution $G(c)/G(c_s)$ as only firms with $c \leq c_s$ produce. Analogously, average revenues and dividends evaluate to

$$\frac{R_s}{N_s} = \int_o^{c_s} r_s(c)dG_s(c) = \frac{\lambda_s L (W_s)^2}{4\gamma}\left((c_s)^2 - (\widetilde{c}_s)^2 - \widetilde{\sigma}_s^2\right) \qquad (18)$$

$$\frac{D_s}{N_s} = \int_o^{c_s} \pi_s(c)dG_s(c) = \frac{\lambda_s L (W_s)^2}{4\gamma}\left((c_s - \widetilde{c}_s)^2 + \widetilde{\sigma}_s^2\right) \qquad (19)$$

as $\widetilde{\sigma}_s^2 + (\widetilde{c}_s)^2 = \int_0^{c_s} c^2 dG_s(c)$ with $\widetilde{\sigma}_s^2$ denoting the conditional variance. Note that, in the above expressions, the conditional mean $\widetilde{c}_s$ and variance $\widetilde{\sigma}_s^2$ are both functions of c_s only.[2]

Finally, (14), (15) and (10) imply the zero cutoff profit condition

$$K_s = \frac{2\gamma f}{\eta} \frac{\alpha - \lambda_s W_s c_s}{\rho_s \lambda_s W_s \left(c_s - \widetilde{c}_s\right)} \qquad (20)$$

[2] Average revenue $\widetilde{r}_s = R_s/N_s$ and average dividend $\widetilde{d}_s = D_s/N_s$ differ from the revenue and profit of the average firm due to additive terms that depend on the variance $\widetilde{\sigma}_s^2$.

which shows that $K_s > 0$ requires $\alpha > \lambda_s W_s c_s$. All the rest given, a larger number of producers (larger $\rho_s K_s/f$) is associated with tougher competition (lower c_s).

3.3 Capital Accumulation and Aggregation

Perfect competition in capital production implies that capital is priced at marginal cost:

$$V_s = W_s f_I \tag{21}$$

while depreciation implies that the capital stock follows the law of motion

$$K_{s+1} = (1-\delta)\left(K_s + \frac{I_s}{V_s}\right) \tag{22}$$

where I_s/V_s is labor employed in capital accumulation.

Investment I_s can be obtained by aggregating the individual dynamic budget constraint

$$B_{s+1}y^c_{s+1} + \frac{V_s K_{s+1} x^c_{s+1}}{1-\delta} - V_s K_s x^c_s + \int_0^{N_s} p_s(\omega) q^c_s(\omega) d\omega = (1+i_s)\, B_s y^c_s + D_s x^c_s + W_s z$$

knowing that aggregate accounting implies $B_{t+1} = B_t = 0$, $\sum_c y^c_{t+1} = \sum_c y^c_t = 1$ and $\sum_c x^c_{t+1} = \sum_c x^c_t = 1$. Aggregation then gives

$$I_s = V_s \left(\frac{K_{s+1}}{1-\delta} - K_s\right) = W_s L z_s - (R_s - D_s)$$

which shows that investment equals the aggregate wage bill minus the wages paid to labor employed in goods production $(R_s - D_s)$. Equivalently, investment is what is left of wage income $W_s L z_s$ and dividend income D_s after paying for consumption expenditure R_s. Accordingly (22) can be rewritten as

$$K_{s+1} = (1-\delta)\left(K_s + \frac{W_s L z_s - (R_s - D_s)}{V_s}\right) \tag{23}$$

3.4 Welfare

Turning to welfare, instantaneous utility (2) can be rewritten as

$$U_s = \alpha Q^c_s - \frac{1}{2}\left(\eta + \frac{\gamma}{N_s}\right)(Q^c_s)^2 - \frac{1}{2}\gamma N_s \mathrm{Var}(q^c_s) \tag{24}$$

which highlights that individual welfare depends on total consumption (Q^c_s), product variety $(N_s = K_s \rho_s / f)$ and the concentration of the consumption bundle across available products (inversely measured by $\mathrm{Var}(q^c_s)$). As marginal utility decreases in the consumption of each product, more concentration (smaller

$\text{Var}(q_s^c)$) is bad for consumption. At the same time it is good for income through a productivity reallocation effect due to the corresponding concentration of resources on the supply of varieties by the most efficient firms. Expression (2) then shows that the latter effect dominates.

3.5 Parametrization of Technology

All the results derived so far hold for any distribution of unit labor requirement draws $G(c)$. However, in order to simplify some of the ensuing analysis, it is useful to introduce a specific and empirically relevant parametrization for this distribution. In particular, it is assumed that individual productivity draws $1/c$ follow a Pareto distribution with lower productivity bound $1/c_M$ and shape parameter $k \geq 1$. This implies a distribution of unit labor requirement draws c given by

$$G(c) = \left(\frac{c}{c_M}\right)^k, \ c \in [0, c_M]. \tag{25}$$

The shape parameter k indexes the dispersion of unit labor requirement draws. When $k = 1$, the unit labor requirement distribution is uniform on $[0, c_M]$ with maximum dispersion. As k increases, the relative number of high unit labor requirement firms increases, and the unit labor requirement distribution becomes more concentrated at these higher unit labor requirement levels. As k goes to infinity, the distribution becomes degenerate at c_M and dispersion vanishes. Hence, k can be interpreted as an inverse measure of firm heterogeneity. Any truncation of the unit labor requirement distribution from above retains the same distribution function and shape parameter k. The productivity distribution of surviving firms is therefore also Pareto with shape k, and the truncated unit labor requirement distribution is given by $G_s(c) = (c/c_s)^k$, $c \in [0, c_s]$.

Given this distributional assumption, the fraction of entrants that produce, their average unit labor requirement and the variance of their unit labor requirements equal

$$\rho_s = \left(\tfrac{c_s}{c_M}\right)^k \quad \widetilde{c}_s = \tfrac{k}{k+1} c_s \quad \widetilde{\sigma}_s^2 = \tfrac{k}{(k+1)^2(k+2)} (c_s)^2 \tag{26}$$

which, together with (15), allows one to rewrite (20), (18) and (19) respectively as

$$K_s = \frac{2\gamma(k+1)(c_M)^k f}{\eta} \frac{\alpha - \lambda_s W_s c_s}{\lambda_s W_s (c_s)^{k+1}} \tag{27}$$

$$D_s = \frac{L}{2\gamma (k+1)(k+2)(c_M)^k f} \lambda_s (W_s)^2 (c_s)^{k+2} K_s \tag{28}$$

$$R_s = (k+1) D_s \tag{29}$$

Accordingly, employment in goods production is $L_s = (R_s - D_s)/W_s = k D_s / W_s$.

Moreover, average price, markup and output from (17) boil down to

$$\begin{array}{c}\widetilde{p}_s = \frac{P_s}{N_s} = \frac{2k+1}{2(k+1)} W_s c_s \\ \widetilde{\mu}_s = \frac{M_s}{N_s} = \frac{W_s c_s}{2(k+1)} \\ \widetilde{q}_s = \frac{Q_s}{N_s} = \frac{L}{2\gamma(k+1)} \lambda_s W_s c_s \end{array} \tag{30}$$

To sum up, at time s, the equilibrium of the model is characterized by seven conditions. Three are the dynamic conditions (8), (13), and (23). The other four are the static conditions (21), (27), (28) and (29). These can be combined to yield the following system of five equations

$$\begin{array}{c}\beta E_s \left\{(1+i_{s+1})\lambda_{s+1}\right\} = \lambda_s \\ \frac{W_{s+1}}{W_s} = \frac{1+i_{s+1}}{1-\delta} - \frac{R_{s+1}}{(k+1) f_I W_s K_{s+1}} \\ K_{s+1} - K_s = (1-\delta)\left(\frac{L z_s}{f_I} - \frac{k}{k+1}\frac{R_s}{f_I W_s}\right) - \delta K_s \end{array} \tag{31}$$

$$\begin{array}{c}R_s = \frac{L}{2\gamma(k+2)(c_M)^k f} \lambda_s \left(W_s\right)^2 \left(c_s\right)^{k+2} K_s \\ K_s = \frac{2\gamma(k+1)(c_M)^k f}{\eta} \frac{\alpha - \lambda_s W_s c_s}{\lambda_s W_s (c_s)^{k+1}} \end{array} \tag{32}$$

There are six endogenous variables (λ, i, W, R, K, c). The characterization of the equilibrium is completed by choosing an efficiency unit of labor as the numeraire good ($W_{s+1} = W_s = 1$).

Turning to welfare, instantaneous indirect utility has a neat expression. In particular, substituting the utility maximizing consumption choices into (2), given the profit maximizing prices and the individual budget constraint, gives

$$U_s = \frac{\lambda_s R_s}{L} + \frac{1}{2\eta}\left(\alpha - \lambda_s c_s\right)\left(\alpha - \frac{k+1}{k+2}\lambda_s c_s\right)$$

which, by (35) and (36), can be transformed into

$$U_s = \frac{1}{2\eta}\left(\alpha - \lambda_s c_s\right)\left(\alpha + \frac{k+1}{k+2}\lambda_s c_s\right) \tag{33}$$

Instantaneous utility U_s is, therefore, a decreasing function of λ_s and c_s.

4 Deterministic Steady State

In steady state $\lambda_{s+1} = \lambda_s = \overline{\lambda}$, $i_{s+1} = i_s = \overline{i}$, $K_{s+1} = K_s = \overline{K}$, $c_{s+1} = c_s = \overline{c}$, $z_s = \overline{z}$, where $\overline{z}$ is the mean value of z_s. Under these conditions, equations (31) determine the unique steady state values of the capital stock and revenues

$$\begin{array}{c}\overline{K} = \frac{1}{\frac{\delta}{1-\delta} + k\frac{1-\beta(1-\delta)}{\beta(1-\delta)}} \frac{L\overline{z}}{f_I} \\ \overline{R} = \frac{\frac{1-\beta(1-\delta)}{\beta(1-\delta)}(k+1)}{\frac{\delta}{1-\delta} + k\frac{1-\beta(1-\delta)}{\beta(1-\delta)}} L\overline{z} \end{array} \tag{34}$$

These imply steady state dividends and employment in goods production are $\overline{D} = \overline{R}/(k+1)$ and $\overline{L} = \overline{R} - \overline{D} = k\overline{D}/\overline{z}$ respectively. In steady state, the number of entrants is then equal to $\overline{K}/f$ while the number of producers evaluates to $\overline{N} = (\overline{c}/c_M)^k \overline{K}/f$, with $\overline{\rho} = (\overline{c}/c_M)^k$ being the success rate of entry. Results (34) also imply

$$\frac{\overline{R}}{\overline{K}} = \frac{1-\beta(1-\delta)}{\beta(1-\delta)}(k+1)f_I$$
$$\frac{\overline{D}}{\overline{K}} = \frac{1-\beta(1-\delta)}{\beta(1-\delta)}f_I$$

Given $\overline{K}$ and $\overline{W}$, equations (32) (implicitly) determine the unique steady-state cutoff unit labor requirement $\overline{c}$ and marginal utility of income $\overline{\lambda}$. To see this, rewrite (32) as

$$\overline{\lambda} = \frac{2\gamma(k+2)(c_M)^k f}{L} \frac{\overline{R}}{\overline{K}} \frac{1}{\overline{c}^{k+2}} \tag{35}$$

$$\overline{\lambda} = \frac{2\alpha\gamma(k+1)(c_M)^k f}{2\gamma(k+1)(c_M)^k f\overline{c} + \eta\overline{K}\overline{c}^{k+1}} \tag{36}$$

Both these expressions represent $\overline{\lambda}$ as positive decreasing functions of $\overline{c}$ with (35) everywhere steeper than (36). Given that the former lies above the latter in a neighbourhood of $\overline{c} = 0$, they must cross and this happens only once at some positive value of $\overline{c}$. This value belongs to the relevant support $[0, c_M]$ provided that c_M is large enough. The formal condition

$$c_M > \frac{k+2}{k+1} \frac{2\gamma(k+1)f + \eta\overline{K}}{\alpha L} \frac{\overline{R}}{\overline{K}} \tag{37}$$

grants existence and uniqueness of the steady state.[3]

As for welfare, in steady state instantaneous indirect utility (33) becomes

$$\overline{U} = \frac{1}{2\eta}\left(\alpha - \overline{\lambda}\overline{c}\right)\left(\alpha + \frac{k+1}{k+2}\overline{\lambda}\overline{c}\right) \tag{38}$$

Then, based on (1), steady state intertemporal indirect utility equals the present value of the constant flow (38) discounted at rate β.

As (32) do not lend themselves to explicit analytical solution, some comparative statics results around the steady state can be obtained graphically after rewriting (35) and (36) as follows

$$\overline{\lambda}\overline{c} = \frac{2\gamma(k+2)(c_M)^k f}{L} \frac{\overline{R}}{\overline{K}} \frac{1}{\overline{c}^{k+1}} \tag{39}$$

$$\overline{\lambda}\overline{c} = \frac{2\alpha\gamma(k+1)(c_M)^k f}{2\gamma(k+1)(c_M)^k f + \eta\overline{K}\overline{c}^{k}} \tag{40}$$

Figures 1 provides a graphical representation of the determination of $\overline{\lambda}\overline{c}$ and $\overline{c}$ for given $\overline{K}$ and $\overline{W}$, with (39) being the steeper curve and (40) being the

[3] Intuitively, the condition for existence and uniqueness of the steady state requires (35) to be below (36) at $\overline{c} = c_M$.

flatter one associating $\overline{\lambda}\overline{c} = \alpha$ to $\overline{c} = 0$. The fact that along the steeper curve (39) $\overline{\lambda}\overline{c}$ goes to infinity when $\overline{c}$ tends to zero confirms that there exist unique equilibrium values for $\overline{\lambda}\overline{c}$ and $\overline{c}$ (and therefore for $\overline{\lambda}$) provided that (37) holds.

The focus here is on the effects of a permanent shock to labor productivity. For concreteness, consider an exogenous increase in $\overline{z}$. The effect of lower $\overline{z}$ will be clearly symmetric. Figure 1 shows the effects of larger $\overline{z}$. The initial situation is represented by the two solid curves. Given (34), larger $\overline{z}$ drives $\overline{K}$ up while the ratio $\overline{R}/\overline{K}$ remains unchanged. This implies that, whereas (39) does not move, (40) shifts downwards to its new dashed position. As a result the equilibrium value of $\overline{\lambda}\overline{c}$ falls whereas the equilibrium value of $\overline{c}$ rises, thus reducing the toughness of competition. Accordingly, higher labor productivity raises the number of entrants ($\overline{K}/f$ increases) as well as the number of producers ($(\overline{c}/c_M)^k\,\overline{K}/f$ increases) due to both more entry (larger $\overline{K}/f$) and a higher survival rate for entrants (larger $(\overline{c}/c_M)^k$). Given (30), by raising $\overline{c}$, higher labor productivity is associated with higher average price and average markup as well as with lower average output. Given (38), by decreasing $\overline{\lambda}\overline{c}$ higher labor productivity is also associated with higher welfare.

5 Local Transitionary Dynamics

The local dynamic properties of the model can be analyzed through linearization around its non-stochastic steady state to show how the economy reacts to productivity shocks.

5.1 Log-Linearization

First, recalling that $W_s = 1$ due to the choice of numeraire and using the fact that $\left(X - \overline{X}\right)/\overline{X} \approx d\ln X$ around the steady state give the following linear approximation for the capital accumulation equation in (31)

$$d\ln K_{s+1} = \left(\delta + k\frac{1-\beta(1-\delta)}{\beta}\right) d\ln z_s - k\frac{1-\beta(1-\delta)}{\beta} d\ln R_s + (1-\delta)\, d\ln K_s \tag{41}$$

Second, under the assumption that $(1+i_s)\,\lambda_s$ is lognormally distributed, the log-linearization of the Euler equation in (31) around the steady state yields:

$$E_{s-1}\left\{\ln \lambda_s\right\} - \ln \lambda_{s-1} = -\left(E_{s-1}\left\{\ln\left(1+i_s\right)\right\} - \ln\left(1+\overline{i}\right)\right) - \chi_0$$

where $\chi_0 \equiv \frac{1}{2}\mathrm{Var}\{\ln\left(\left(1+i_s\right)\lambda_s\right)\}$ is a constant due to lognormality and $\ln\left(1+\overline{i}\right) = -\ln\beta$. The focus is on the dynamic response to shocks rather than on trend movements, so the constant χ_0 is omitted henceforth. In terms of deviations from steady state one can then rewrite

$$E_{s-1}\left\{d\ln \lambda_s\right\} - d\ln \lambda_{s-1} = -\left(1-\beta(1-\delta)\right)\left(E_{s-1}\left\{d\ln R_s\right\} - d\ln K_s\right) \tag{42}$$

where the interest rate has been substituted out using the following linear approximation of the no-arbitrage condition in (31) around the steady state

$$d\ln(1+i_s) = (1-\beta(1-\delta))\,(d\ln R_s - d\ln K_s)$$

Third, linearization of (32) around the steady state yields the system

$$\begin{aligned} d\ln R_s &= d\ln\lambda_s + (k+2)\,d\ln c_s + d\ln K_s \\ d\ln K_s &= -\left(\tfrac{\alpha}{\alpha-\overline{\lambda}\overline{c}} + k\right) d\ln c_s - \tfrac{\alpha}{\alpha-\overline{\lambda}\overline{c}} d\ln\lambda_s \end{aligned}$$

which can be solve by Cramer rule to obtain

$$\begin{aligned} d\ln c_s &= -\frac{\alpha}{(k+1)\alpha - k\overline{\lambda}\overline{c}} d\ln\lambda_s - \frac{\alpha - \overline{\lambda}\overline{c}}{(k+1)\alpha - k\overline{\lambda}\overline{c}} d\ln K_s \\ d\ln R_s &= -\frac{\alpha + k\overline{\lambda}\overline{c}}{(k+1)\alpha - k\overline{\lambda}\overline{c}} d\ln\lambda_s - \frac{\alpha - 2\overline{\lambda}\overline{c}}{(k+1)\alpha - k\overline{\lambda}\overline{c}} d\ln K_s \end{aligned} \tag{43}$$

The latter can then be substituted into (41) and (42) to yield

$$\begin{aligned} d\ln K_{s+1} - d\ln K_s &= \frac{1-\beta(1-\delta)}{\beta} \frac{k\left(\alpha + k\overline{\lambda}\overline{c}\right)}{(k+1)\alpha - k\overline{\lambda}\overline{c}} d\ln\lambda_s \\ &\quad + \left(\frac{1-\beta(1-\delta)}{\beta} \frac{k\left(\alpha - 2\overline{\lambda}\overline{c}\right)}{(k+1)\alpha - k\overline{\lambda}\overline{c}} - \delta\right) d\ln K_s \\ &\quad + \left(\delta + k\frac{1-\beta(1-\delta)}{\beta}\right) d\ln z_s \end{aligned} \tag{44}$$

$$\begin{aligned} E_{s-1}\{d\ln\lambda_s\} - d\ln\lambda_{s-1} &= \frac{(1-\beta(1-\delta))\,\frac{\alpha+k\overline{\lambda}\overline{c}}{(k+1)\alpha-k\overline{\lambda}\overline{c}}}{1-(1-\beta(1-\delta))\,\frac{\alpha+k\overline{\lambda}\overline{c}}{(k+1)\alpha-k\overline{\lambda}\overline{c}}} d\ln\lambda_{s-1} \\ &\quad + \frac{(1-\beta(1-\delta))\,\frac{(k+2)\left(\alpha-\overline{\lambda}\overline{c}\right)}{(k+1)\alpha-k\overline{\lambda}\overline{c}}}{1-(1-\beta(1-\delta))\,\frac{\alpha+k\overline{\lambda}\overline{c}}{(k+1)\alpha-k\overline{\lambda}\overline{c}}} d\ln K_s \end{aligned} \tag{45}$$

Equations (44) and (45) constitute a system of two linear stochastic difference equations in the percentage deviations of the capital stock ($d\ln K_s$) and of the marginal utility of income ($d\ln\lambda_s$) from steady state with the percentage productivity shock ($d\ln z_s$) as exogenous forcing variable. The system is closed by assuming the following specific process for this shock

$$d\ln z_s = \theta_z d\ln z_{s-1} + \xi_s \tag{46}$$

where ξ_s is independent normally distributed white noise disturbance with mean zero.

5.2 Stability

Forwarding (45) by one period and substituting for $d\ln K_{s+1}$ from (44) in the forwarded expression yields

$$\begin{aligned} &E_s\left\{d\ln\lambda_{s+1}\right\} - d\ln\lambda_s = \\ &\frac{(1-\beta(1-\delta))\frac{(k+2)\left(\alpha-\overline{\lambda}\overline{c}\right)}{(k+1)\alpha-k\overline{\lambda}\overline{c}}}{1-(1-\beta(1-\delta))\frac{\alpha+k\overline{\lambda}\overline{c}}{(k+1)\alpha-k\overline{\lambda}\overline{c}}}\left(1+\frac{1-\beta(1-\delta)}{\beta}\frac{k(k+2)\left(\alpha-\overline{\lambda}\overline{c}\right)}{(k+1)\alpha-k\overline{\lambda}\overline{c}}\right)\frac{\alpha+k\overline{\lambda}\overline{c}}{(k+2)\left(\alpha-\overline{\lambda}\overline{c}\right)}d\ln\lambda_s \\ &+\frac{(k+2)\left(\alpha-\overline{\lambda}\overline{c}\right)}{(k+1)\alpha-k\overline{\lambda}\overline{c}}\frac{1-\beta(1-\delta)}{\beta}d\ln K_s+\frac{(1-\beta(1-\delta))\frac{(k+2)\left(\alpha-\overline{\lambda}\overline{c}\right)}{(k+1)\alpha-k\overline{\lambda}\overline{c}}}{1-(1-\beta(1-\delta))\frac{\alpha+k\overline{\lambda}\overline{c}}{(k+1)\alpha-k\overline{\lambda}\overline{c}}}\left(\delta+k\frac{1-\beta(1-\delta)}{\beta}\right)d\ln z_s \end{aligned} \tag{47}$$

Abstracting from the random shocks (i.e. imposing $d\ln z_s = 0$), standard phase diagram analysis reveals that the dynamics of the system defined by (44) and (47) exhibit saddle path properties whenever the slope of the locus (K_s, λ_s) such that $d\ln K_{s+1} = d\ln K_s$ is larger than the slope of the locus (K_s, λ_s) such that $d\ln\lambda_{s+1} = d\ln\lambda_s$. This is always the case given that $\alpha > \overline{\lambda}\overline{c}$ has to hold for $\overline{K} > 0$ (see (20)).

5.3 Undetermined coefficients

The stochastic system can be solved by the method of undetermined coefficients conjecturing a solution of the form

$$d\ln\lambda_s = \phi_K d\ln K_s + \phi_z d\ln z_s$$

To see this, define the bundling parameters a_λ, a_k, a_z and b_λ, b_k, b_z that allow (44) and (47) to be rewritten as

$$\begin{aligned} d\ln K_{s+1} &= a_\lambda d\ln\lambda_s + (1+a_k)\,d\ln K_s + a_z d\ln z_s \\ E_s\left\{d\ln\lambda_{s+1}\right\} &= (1+b_\lambda)\,d\ln\lambda_s + b_k d\ln K_s + b_z d\ln z_s \end{aligned}$$

Substituting in the conjectured solution and the processes of the productivity shock (46) yields

$$\begin{aligned} d\ln K_{s+1} &= (1+a_k+a_\lambda\phi_K)\,d\ln K_s + (a_z+a_\lambda\phi_z)\,d\ln z_s \\ d\ln K_{s+1} &= \frac{b_k+(1+b_\lambda)\,\phi_K}{\phi_K}d\ln K_s + \frac{(b_z-\phi_z\theta_z)+(1+b_\lambda)\,\phi_z}{\phi_K}d\ln z_s \end{aligned}$$

Then the conjectured solution is indeed correct if its coefficients ϕ_K and ϕ_z jointly satisfy the following conditions

$$\begin{aligned} 1+a_k+a_\lambda\phi_K &= \frac{b_k+(1+b_\lambda)\,\phi_K}{\phi_K} \\ a_z+a_\lambda\phi_z &= \frac{(b_z-\phi_z\theta_z)+(1+b_\lambda)\,\phi_z}{\phi_K} \end{aligned} \tag{48}$$

These define a system of two equations in two unknowns, ϕ_K and ϕ_z. Using the values that solve this system, one can characterize the transitionary dynamics in a neighbourhood of the steady state forced by the productivity shock $d\ln z_s$. Then, the evolution of the capital stock and the marginal utility of income r periods after the shock can be described by

$$d\ln K_s = 0;\ d\ln K_{s+r} = \xi_s \left(a_z + a_\lambda \phi_z\right) e^{(a_k + a_\lambda \phi_K + \theta_z)(r-1)},\ r = 1, ..., \infty \quad (49)$$

$$d\ln \lambda_{s+r} = \phi_K d\ln K_{s+r} + \phi_z d\ln z_{s+r},\ r = 0, ..., \infty \quad (50)$$

with

$$d\ln z_s = \xi_s;\ d\ln z_{s+r} = \left\{ \begin{array}{cc} 0 & \text{if } \theta_z = 0 \\ \xi_s e^{(\theta_z - 1)r} & \text{if } \theta_z \neq 0 \end{array} \right.,\ r = 1, .., \infty$$

where $\theta_z = 0$ corresponds to a temporary shock

5.4 Numerical analysis

The aim of this section is not to perform a full-fledged calibration exercise. It is rather to propose a numerical exploration of how firm heterogeneity affects the propagation of exogenous technology shocks.

Firm heterogeneity is regulated by the shape parameter k of the Pareto distribution: the larger k, the smaller the degree of firm heterogeneity as the population of firms becomes increasingly concentrated at low efficiency levels. Figures 2 to 8 depict the effects of a 10% temporary increase in labor productivity ξ_s. The figures are drawn by simulating the model for the following parameter values: $\alpha = 10$, $\gamma = 10$, $\eta = 10$, $c_M = 10$, $f = 1$, $L = 100$, $\delta = 0.025$, $\beta = 0.99$, $f_I = 1$, $\overline{z} = 1$. The values of δ and β are borrowed from Bilbiie, Ghironi and Melitz (2007). They correspond to the interpretation of a time period as a quarter and have been chosen for expositionary concreteness. In all figures the effects of the productivity shock are described for three alternative values of the Pareto shape parameter, corresponding to $k = 2$, $k = 3$ and $k = 4$.

Formally, Figures 2, 3 and 4 are based on (49), (50) and (43) respectively. Figures 5 plots $d\ln \rho_s = kd\ln c_s$, as implied by (26). Figure 6 portrays $d\ln N_s = kd\ln c_s + d\ln K_s$, as implied by $N_s = (c_s/c_M)^k K_s$. The Herfindahl Index of concentration in Figure 7 is calculated as

$$H_s = N_s \int_0^{c_s} s_s\left(c\right)^2 dG_s(c) = \frac{2}{N_s}\frac{k+2}{k+4}$$

where $s_s\left(c\right) = r_s(c)/R_s$ is the market share of a firm with unit labor requirement c. Finally, Figure 8 represents

$$d\ln U_s = -\frac{\left(\alpha + 2(k+1)\overline{\lambda}\overline{c}\right)\overline{\lambda}\overline{c}}{\left(\alpha - \overline{\lambda}\overline{c}\right)\left(\alpha\left(k+2\right) + \left(k+1\right)\overline{\lambda}\overline{c}\right)}\left(d\ln \lambda_s + d\ln c_s\right)$$

as obtained by linearizing (33) around the steady state.

On impact (period 0 in the figures) the capital stock and, therefore, the number of entrants is fixed. All adjustment is then loaded onto the marginal utility of income that stimulates demand by falling below its steady state level. Higher demand raises profits, allowing additional less productive entrants to produce. The cutoff and the success rate rise accordingly. Even if the number of entrants is unchanged, the higher success rate implies a larger number of producers. As a result, market concentration falls and welfare rises. Higher profits also raise investment in the accumulation of new capital that will become available at the beginning of the next period.

In the first period after the shock (period 1 in the figures), newly accumulated capital is in place making additional entry possible. This feeds into additional producers, lower market concentration and higher welfare, even though additional entry pushes both the cutoff and the success rate below their steady state levels. At the same time, the marginal utility of income starts increasing towards its long run level.

In the subsequent periods the marginal utility of income keeps on increasing and demand keeps on falling. Producers then leave the market; the cutoff, the success rate and market concentration rise; the capital stock, the number of entrants and welfare fall back to their steady state values.

Comparing the curves corresponding to different values of k across the figures allows one to gauge the role played by firm heterogeneity. In all figures larger values of k are associated with fluctuations of smaller amplitude, revealing that more heterogeneity is associated with smaller effects of the exogenous productivity shocks. Accordingly, existing studies finding that with representative firms procyclical entry and exit in imperfectly competitive markets can act as a powerful propagation mechanism of technology shocks (see, e.g., Jaimovich and Floetotto, 2008, and Bilbiie, Ghironi and Melitz, 2007) may actually overstate the role of entry and exit as an amplification mechanism during the business cycle. The reason is that procyclical entry and countercyclical exit disproportionately involve low efficiency firms, whose importance for the aggregate economy is limited by their small size. More generally, these findings suggest that the role of entry and exit as a propagation mechanism of technology shocks cannot be fully assessed without simultaneously considering firm heterogeneity.

As a final comment, it is interesting to stress the evolution of the average markup, average firm efficiency and CPI-deflated wage per worker. Both the average markup and average firm efficiency change one to one with the cutoff c_s. However, whereas the former increases when the cutoff increases, the latter decreases when the cutoff increases. As the cutoff moves procyclically on impact, countercyclically after one period, and procyclically again afterwards, also the average markup follows the same evolution. By symmetry, average efficiency moves instead countercyclically on impact, procyclically after one period and countercyclically again afterwards. Turning to wage per worker, its nominal value is $W_s z_s$. The CPI-deflator is defined as $CPI_s = N_s \int_0^{c_s} p_s(c) s_s(c)\, dG_s(c)$ with $s_s(c) = r_s(c)/R_s$. Under the Pareto assumptions, it boils down to $CPI_s = c_s \left[6k + 2k^2 + 3\right] / \left[2(k+1)(k+3)\right]$. Hence, the percentage deviation of the CPI-deflated real wage $W_s z_s / CPI_s$ from steady state evaluates locally to $d \ln z_s - d \ln c_s = \xi_s - d \ln c_s$ upon impact and $-d \ln c_s$

afterwards. Its evolution is depicted in Figure 9, which shows that the real wage moves procyclically on impact and countercyclically afterwards. These alternating responses due to compositional changes in the presence of heterogeneous imperfectly competitive firms may help explain why the empirical debate on the cyclical properties of markups and real wages is still pretty much open.

6 Conclusion

This paper has Introduced firm heterogeneity in a standard two-sector growth model to study the propagation of exogenous technology shocks. Focusing on how the proposed model reacts changes in the number of efficiency units per worker has revealed that, during business cycle upswings, on impact there is more entry, more survival after entry, and surviving firms are on average less efficient and smaller. The opposite is true during downswings. This is due to the fact that new producers and quitters are less efficient than incumbents. The resulting pro-cyclical entry and the countercyclical exit of less efficient firms reduce the impact of exogenous technology shocks on aggregate output per worker and welfare. Due to variable demand elasticity, the dampening effect of firm selection on aggregate productivity and welfare works also through a second channel. Holding the number of incumbents constant, in an upswing industry concentration decreases as market shares are reallocated towards less efficient firms due to the fact that the elasticity of demand falls more for high-price firms than for low-price ones. In a downswing the opposite happens.

It has been shown that the dampening effect of firm selection depends on the degree of firm heterogeneity, with the impact of changes in the number of efficiency units per worker on aggregate output per worker and welfare being stronger the less heterogeneous firms are. This reveals a way through which microeconomic heterogeneity may crucially affect the propagation of aggregate shocks. Specifically, it implies that existing models with representative firms may overstate the actual role of procyclical entry and exit in imperfectly competitive markets as a propagation mechanism of technology shocks. The reason is that pro-cyclical entry and countercyclical exit disproportionately involve less efficient firms whose impact on the aggregate is hampered by their smaller size.

These finding need not be taken to imply that, as inefficient firms are typically small, variations in their number are of limited practical relevance. Indeed, variations in the number of firms are only one of the channels that generate actual changes in the number of competitors. In particular, a new establishment or franchise by an existing firm increase the number of competitors without affecting the number of active firms. What the findings do imply is, instead, that the role of entry and exit as a propagation mechanism of technology shocks cannot be fully assessed without simultaneously considering firm heterogeneity.

The analysis has focused on a numerical investigation of transitionary dynamics that falls short of a full-fledged calibration exercise. It would be natural to explore the ability of the model to reproduce key moments of the data. It would also be interesting to extend the model to the open economy in order to study the international transmissions of country specific shocks.

References

[1] Abraham, K., and J. Haltiwanger (1995) Real Wages and the Business Cycle, *Journal of Economic Literature* 33, 1215-1264.

[2] Bernard, A., B. Jensen, S. Redding and P. Schott (2007) Firms in International Trade, *Journal of Economic Perspective*s 21, 105–130.

[3] Bilbiie, F., F. Ghironi and M. Melitz (2007) Endogenous Entry, Product Variety, and Business Cycles, NBER Working Paper No. 13646.

[4] Broda and Weinstein (2007) Product Creation and Destruction:Evidence and Price Implications, NBER Working Paper No. 13041.

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

[6] Davis, S., and J. Haltiwanger (1990) Gross Job Creation and Destruction: Microeconomic Evidence and Macroeconomic Implications, in O. Blanchard and S. Fischer, eds., *NBER Macroeconomics Annual*, MIT Press, Cambridge MA, 123-168.

[7] Davis, S., and J. Haltiwanger (1992) Gross Job Creation, Gross Job Destruction, and Employment Reallocation, *Quarterly Journal of Economics* 107, 818-863.

[8] Davis, S., R. Faberman and J. Haltiwanger (2006) The Flow Approach to Labor Markets: New Data Sources and Micro-Macro Links, *Journal of Economic Perspectives* 20, 3-26.

[9] Gali J. (1994).Monopolistic Competition, Business Cycles, and the Composition of Aggregate Demand, *Journal of Economic Theory* 63,73–96.

[10] Ghironi, F., and M. Melitz (2005) International Trade and Macroeconomic Dynamics with Heterogeneous Firms, *Quarterly Journal of Economics* 120, 865-915.

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

[12] Jaimovich, N., and M. Floetotto (2008) Firm Dynamics, Markup Variations,and the Business Cycle, *Journal of Monetary Economics* 55,1238–1252.

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

[14] Melitz, M. and G. Ottaviano (2008) Market Size, Trade and Productivity, *Review of Economic Studies* 75, 295-316.

[15] Mortensen, D., and C. Pissarides (1994) Job Creation and Job Destruction in the Theory of Unemployment, *Review of Economic Studies* 61, 397-415.

[16] Neary, P. (2007) Cross-Border Mergers as Instruments of Comparative Advantage, *Review of Economic Studies* 74, 1229-1257.

[17] Rotemberg, J. and M. Woodford (1991).Markups and the Business Cycle,.*NBER Macroeconomics Annual*, MIT Press, Cambridge MA, 63–129.

[18] Rotemberg, J. , and M. Woodford (1992) Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity, *Journal of Political Economy* 100, 1153–1207.

[19] Rotemberg, J., and M. Woodford (1995) Dynamic General Equilibrium Models with Imperfectly Competitive Product Markets, in T. Cooley, ed., *Frontiers of Business Cycle Research*, Princeton University Press, Princeton, 243–293.

[20] Rotemberg, J., and M. Woodford (1999) The Cyclical Behavior of Prices and Costs, in J. Taylor and M. Woodford, eds., *Handbook of Macroeconomics*, vol. 1B, North-Holland, Amsterdam, 1051-1135.

[21] Tybout, J. (2003) Plant- and Firm-Level Evidence on "New" Trade Theories, in E. Choi and J. Harrigan, eds., *Handbook of International Economics*, Vol. 38, Blackwell, Oxford.

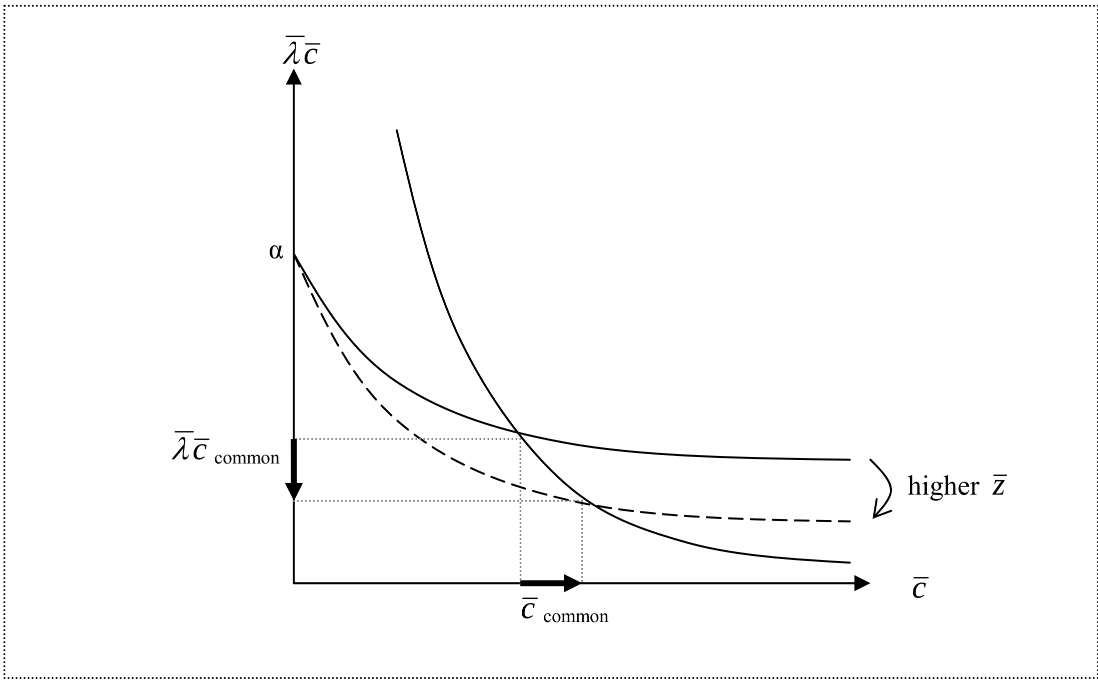

Figure 1 - Higher Labor Productivity:
Comparative Statics around the Steady State

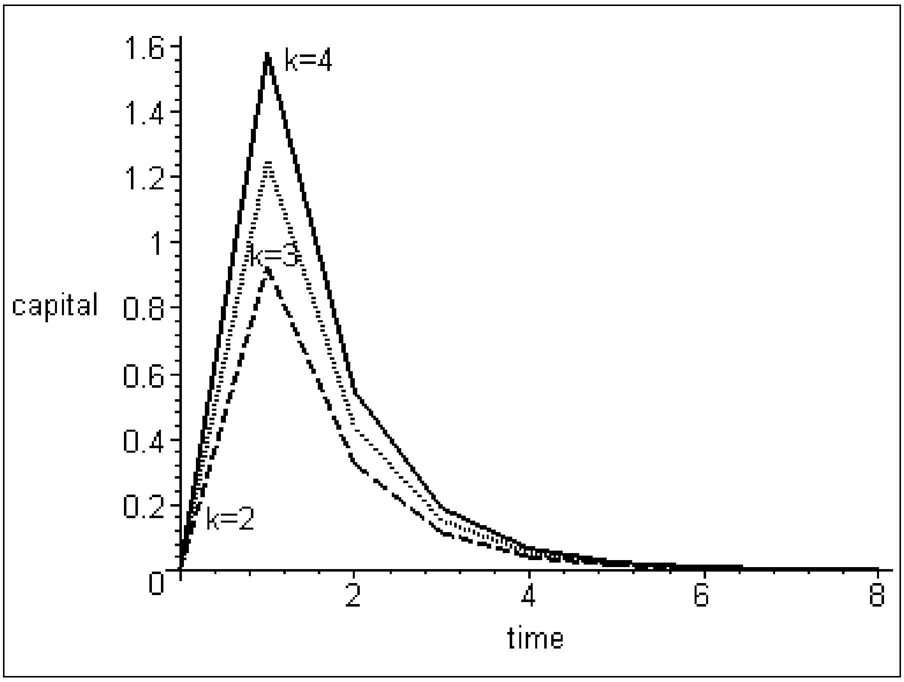

Figure 2 – Transitory Positive Labor Productivity Shock:
Response of the Capital Stock (Entry)

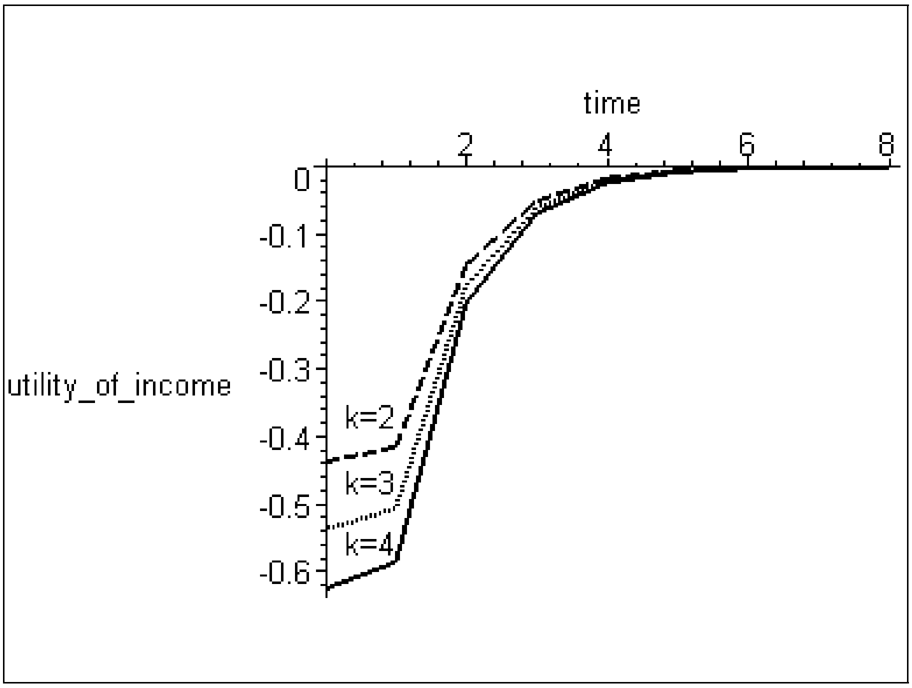

Figure 3 – Transitory Positive Labor Productivity Shock: Response of the Marginal Utility of Income

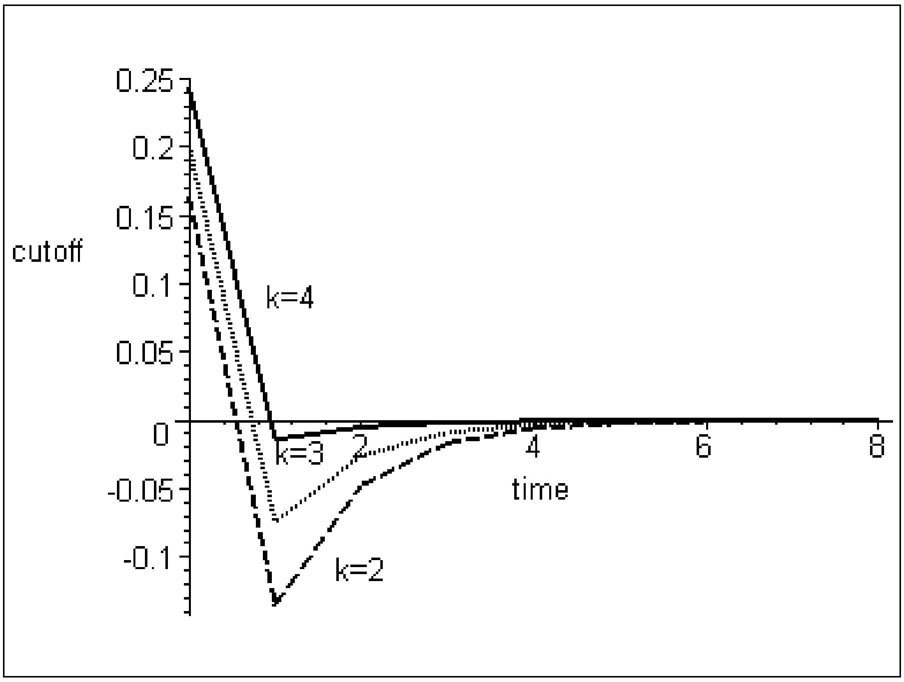

Figure 4 – Transitory Positive Labor Productivity Shock: Response of Cutoff (Markup)

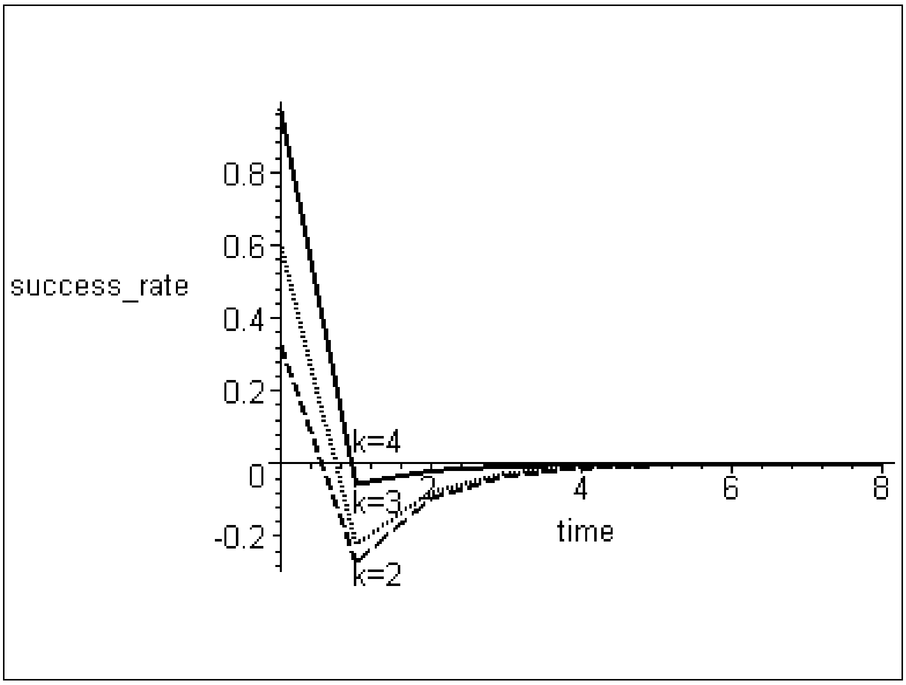

Figure 5 – Transitory Positive Labor Productivity Shock: Response of the Success Rate of Entry

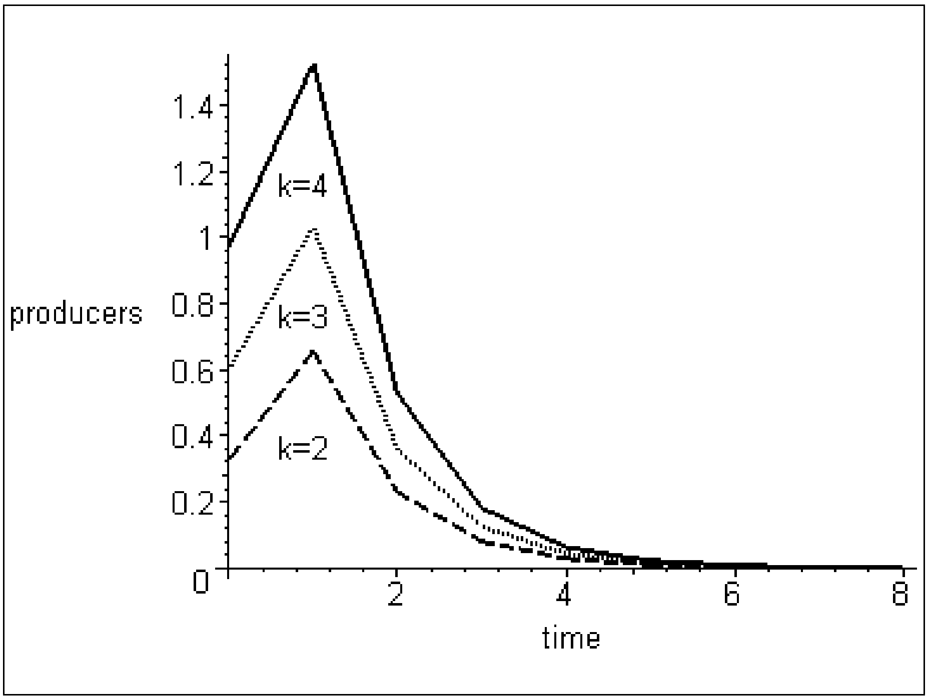

Figure 6 – Transitory Positive Labor Productivity Shock: Response of the Number of Producers

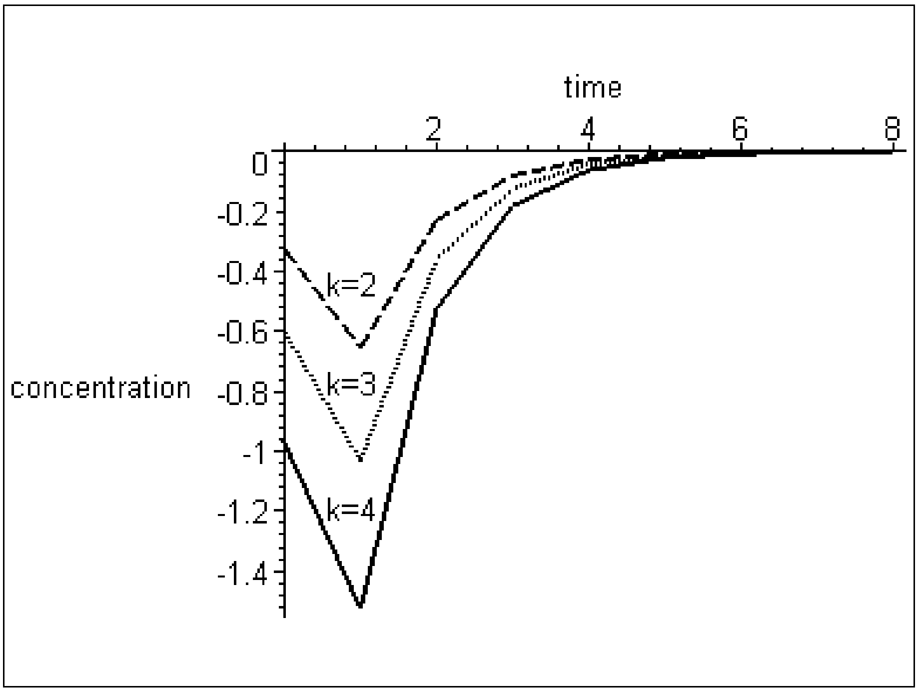

Figure 7 – Transitory Positive Labor Productivity Shock: Response of Market Concentration (Herfindahl Index)

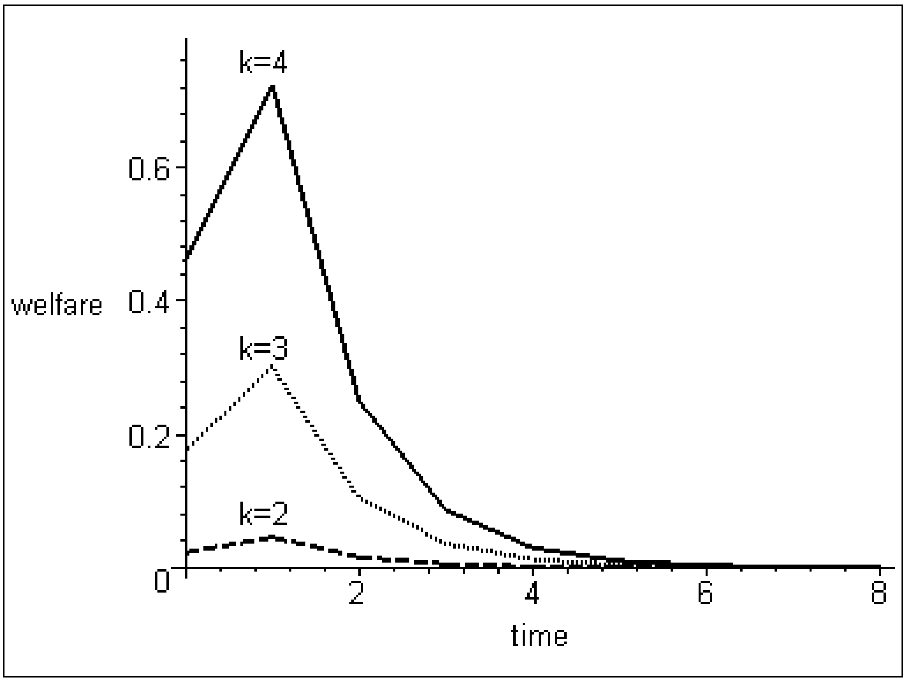

Figure 8 – Transitory Positive Labor Productivity Shock: Response of the Instantaneous Utility

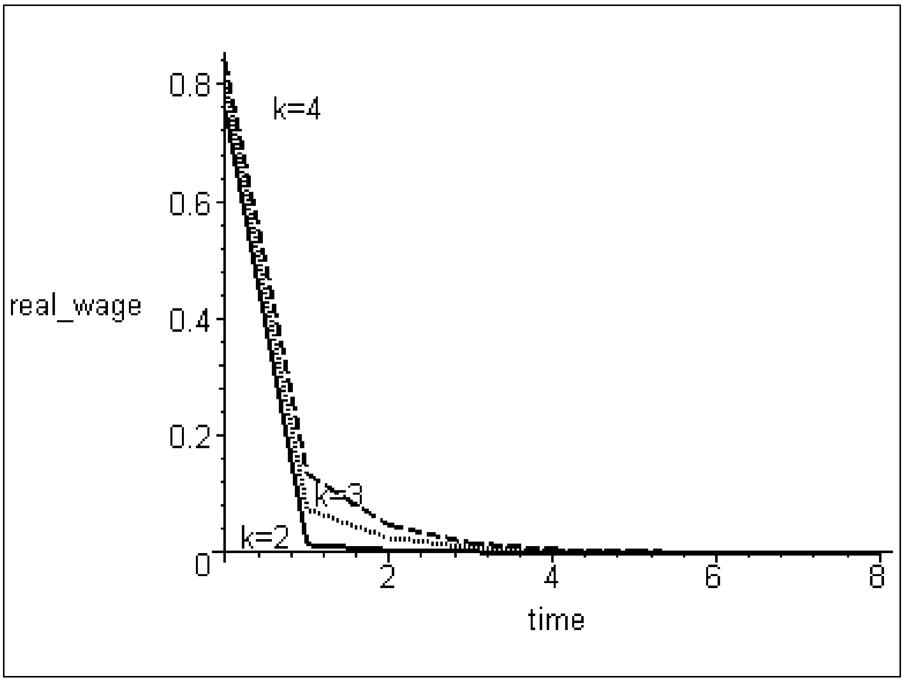

Figure 9 – Transitory Positive Labor Productivity Shock: Response of the Real Wage