Pricing-to-market, Intra-Industry Reallocations and Macroeconomic Dynamics

Gianmarco I.P. Ottaviano, Bocconi University and CEPR

VERY PRELIMINARY AND INCOMPLETE.
DO NOT CITE WITHOUT PERMISSION

June 10, 2011

Abstract

This paper presents a theory of endogenous aggregate productivity through firm selection. The focus is on the determination of aggregate output per worker when firms with heterogenous efficiency react in terms of entry and exit to exogenous changes in labor productivity. The main result is that higher labor productivity makes survival easier for a larger number of less efficient firms. Accordingly, both the number of entrants and the number of producers increase. As the average firm efficiency decreases due to a composition effect, higher labor productivity is associated with higher average price and average markup as well as with lower average output. These findings carry over to the open economy. Crucially, however, the selection effect is stronger in open economy than in closed economy and its strength increases as trade gets freer and the scope for price discrimination, and therefore pricing-to-market, gets smaller.

Keywords: aggregate productivity, firm selection, pricing-to-market, trade liberalization.

J.E.L. Classification: F4, F12.

1 Introduction

The impact of recent models with heterogenous firms on international trade theory and empirics has been huge. Their hallmark result is that that trade liberalization has important effects on aggregate productivity and thus on welfare through firm selection. Earlier works have focused on single-product firms highlighting the aggregate productivity gains stemming from the expulsion of less productive firms from the market as trade gets freer (Melitz, 2003; Bernard, Eaton, Kortum and Schott, 2003). Later contributions have also stressed the aggregate productivity gains deriving from the rationalization of product lines by multiproduct firms (Eckel and Neary, 2010; Bernard, Redding and Schott, 2010). Most contributions have focused on CES demand systems highlighting firm exit and product demise as the key sources of aggregate productivity changes. On the other hand, contributions featuring variable demand elasticity have argued that aggregate productivity gains do not necessarily require the exit of less efficient firms and products. Indeed, aggregate productivity may increase even for a given population of firms and products thanks to the reallocation of productive resources away from less efficient firms and products (Melitz and Ottaviano, 2008; Mayer, Melitz and Ottaviano, 2011).

Notwithstanding the large body of empirical evidence in favor of the relevance of selection effects in inducing endogenous aggregate productivity changes, the argument has had so far limited impact on the macroeconomic literature. On the one hand, this may be due to the irrelevance results inherent in the CES approach, according to which aggregate trade balance measures are sufficient statistics for trade gains, no matter whether selection or some other mechanism generate them (Arkolakis, Costinot and Rodriguez-Clare, 2010). On the other hand, the static models with selection proposed by trade economists are rather distant from those of interest to macroeconomists.

There are very few exceptions. Ghironi and Melitz (2005) propose a dynamic version of Melitz (2003) to argue that firm selection into export status makes the tradeability of goods endogenous and this may help explaining the persistence of deviations from purchasing power parity. Endogenous tradeability is also at the centre of Bergin and Glick (2003), who discuss the role of heterogenous trade costs in accounting for the behavior of international relative prices. Atkenson and Burstein (2008) embed a model of imperfect competition and variable markups in some of the recently developed quantitative models of international trade to examine whether such models can reproduce the main features of the fluctuations in international relative prices. They also discuss how price discrimination ("pricing-to-market") depends on the presence of international trade costs and various features of market structure. Building on the insight that export participation decisions alter the comovement of net exports with the real exchange rate, Alessandria and Choi (2007) propose a model whose business cycle exporter dynamics are consistent with that of U.S. exporters.

A common feature of these contributions is that they focus on the exporting decisions of a given number of incumbent firms. As entry and exit are blocked, the impact of firm selection on productivity is completely neutralized. Hence, while tradeability is endogenous in those models and reacts to exogenous productivity shocks, aggregate productivity itself remains exogenously determined.

The aim of the present paper is to propose a theory of endogenous aggregate productivity through firm selection. The focus is on the determination of output per worker when firms with heterogenous efficiency react in terms of entry and exit to exogenous changes in efficiency units per worker. As a result, aggregate productivity (measured as aggregate output per worker) is jointly determined by exogenous variations in labor productivity (measured as efficiency units per worker) and selection-driven endogenous reactions in firm productivity (measured as firm output per worker). The theory is crucially based on the assumption of variable demand elasticity and endogenous markups as in Melitz and Ottaviano (2008). It therefore has implications for the role of pricing-to-market in translating labor productivity shocks into aggregate productivity when firm productivity varies endogenously. The assumption of variable elasticity is not only empirically grounded but it also insulates the theory from the irrelevance results pointed out by Arkolakis, Costinot and Rodriguez-Clare (2010) in the case of constant demand elasticity.

The theory proposes a two-sector growth model in the spirit of Grossman and Helpman (1991). 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 in terms of unit labor requirements. Heterogeneity is itself endogenous due to selection as in Melitz (2003). In particular, to enter the market firms have to hire the required fixed amount of capital. After paying the corresponding rental price, they draw their unit labor requirements from some common probability distribution. Then, knowing their own labor productivity as well as 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 low enough unit labor requirements become producers; all other entrants leave the market without even starting production.

On the demand side, the proposed model borrows its instantaneous utility from the static setup of Melitz and Ottaviano (2008). However, by removing the linear component of their quasi-linear quadratic utility, it crucially introduces income effects and variable marginal utility of income as in Neary (2007). The demand system maintains the property of variable elasticity, implying that less productive firms (i.e. firms with higher unit labor requirements) face higher demand elasticity. Accordingly, they quote lower markups. As this is not enough to compensate their inefficiency, they quote higher prices and are smaller in terms of output, revenues, and profits. All these implications comply with the empirical evidence collected by the trade literature.

Though transitionary dynamics are fully characterized, the focus is on steady state and on how aggregate output per worker reacts to permanent changes in labor productivity. The main result is that higher labor productivity makes survival easier for a larger number of less productive firms. Accordingly, both the number of entrants and the number of producers increase. As these producers are less productive, higher labor productivity (i.e. more efficiency units per worker) is associated with higher prices and markups as well as with smaller output per firm.

A way to read these findings from business cycle point of view is that, during upswings in labor productivity, there is more entry and more survival after entry. Surving firms are, however, on average less efficient and smaller. The opposite is true during downswings. Hence, the positive impact of higher labor productivity on aggregate output per worker is reduced by the pro-cyclical entry of less efficient firms. Due to variable demand elasticity, such a stabilizing effect of firm selection is reinforced by the fact that, holding the number of active firms constant, during an upswing market shares are reallocated towards less efficient firms as demand elasticity falls more for high-price firms than for low-price ones. In an downswing the opposite happens. This reallocation of market shares for a given number of firms would be muted if demand exhibited constant elasticity.

It is also interesting to point out that the stabilizing effect of firm selection depends on the degree of firm heterogeneity: the impact of labor productivity on aggregate productivity is stronger the more heterogeneous firms are. This reveals a way through which microeconomic heterogeneity may crucially affect macroeconomic performance as this is determined not only by the first moment but also by higher moments of the distribution of firms across productivity levels. This would not be true in the case of CES demand.

These findings, initially derived for a closed economy, carry over to an open economy. For simplicity, this is shown in the case of two identical countries. The main insights is that the stabilizing selection effect is stronger in open economy than in closed economy, and its strength increases as trade gets freer. The reason lies in the increase in demand elasticity brought by trade liberalization, together with a reduced scope for price discrimination and pricing-to-market.

The rest of the paper is organized in two sections. The first is devoted to the closed economy. The second deals with the open economy. An additional section presents some concluding remarks.

2 Closed Economy

2.1 Endowments

There are L identical workers each supplying z units of labor inelastically every period. Accordingly, $L = Lz$ is the number of units of labor available each period and z can be interpreted as an aggregate labor productivity parameter. 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 utility function

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

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

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

with N_s and $q_s^c(\omega)$ denoting the measure ("number") of available products and the individual consumption level of product ω respectively. 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

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

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 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, 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 random draws 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

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

2.4 Consumption and investment

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

$$\sum_{s=t}^{\infty}\beta^{s-t}\left[\alpha\int_{0}^{N_s} q_s^c(\omega)d\omega-\frac{1}{2}\gamma\int_{0}^{N_s}\left(q_s^c(\omega)\right)^2 d\omega-\frac{1}{2}\eta\left(\int_{0}^{N_s} q_s^c(\omega)d\omega\right)^2\right]$$
$$-\lambda\left[\sum_{s=t}^{\infty}F_{t,s}\left(\frac{V_sK_{s+1}x_{s+1}^c}{1-\delta}-V_sK_sx_s^c+\int_0^{N_s}p_s(\omega)q_s^c(\omega)d\omega\right)\right.$$
$$\left.-(1+i_s)B_ty_t-\sum_{s=t}^{\infty}F_{t,s}\left(D_sx_s^c+W_sz\right)\right]$$

2.4.1 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

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

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

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

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

$$Q_s^c=\frac{N_s\alpha-\lambda_sP_s}{\gamma+\eta N_s}$$

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

Substituting this expression in (7) gives

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

Hence, products priced above the choke price

$$p_s = \frac{\frac{\alpha}{\lambda_s}\gamma + \eta P_s}{\gamma + \eta N_s} \tag{9}$$

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

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

$$p_s(\omega) = p_s - \frac{\gamma}{\lambda_s L} q_s(\omega) \tag{10}$$

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:

$$\left|\frac{dq_s(\omega)}{dp_s(\omega)}\frac{p_s(\omega)}{q_s(\omega)}\right| = \left(\frac{p_s}{p_s(\omega)} - 1\right)^{-1} = \left(\frac{p_s}{p_s - \frac{\gamma}{\lambda_s L} q_s(\omega)} - 1\right)^{-1} \tag{11}$$

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 also that the impact of changing p_s is stronger for lower $p_s(\omega)$.

2.4.2 Investment Decision

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

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

which, by (5), can be rewritten as

$$1 + i_{s+1} = (1-\delta)\left(\frac{V_{s+1}}{V_s} + \frac{D_{s+1}}{V_s K_{s+1}}\right) \tag{12}$$

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

2.5 Goods Production and Dividends

Profit maximization in goods production requires marginal revenue to match marginal cost. Given total inverse demand (10), the FOC for profit maximization in period s by a firm with unit labor requirement c implies output

$$q_s(c) = \frac{\lambda_s L}{2\gamma}\left(p_s^* - W_s c\right)$$

This uniquely identifies a cutoff unit labor requirement

$$c_s = \frac{p_s}{W_s} \tag{13}$$

such $q_s(c_s) = 0$ so that 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

$$N_s = \rho_s \frac{K_s}{f} \tag{14}$$

Expression (13) can be used to rewrite output as

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

which can be plugged into total inverse demand (10) to obtain the corresponding price, markup, revenue and profit:

$$\begin{array}{rlrl} p_s(c) = & \frac{W_s}{2}\left(c_s + c\right) & \mu_s(c) = & \frac{W_s}{2}\left(c_s - c\right) \\ r_s(c) = & \frac{\lambda_s L (W_s)^2}{4\gamma}\left((c_s)^2 - (c)^2\right) & \pi_s(c) = & \frac{\lambda_s L (W_s)^2}{4\gamma}\left(c_s - c\right)^2 \end{array} \tag{15}$$

Profit is equally shared as dividend 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 (15), average price, average markup and average output evaluate to

$$\begin{aligned} \frac{P_s}{N_s} &= \int_o^{c_s} p_s(c)dG_s^*(c) = \frac{W_s}{2}\left(c_s + \widetilde{c}_s\right) \\ \frac{M_s}{N_s} &= \int_o^{c_s} \mu_s(c)dG_s^*(c) = \frac{W_s}{2}\left(c_s - \widetilde{c}_s\right) \\ \frac{Q_s}{N_s} &= \int_o^{c_s} q_s(c)dG_s^*(c) = \frac{\lambda_s W_s L}{2\gamma}\left(c_s - \widetilde{c}_s\right) \end{aligned} \tag{16}$$

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

$$\frac{R_s}{N_s} = \int_o^{c_s} r_s(c) dG_s^*(c) = \frac{\lambda_s L (W_s)^2}{4\gamma} \left((c_s)^2 - (\widetilde{c}_s)^2 - \widetilde{\sigma}_s^2 \right) \tag{17}$$

$$\frac{D_s}{N_s} = \int_o^{c_s} \pi_s(c) dG_s^*(c) = \frac{\lambda_s L (W_s)^2}{4\gamma} \left((c_s - \widetilde{c}_s)^2 + \widetilde{\sigma}_s^2 \right) \tag{18}$$

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, (13), (14) and (9) imply the zero cutoff profit condition

$$K_s = \frac{2\gamma f}{\eta} \frac{\alpha - \lambda_s W_s c_s}{\rho_s \lambda_s W_s (c_s - \widetilde{c}_s)} \tag{19}$$

All the rest given, a larger number of producers (larger $\rho_s K_s / f$) is associated with tougher competition (lower c_s).

2.6 Capital Accumulation and Aggregation

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

$$V_s = W_s f_I \tag{20}$$

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

$$K_{s+1} = (1 - \delta) \left(K_s + \frac{I_s}{V_s} \right) \tag{21}$$

where I_s / V_s is labor employed in capital accumulation.

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

$$B_{s+1} y_{s+1}^c + \frac{V_s K_{s+1} x_{s+1}^c}{1 - \delta} - V_s K_s x_s^c + \int_0^{N_s} p_s(\omega) q_s^c(\omega) d\omega = (1 + i_s) B_s y_s^c + D_s x_s^c + W_s z$$

knowing that aggregate accounting implies $B_{t+1} = B_t = 0$, $\sum_c y_{t+1}^c = \sum_c y_t^c = 1$ and $\sum_c x_{t+1}^c = \sum_c x_t^c = 1$. Aggregation then gives

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

[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 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 $W_s Lz$ and dividend income D_s after paying for consumption expenditure R_s. Accordingly (21) can be rewritten as

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

2.7 Parametrization of Technology

All the results derived so far hold for any distribution of cost 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

$$G(c) = \left(\frac{c}{c_M}\right)^k, \ c \in [0, c_M]. \tag{23}$$

The shape parameter k indexes the dispersion of unit labor requirement draws. When $k = 1$, the unit labor requirement distribution is uniform on $[0, c_M]$. As k increases, the relative number of high unit labor requirement firms increases, and the unit labor requirement distribution is more concentrated at these higher unit labor requirement levels. As k goes to infinity, the distribution becomes degenerate at c_M. 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

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

which, together with (14), allows one to rewrite (19), (17) and (18) respectively as

$$K_s = \frac{2\gamma(k+1)\left(c_M\right)^k f}{\eta} \frac{\alpha - \lambda_s W_s c_s}{\lambda_s W_s \left(c_s\right)^{k+1}} \tag{24}$$

$$D_s = \frac{L}{2\gamma\left(k+1\right)\left(k+2\right)\left(c_M\right)^k f} \lambda_s \left(W_s\right)^2 \left(c_s\right)^{k+2} K_s \tag{25}$$

$$R_s = (k+1) D_s \tag{26}$$

Accordingly, employment in goods production is $L_s = (R_s - D_s)/W_s = kD_s/W_s$. Moreover, average price, markup and output from (16) boil down to

$$\begin{aligned} \widetilde{p}_s &= \tfrac{P_s}{N_s} = \tfrac{2k+1}{2(k+1)} W_s c_s \\ \widetilde{\mu}_s &= \tfrac{M_s}{N_s} = \tfrac{W_s c_s}{2(k+1)} \\ \widetilde{q}_s &= \tfrac{Q_s}{N_s} = \tfrac{L}{2\gamma(k+1)} \lambda_s W_s c_s \end{aligned} \tag{27}$$

2.8 Summary of Equilibrium Conditions

At time s, the equilibrium of the model is characterized by seven conditions. Three are the dynamic conditions (8), (12), and (22). The other four are the static conditions (20), (24), (25) and (26). These can be combined to yield the following system of five equations

$$\begin{aligned} \tfrac{\lambda_{s+1}}{\lambda_s} &= \tfrac{1}{\beta(1+i_{s+1})} \\ \tfrac{W_{s+1}}{W_s} &= \tfrac{1+i_{s+1}}{1-\delta} - \tfrac{R_{s+1}}{(k+1) f_I W_s K_{s+1}} \\ K_{s+1} - K_s &= (1-\delta)\left(\tfrac{Lz}{f_I} - \tfrac{k}{k+1} \tfrac{R_s}{f_I W_s}\right) - \delta K_s \end{aligned} \tag{28}$$

$$\begin{aligned} R_s &= \tfrac{L}{2\gamma(k+2)(c_M)^k f} \lambda_s (W_s)^2 (c_s)^{k+2} K_s \\ K_s &= \tfrac{2\gamma(k+1)(c_M)^k f}{\eta} \tfrac{\alpha - \lambda_s W_s c_s}{\lambda_s W_s (c_s)^{k+1}} \end{aligned} \tag{29}$$

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 ($W_{s+1} = W_s = 1$).

2.9 Steady State

In steady state $\lambda_{s+1} = \lambda_s = \overline{\lambda}$, $i_{s+1} = i_s = \overline{i}$, $K_{s+1} = K_s = \overline{K}$, $c_{s+1} = c_s = \overline{c}$. Under these conditions, equations (28) determine the unique steady state values of the capital stock and revenues

$$\begin{aligned} \overline{K} &= \tfrac{1}{\frac{\delta}{1-\delta} + k\frac{1-\beta(1-\delta)}{\beta(1-\delta)}} \tfrac{Lz}{f_I} \\ \overline{R} &= \tfrac{\frac{1-\beta(1-\delta)}{\beta(1-\delta)}(k+1)}{\frac{\delta}{1-\delta} + k\frac{1-\beta(1-\delta)}{\beta(1-\delta)}} Lz \end{aligned} \tag{30}$$

so that steady state dividends and employment in goods production are $\overline{D} = \overline{R}/(k+1)$ and $\overline{L} = \overline{R} - \overline{D} = k\overline{D}/z$ respectively. In steady state, the number of entrants is therefore $\overline{K}/f$ while the number of producers is $\overline{N} = (\overline{c}/c_M)^k \, \overline{K}/f$, where $\overline{\rho} = (\overline{c}/c_M)^k$ is the success rate of entry. Results (30) also imply

$$\begin{aligned} \tfrac{\overline{R}}{\overline{K}} &= \tfrac{1-\beta(1-\delta)}{\beta(1-\delta)}(k+1) f_I \\ \tfrac{\overline{D}}{\overline{K}} &= \tfrac{1-\beta(1-\delta)}{\beta(1-\delta)} f_I \end{aligned}$$

Given $\overline{K}$ and $\overline{W}$, equations (29) (implicitly) determine the unique steady-state cutoff unit labor requirement $\overline{c}$ and marginal utility of income $\overline{\lambda}$. To see this, rewrite (29) as

$$\overline{\lambda} = \frac{2\gamma\left(k+2\right)\left(c_M\right)^k f}{L}\frac{\overline{R}}{\overline{\overline{K}}}\frac{1}{\overline{c}^{k+2}} \tag{31}$$

$$\overline{\lambda} = \frac{2\alpha\gamma(k+1)\left(c_M\right)^k f}{2\gamma(k+1)\left(c_M\right)^k f\overline{c} + \eta\overline{K}\overline{c}^{k+1}} \tag{32}$$

Both these expressions represent $\overline{\lambda}$ as positive decreasing functions of $\overline{c}$ with (31) everywhere steeper than (32). 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

$$c_M > \frac{k+2}{k+1}\frac{2\gamma(k+1)f + \eta\overline{K}}{\alpha L}\frac{\overline{R}}{\overline{\overline{K}}} \tag{33}$$

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

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

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

which, by (31) and (32), can be transformed into

$$\overline{U} = \frac{1}{2\eta}\left(\alpha - \overline{\lambda}\overline{c}\right)\left(\alpha + \frac{k+1}{k+2}\overline{\lambda}\overline{c}\right) \tag{34}$$

This is a decreasing function of $\overline{\lambda}\overline{c}$. Based on (1), steady state intertemporal indirect utility equals the present value of the constant flow (34) discounted at rate β.

2.10 Comparative Statics and Aggregate Productivity

As (29) do not lend themselves to explicit analytical solution, some comparative statics results around the steady state can be obtained graphically after rewriting (31) and (32) as follows

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

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

[3]Intuitively, the condition for existence and uniqueness of the steady state requires (31) to be below (32) at $\overline{c} = c_M$.

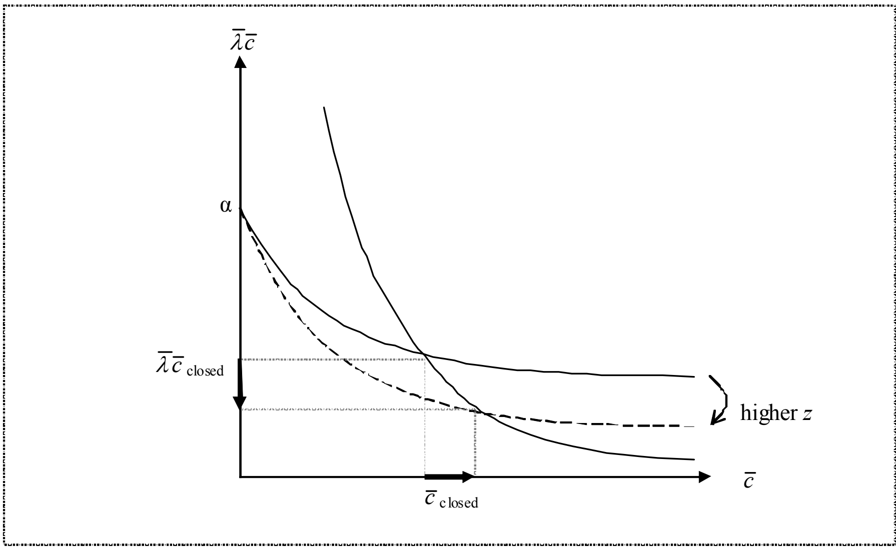

Figure 1: Closed Economy

Figure 1 provides a graphical representation of the determination of $\overline{\lambda}\overline{c}$ and $\overline{c}$ for given $\overline{K}$ and $\overline{W}$, with (35) being the steeper curve and (36) being the flatter one associating $\overline{\lambda}\overline{c} = \alpha$ to $\overline{c} = 0$. The fact that along the steeper curve (35) $\overline{\lambda}\overline{c}$ 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 (33) holds.

The focus is on the effect of an aggregate shock to labor productivity z. For concreteness, consider an exogenous increase in z; the impact of lower z will be clearly symmetric. The initial situation is represented by the two solid curves. Given (30), larger z drives $\overline{K}$ up while $\overline{R}/\overline{K}$ remains unchanged. This implies that, whereas (35) does not move, (36) 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 increases the number of entrants (larger $\overline{K}/f$) as well as the number of producers (larger $(\overline{c}/c_M)^k\,\overline{K}/f$) due to both more entry (larger $\overline{K}/f$) and a higher survival rate for entrants (larger $(\overline{c}/c_M)^k$). Given (27), by raising $\overline{c}$, higher labor productivity is associated with higher average price and average markup as well as with lower average output. Given (34), by decreasing $\overline{\lambda}\overline{c}$ higher labor productivity is also associated with higher welfare.

Higher labor productivity makes firm survival easier allowing an additional margin of less efficient firms to survive. This has implications on steady state aggregate productivity measured as ouput per worker in goods production

$$\frac{\overline{Q}}{\overline{L}} = \frac{\overline{K}}{\overline{L}} \frac{L}{2\gamma\,(k+1)\,(c_M)^k} \overline{\lambda}\overline{c}^{k+1} = \frac{k+2}{k}\frac{z}{\overline{c}} \tag{37}$$

where the first equality is granted by (27) while the second one is granted by (30) and (35). Expression (37) highlights how the exogenous labor productivity shock interacts with firm selection in endogenously determining aggregate productivity: the impact of higher labor productivity (larger z) on aggregate productivity $(\overline{Q}/\overline{L})$ is reduced by the the entry of less efficient firms (larger $\overline{c}$).

The stabilizing effect of adverse firm selection works through two channels. First, weaker competition leads to the entry and survival of an additional margin of less efficient firms. Second, holding the number of active firms constant, it also leads to a reallocation of market shares towards less efficient firms due to the fact that the elasticity of demand falls more for high-price firms than for low-price ones (see (11)). Obviously, the latter channel would be muted if demand exhibited constant elasticity.

It is also interesting to point out that the offsetting effect of adverse firm selection is stronger the larger is k as larger values of k are associated with a distribution of unit labor requirements that is more skewed towards high draws and therefore generates a fatter tail of less efficient firms. In other words, as larger k reduces firm heterogeneity, the impact of higher labor productivity on aggregate productivity is stronger the more heterogeneous firms are. This reveals a way through which microeconomic heterogeneity may crucially affect macroeconomic performance as this is determined not only by the first moment but also by higher moments of the distribution of firms across unit labor requirement levels. This would not be true in the case of CES demand.

3 Open Economy

Variable elasticity of demand naturally leads to price discrimination. It is therefore interesting to extend the model to an open economy to discuss the role of pricing-to-market. This is readily done in the case of two identical countries (Home and Foreign) with partially integrated goods markets where international trade is hampered by iceberg trade barriers: the delivery of a unit of goods requires the shipment of $\tau > 1$ units because a fraction $\tau - 1$ of the shipped quantity melts in transit.[4]

3.1 Two Symmetric Countries

A home efficiency unit of labor is taken as numeraire, so wage per efficiency unit equals one ($W_s = 1$). Due to symmetry, also foreign wage per efficiency unit equals one. Accordingly, symmetry implies that there is no need to introduce differentiated notations for Home and Foreign variables. The equilibrium in each country is characterized by modified versions of (28) and (29). A first difference with respect to the closed economy is that, though restricted, trade in goods now implies that the revenues of domestic firms derive not only from

[4] Perfect symmetry between countries implies that the distinction between financial autarky and financial integration is immaterial.

domestic sales but also from export sales

$$R_s = \frac{\left(1+\tau^{-k}\right) L}{2\gamma\,(k+2)\,(c_M)^k\, f} \lambda_s\, (c_s)^{k+2}\, K_s \tag{38}$$

In the limit case of autarky ($\tau \to \infty$), (38) boils down to the closed economy case. At the other extreme, when trade is free ($\tau = 1$), having access to the external market generates twice as much revenues. In the intermediate case ($\tau \in (1,\infty)$), export revenues are a fraction $\tau^{-k} < 1$ of domestic revenues. Due to the combination of the distributional assumption and iceberg trade barriers, also (27) is unaffected by the opening up of the economy, the only caveat being that $\widetilde{q}_s$ refers to the average output of producers for their own domestic market. Hence, their average output inclusive of exports equals $\widetilde{q}_s \left(1+\tau^{-k}\right)$.

Under symmetry, in the open economy the steady state values of the capital stock and revenues in each country are still given by $\overline{K}$ and $\overline{R}$ in (30). Conditions (35) and (36) are, instead, affected by the shift from closed to open economy. In particular, by (38), the new expression for (35) can be written as

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

which shows again $\overline{\lambda c}$ as a decreasing function of $\overline{c}$. However, for any $\overline{c}$, the

corresponding value of $\overline{\lambda c}$ is smaller for (39) than for (35) due to the presence of the extra term $1/\left(1+\tau^{-k}\right)$. As to be expected, such gap disappears in autarky ($\tau \to \infty$). The extra term in (39) derives from the fact that a fraction τ^{-k} of producers in a country are able not only to sell to the domestic market but also to export.

In the case of (36), the new expression becomes

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

which also shows $\overline{\lambda c}$ as a decreasing function of $\overline{c}$ such that, for any $\overline{c}$, the corresponding value of $\overline{\lambda c}$ is smaller for (40) than for (36) due to the presence of the extra term $\left(1+\tau^{-k}\right)$ at the denominator that equals one in autarky ($\tau \to \infty$). The extra term in (36) derives from the fact that sellers to a country consist of domestic producers and a fraction τ^{-k} of foreign producers that are able to export.

Figure 2 is the analogue of Figure 1. It compares the open economy with the closed economy. Dashed curves represent the relations between $\overline{\lambda c}$ and $\overline{c}$ implied by (39) and (40). The open economy equilibrium is at a crossing of the two dashed curves. This can be compared to the closed economy equilibrium at the crossing of the two solid curves corresponding to (35) and (36). Crucially, the function mapping the equilibrium values of $\overline{\lambda c}$ and $\overline{c}$ into welfare is given by (34) both in the closed and the open economies.

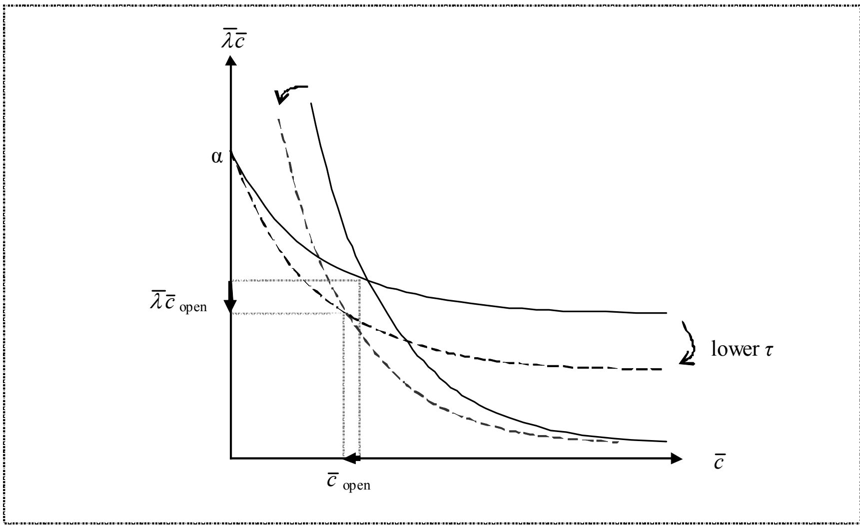

Figure 2: Trade Liberalization

3.2 Trade Liberalization

Figure 2 can be used to reveal the impact of trade liberalization (smaller τ) on firm efficiency $\overline{c}$ for given labor productivity z. As τ falls, (30) shows that $\overline{K}$ and $\overline{R}$ do not change. Differently, both (39) and (36) shift downwards leading to smaller equilibrium values for both $\overline{\lambda}\overline{c}$ and $\overline{c}$. This brings tougher competition. As a result, whereas trade liberalization has no effect on the number of entrants ($\overline{K}/f$ is unchanged), the number of producers falls (smaller $(\overline{c}/c_M)^k\,\overline{K}/f$) due to a lower survival rate for entrants (smaller $(\overline{c}/c_M)^k$). Given (27), by reducing $\overline{c}$ trade liberalization leads to lower average price and average markup. Given (34), by decreasing $\overline{\lambda}\overline{c}$ higher labor productivity is also associated with higher welfare.

Trade liberalization makes firm survival tougher expelling a margin of less efficient firms from the market. This has implications on steady state aggregate productivity measured as ouput per worker in goods production:

$$\frac{\overline{Q}}{\overline{L}} = \frac{\overline{K}}{\overline{L}} \frac{L}{2\gamma\,(k+1)\,(c_M)^k} \overline{\lambda}\overline{c}^{k+1}(1+\tau^{-k}) = \frac{k+2}{k}\frac{z}{\overline{c}} \tag{41}$$

where the first equality is granted by the fact that average output inclusive of exports equals $\widetilde{q}_s\left(1+\tau^{-k}\right)$ while the second one is granted by (30) and (39). Expression (41) shows that in the open economy the functional relation between output per worker in goods production and the cutoff is the same as in the closed economy. Accordingly, by reducing $\overline{c}$ trade liberalization boosts aggregate productivity through a selection effect. The impact is stronger the higher exogenous labor productivity (the larger z) and firm heterogeneity (the lower k) are.

The average markup and average price of domestic sales are $\widetilde{\mu}_s$ and $\widetilde{p}_s$ whereas the average markup and average price of foreign sales equal $\widetilde{\mu}_s \tau^{-k}$ and $\widetilde{\mu}_s \tau^{-k}$. Hence, as to be expected, trade liberalization (lower τ) reduces the scope for price discrimination and thus pricing-to-market. The reason for this is that, as trade barriers fall, producers face increasingly similar demand elasticities in the domestic and export markets.

3.3 Aggregate Productivity

Consider now the effect of an aggregate productivity shock to labor productivity z, focusing again on an exogenous increase in z for concreteness. Given (30), larger z drives $\overline{K}$ up while $\overline{R}/\overline{K}$ remains unchanged. This implies that, whereas (39) does not move, (40) shifts downwards. As a result, the equilibrium value of $\overline{\lambda c}$ falls whereas the equilibrium value of $\overline{c}$ rises, thus reducing the toughness of competition. Accordingly, higher labor productivity increases the number of entrants (larger $\overline{K}/f$) as well as the number of producers (larger $(\overline{c}/c_M)^k \, \overline{K}/f$) due to both more entry (larger $\overline{K}/f$) and a higher survival rate for entrants (larger $(\overline{c}/c_M)^k$). Given (34), by decreasing $\overline{\lambda c}$ higher labor productivity is associated with higher welfare. Given (27), by raising $\overline{c}$ higher labor productivity is also associated with higher average price and average markup as well as with lower average output. Hence, just like in the closed economy, also in the open economy the transmission of a positive labor productivity shock to aggregate productivity is hampered by adverse firm selection.

Is the impact of the labor productivity shock larger or smaller in open than in closed economy? The answer to this question lies in two facts. First, by (30) larger z drives $\overline{K}$ up by the same amount in open and closed economy. This translates, however, into a larger downward shift in the open economy curve (40) than in the closed economy curve (36) given that the term $(1 + \tau^{-k}) > 1$ multiplies $\overline{K}$ at the denominator of the former. Second, (39) is flatter than (35). The two facts together imply that the same increase in z makes the equilibrium value of $\overline{\lambda c}$ fall more and the equilibrium value of $\overline{c}$ rise more in the open than in the closed economy. The more so, the lower τ. Accordingly, the stabilizing selection effect is stronger in open than in closed economy and its strength increases as trade gets freer. The reason for a stronger selection effect is the increase in demand elasticity brought by trade liberalization together with a reduced scope for price discrimination and therefore pricing-to-market.

4 Conclusion

This paper has presented a theory of (partially) endogenous aggregate productivity with firm selection. The focus has been on the determination of output per worker when firms with heterogenous efficiency react in terms of entry and exit to exogenous changes in efficiency units per worker.

Though transitionary dynamics have been fully characterized, the focus has been on the steady state. The main result is that more efficiency units per worker make survival easier for a larger number of less efficient firms. Accordingly, both the number of entrants and the number of producers increase. More efficiency units per worker are also associated with higher average price and average markup as well as with lower output per firm.

A way to read these findings from a business cycle point of view is that, during labor productivity driven upswings, there is more entry, more survival after entry and surving firms are on average less efficient and smaller. The opposite is true during downswings. Hence, the impact of changing efficiency units per worker on aggregate output per worker is reduced by the pro-cyclical entry and exit of the least efficient firms. Due to variable demand elasticity, the stabilizing effect of firm selection works also through a second channel. Holding the number of active firms constant, in an upswing 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 an downswing the opposite happens. This second channel would be muted if demand exhibited constant elasticity.

It has also been pointed out that the stabilizing effect of selection depends on the degree of firm heterogeneity: the impact of a change in efficiency units per worker on aggregate output per worker is stronger the more heterogeneous firms are. This reveals a way through which microeconomic heterogeneity may crucially affect macroeconomic performance as this is determined not only by the first moment but also by higher moments of the distribution of firms across productivity levels. This would not be true in the case of CES demand.

These findings have been shown to carry over to the open economy. Crucially, however, the selection effect is stronger in open economy than in closed economy and its strength increases as trade gets freer. The reason lies in the increase in demand elasticity brought by trade liberalization together with a reduced scope for price discrimination and therefore pricing-to-market.

The analysis has currently focused on steady state reactions to permanent shocks, and on two identical countries in the case of the open economy. It would be natural to introduce country asymmetries in order to study the international transmissions of country specific shocks and explore the transitionary dynamics.

References

[1] Arkolakis, C., Costinot, A., and A. Rodriguez-Clare (2010) New Trade Models, Same Old Gains?, *American Economic Review*, forthcoming.

[2] Atkenson, A., and A. Burstein (2008) Pricing-to-Market, Trade Costs, and International Relative Prices, *American Economic Review* 98, 1998-2031.

[3] Alessandria, G., and H. Choi (2007) Do Sunk Costs of Exporting Matter for Net Export Dynamics?, *Quarterly Journal of Economics* 122, 289-336.

[4] Bergin, P., and R. Glick (2003) Endogenous Tradability and Macroeconomic Implications, National Bureau of Economic Research, NBER Working Papers 9739.

[5] Bernard, Andrew B., Stephen J. Redding and Peter K. Schott (2010) Multiproduct Firms and Trade Liberalization, *Quarterly Journal of Economics*, forthcoming.

[6] Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum (2003), Plants and Productivity in International Trade, *American Economic Review* 93, 1268-1290.

[7] Eckel, Carsten and J. Peter Neary (2010) Multi-Product Firms and Flexible Manufacturing in the Global Economy, *Review of Economic Studies* 77, 188-217.

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

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

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

[11] Mayer, T., Melitz, M.J., and G.I.P. Ottaviano (2011) Market Size, Competition, and the Product Mix of Exporters, National Bureau of Economic Research, NBER Working Papers 16959.

[12] Melitz, Marc J. and Gianmarco I.P. Ottaviano (2008) Market Size, Trade and Productivity, *Review of Economic Studies* 75, 295 -316.

[13] Neary, Peter J. (2007) Cross-border Mergers as Instruments of Comparative Advantage, *Review of Economic Studies* 74, 1229-1257.