Liberalising trade in the shadow of superstar firms

Chara Vavoura*

June 2016

Preliminary and Incomplete.

Abstract

Strong empirical evidence points towards an extremely skewed distribution of exporters, corresponding to a few "superstar" firms functioning alongside a fringe of small competitors. Motivated by this pattern, we develop a model where all firms enter symmetrically but some grow to the extent that they can drive market aggregates whereas others remain negligible throughout their life cycle. Conditions are derived for such strategic market power differences to arise endogenously as a subgame perfect equilibrium: heterogeneity in productivity results in different abilities to invest in cost-reducing innovation which increases profitability to the point where it becomes optimal to incorporate one's ability to influence the market. We then use this model to investigate the welfare impact and distributional consequences of trade liberalisation. We show that the coexistence of firms that differ not only in their productivity but also in their strategic behaviour reduces the pro-competitive welfare gains from trade: trade liberalisation reallocates market share from smaller towards larger players resulting in the fringe firms largely absorbing the competitive pressures from increased import penetration. In fact, the ability of superstars to price-to-market and shield themselves from competition gives rise to an inefficiency in the presence of which size-dependent policies might have a welfare-enhancing role to play.

*School of Economics, University of Nottingham. E-mail: Charikleia.Vavoura@nottingham.ac.uk.

1 Introduction

Global markets are extremely concentrated: the largest 1% of trading firms -formally defined as the *superstar* firms- consistently account for at least half of a country's total exports with the precise market share of this top percentile increasing with a country's stage of development (Freund and Pierola, 2015; Mayer and Ottaviano, 2008). For example, in the case of the U.S. superstars are responsible for no less than 80% of the value of total trade (Bernard et al., 2007). Being the ones that drive trade volumes, superstars have the ability to shape both domestic and export market aggregates. This property, however, is largely uncharacteristic of their smaller counterparts. In this paper we examine two questions. First, how do some firms grow into superstars whereas most of their competitors remain small throughout their life cycle? And second, incorporating the co-existence of firms with different abilities to influence market outcomes, what welfare gains and distributional consequences should we expect from trade liberalisation?

To address our first question we need to specify *how* are superstars different. Superstars are found to be significantly more productive compared to the average firm and they appear to charge a higher markup of price over marginal cost[1]. This observation is consistent with the class of trade models with firm heterogeneity following Melitz (2003): firms differ in their productivity and it is solely due to their productivity advantage that some firms have the ability to be more profitable. In these models trade liberalisation triggers a *selection* of the most efficient producers leading to an increase in average productivity which creates welfare gains from trade.

This standard paradigm attributes any differences in firms' behaviour to differences in productive efficiency. However, there is a growing body of available plant- and firm-level data which indicates that there is a significant and persistent markup gap between firms operating in the same industry which, although consistent with the well-substantiated productivity premia, it can only partly be explained by heterogeneous productivity. In fact, De Loecker and Warzynski (2012) estimate that, contrary to the predictions of the Melitz (2003)-type models, controlling for productivity differences leaves out a significant *residual* markup inequality. These findings suggest that, besides the differences in productivity, there exist additional important sources of markup heterogeneity. In their study, Tamminen and Chang (2013) report that "great markup dispersion within narrowly defined sectors would imply high variety in the level of market power between companies". This conclusion is in line with a long tradition in industrial organisation literature[2] that treats the ability to charge different markups as an indication of differential *strategic market power*[3].

Theoretical new trade research is based on one of two modes of competition: either monopolistic competition or oligopoly[4]. In both of these market structures all firms have the same ability to shape market aggregates and, thus, their predictions are built upon a uniform supply-side market power assumption which is put in doubt by the mere co-existence of superstars and infinitesimal firms that is so often witnessed in global markets. Echoing a number of recent empirical works like Bernard et al. (2015) and De Loecker et al. (2015) which call for the need to abstract from this assumption,

[1] See for example Mayer and Ottaviano (2008) and De Loecker and Warzynski (2012).

[2] See for example Hall (1986) and Shapiro (1987).

[3] This term only refers to the supply-driven market power that relates to whether a firm internalises the effect of its decisions on market aggregates.

[4] For a survey of the literature on trade with firm heterogeneity see for example Melitz and Redding (2014). For a survey of the oligopolistic approach see Neary (2010).

our study introduces market power differences jointly with productivity heterogeneity in an integrated framework.

To address our first question on how do some firms acquire market power, we construct a two-period general equilibrium model with free entry in order to mimic firm dynamics and illustrate the simplest possible way in which a small number of firms can choose to grow into superstars. The idea is simple: in a dynamic setup initial efficiency differences are endogenously magnified leading to differences in strategic behaviour. On the producer side, in the first period, we assume an industry (that can be thought of as newly-created industry) populated by monopolistically competitive firms which must incur a sunk entry cost before they can draw a productivity from a commonly known distribution and shape accordingly their marginal cost of producing a variety of a heterogeneous good. Some firms get lucky in their draw and, benefiting from their luck, they grow more and make a higher profit. They, then, choose to invest part of their accumulated profit in order to innovate and further increase the productivity gap that separates them from their competitors. We model this turning point for firm dynamics as the second period of our model (which can be rationalised as a mature industry). Successful innovators become extremely efficient and thrive. Being responsible for a disproportionately large market share, a handful of firms can affect the market aggregates. We derive the conditions under which it is optimal for these firms to internalise this effect. As a result, the behaviour of these happy few changes endogenously. Given their innovation decision, they choose to switch from acting as non-strategic monopolistic competitors to behaving like Cournot oligopolists. At the same time, firms that got unlucky and drew a low initial productivity remain small throughout their lives and unable to manipulate these market aggregates, treating them parametrically when they form their decisions. Oligopolists compete with their "unlucky" counterparts who have remained small monopolistic competitors, in a *mixed market* structure consisting of two subsectors governed by distinct forms of competition, namely monopolistic competition and oligopoly. On the consumer side, we consider a CES demand system. Thus, we are able to "shut down" any demand-driven markup heterogeneity. As a result, this mixed market mechanism creates a markup dispersion than, unlike models like Melitz and Ottaviano (2008), is not generated via a specific demand-side assumption (namely linear demand) but is a corollary of the underlying process of firm dynamics.

Our paper draws heavily on the mixed market structure developed by Shimomura and Thisse (2012) and introduced in trade literature by Parenti (2013)[5]. Shimomura and Thisse construct a general equilibrium model of static mixed market competition with a given finite number of firms born oligopolistic and a continuum of firms born monopolistically competitive whose total mass adjusts according to a zero profit condition. Our model incorporates this setup as its second-period market structure relaxing a number of its restrictive assumptions and extending it from an ad hoc closed economy framework to an open economy firm dynamics result. In our model firms are not born large or small. Firms grow not only because of luck but also because of successful innovation. Differences in productivity and market power are obtained as the necessary conditions for the mixed market structure to be a subgame perfect equilibrium in pure strategies. This intuition is in line with Peters (2013) who argues that, under imperfect competition, high productivity is an indication of market power and

[5]Parenti (2013) develops a model with single-product monopolistic competitors and multi-product oligopolists. Although his approach is methodologically very different, some of our key findings are in agreement and they will be presented in the remainder of the paper.

that the presence of market power affects markets not only statically but also dynamically, since it influences the innovation decisions of firms.

Giving up the assumption of a fixed number of oligopolists our model is, to the best of our knowledge, the first to incorporate a free entry condition with the equilibrium total number of firms adjusting to the point where the total expected profits prior to entry (from both periods additively) are set equal to the entry cost. We use our assumption of free entry in order to illustrate the relationship between product innovation and competition. We show that as markets become more competitive (in the sense that they become less heterogeneous), Romer (1990)-type innovation (as it is reflected in the number of differentiated varieties produced in the economy) follows an inverted-U pattern that peaks at relatively low levels of competition. Although our modeling approach is completely different, these findings bear a strong resemblance to the ones of the endogenous growth literature following Aghion et al. (2002).

To answer our second question on the magnitude and the distribution of welfare gains from trade, in the second period of our model, we introduce a foreign perfectly symmetric country and explore the impact of trade liberalisation on aggregate welfare and the implications of costly trade for the two different types of firms. Under CES, any markup gap within the domestic market and between the domestic and foreign market can only be attributed to differences in strategic (supply-side) market power. In autarky, we find that superstars charge lower prices and higher markups. But, allowing for international trade, when firms are large, not only do they decide to charge a higher markup in their domestic market, in comparison to the fringe firms, but also, unlike them, they are capable of pricing-to-market, with their selected markups varying across markets, according to the respective sales shares. In the presence of trade costs, large firms will select a lower markup in the foreign market, engaging in reciprocal dumping which is a recurring theme in the oligopolistic trade models introduced by Brander and Krugman (1983). The ability of large firms to charge variable markups implies that trade liberalisation in the form of a decrease in the cost to export will only partly be passed on to the prices. On the contrary, as in Krugman (1980), monopolistically competitive firms will always charge the same markup at home and abroad.

The idea of incomplete pass-through has been explored in a number of theoretical and empirical works. The pattern is generated under both oligopolistic market structure, as in Atkeson and Burnstein (2008), and under non-CES monopolistic competition, as in Melitz and Ottaviano (2008). However, although there is strong empirical evidence to support the existence of variable markups, exactly how trade influences markups is less clear. De Loecker et al. (2015) find that markups actually increase as a result of trade liberalisation. This is because trade liberalisation not only reduces a firm's cost of exporting its output but it also reduces the cost of imported inputs. This result goes in the opposite way from Mayer et al. (2011) and Edmond et al. (2012) who find that markups decrease due to large pro-competitive gains from trade.

We show that, bringing together the two established but hitherto unrelated market structures of oligopoly and monopolistic competition in an open economy setup, the result is not a simple combination of the variety gains predicted by the Krugman (1980)-type monopolistically competitive models with the pro-competitive gains which are expected to materialise under pure oligopoly, as in Brander and Krugman (1983). In fact, the coexistence of firms that differ in their strategic behaviour gives rise to a composition effect of these two traditional sources of gains resulting in a market share reallocation from smaller towards larger players which dampens the pro-competitive effect of openness and reduces welfare gains from trade liberalisation. Globalisation disproportionately benefits large firms whose gains from the decrease in the cost of exports dominates their losses from increased competition in the domestic market. This is due to the presence of small firms who largely absorb the competitive pressures, shielding large firms from increased import penetration. In the end, average markups decrease and welfare increases but the magnitude of these changes is significantly smaller than what standard trade models à la Brander and Krugman would predict.

Reconciling this result with the empirical literature on incomplete pass-through mentioned above, we find that, in the presence of firm productivity and market power heterogeneity, markups decrease as in Mayer et al. (2011) and Edmond et al. (2012) but, differently from their predictions, they are unlikely to decrease significantly. Gains from trade are always positive but they decrease dramatically with market concentration. This result highlights the importance of incorporating asymmetric market power in estimating the distributional consequences of trade liberalisation. This insight is present in the press where large enterprises appear to be much more pro-globalisation. For example, in the UK, business leaders almost unanimously opposed Brexit, as opposed to smaller companies who appeared to be divided[6].

Finally, we use our model to evaluate the desirability of size-dependent trade policies. We show that trade increases welfare more in less concentrated industries. The reason is that trade triggers two different types of market share reallocation away from the smaller and towards the larger firms: a reallocation which resembles the Melitz-type selection and is due to the efficiency advantage of superstar firms and a reallocation that is solely attributed to the differential abilities of the two types of firms to charge variable markups. The latter form of reallocation is inefficient. This inefficiency creates a role for size-dependent export policies which could turn out to be welfare-enhancing. To sum up, our model highlights the trade-off between the efficiency gains from innovation undertaken by superstars and the welfare losses due to decreased competition in the presence of few powerful firms. We conclude that trade and competition policy are complementary and should be jointly pursued, aiming at lowering entry costs and the cost of innovation.

The structure of our paper is as follows. We first present the closed-economy version of our model. To illustrate our results, we conduct a comparative statics analysis of the closed-economy equilibrium. We go on to present the open economy version of the model with trade between two symmetric countries. We then discuss our findings and explore the impact of trade liberalisation by illustrating the behaviour of our model as we change the trade cost. Finally, we present our policy implications and conclude.

2 Setup of the model

2.1 Preferences and demand

In order to mimic firm dynamics in the simplest possible way we analyse a two-period game, where the first period corresponds to a newly-created industry and the second period represents a mature

[6] See for example Financial Times (2016).

one. At each point in time ($t = 1, 2$) the economy involves one horizontally differentiated and one homogeneous good the production of which only requires labour, which is perfectly mobile.

Following Shimomura and Thisse (2012), the horizontally differentiated good (H) is produced under increasing returns. At time t, H_t can be supplied either by large oligopolistists or by small monopolistically competitive firms (MC firms). We assume that the differentiated good is formed by two CES-composite goods, X and Y, which at time t are defined as follows:

$$X_t = \left(\sum_{i=1}^{n_{xt}} x_{it}^a \right)^{\frac{1}{a}}$$

$$Y_t = \left(\sum_{i=1}^{n_{yt}} y_{it}^a \right)^{\frac{1}{a}}$$

where x_{it} is the output level of large firm $i = 1, 2, \ldots, n_{xt}$ at time t and y_{it} is the output level of small firm $i = 1, 2, \ldots, n_{yt}$ at time t. The parameter $a \in (0, 1)$, implying that the goods are imperfect substitutes. The elasticity of substitution between any two goods is given by $\sigma \equiv \frac{1}{(1-a)} > 1$. Firms can costlessly differentiate their products and all products enter symmetrically into demand. Hence, in the eyes of the consumers, large and small firms produce equally substitutable goods. Pan and Hanazono (2015) examine how allowing for different substitutabilities would alter the predictions of the mixed market model. Wanting to shut down any demand driven heterogeneity we assume that the elasticity of substitution is equal to σ both across and within subsectors.

In pure monopolistic competition, consumers' utility only depends on Y, whereas in pure oligopoly it only depends on X. Combining X and Y we obtain H which corresponds to the output index of the entire differentiated sector:

$$H_t = (X_t^a + Y_t^a)^{\frac{1}{a}}$$

The homogeneous good (O) sums up the rest of the economy and serves in placing the analysis in general equilibrium. O is produced under constant returns and supplied by perfectly competitive firms.

Given CES preferences, we can assume that there is a representative consumer who describes the aggregate behaviour of consumers having different tastes (Anderson et al. 1992). This agent is endowed with L units of labour and holds the shares of all firms. Preferences of the representative consumer over the homogeneous and the differentiated good are represented by a logarithmic upper-tier instantaneous utility function. At every period $t = 1, 2$, the representative consumer will maximise the utility function:

$$U_t = \ln H_t + \beta \ln O_t$$

subject to the budget constraint:

$$O_t + E_t = I_t$$

In the utility function, the parameter $\beta > 0$ expresses the weight of the homogeneous good. A higher

β shifts consumption away from the differentiated and towards the homogeneous industry.

In the budget constraint, E represents the expenditure on the differentiated good, which at at time t is defined as $E_t = P_t H_t$, and I is the income level (in terms of the numeraire). Since income is given by the wage bill wL plus any uninvested profits and given that profits are determined at the equilibrium, I is endogenous. Without loss of generality, we assume that one unit of labour produces one unit of the homogeneous good. We take O as the numeraire commodity and, therefore, the equilibrium wage will be equal to 1 and $wL = L$. We move on to P_t, which, following Dixit and Stiglitz (1977), can be defined as the aggregate price of the differentiated sector:

$$P_t = \left(\sum_{i=1}^{n_{xt}} p_{xit}^{\frac{-a}{1-a}} + \sum_{i=1}^{n_{yt}} p_{yit}^{\frac{-a}{1-a}} \right)^{-\frac{1-a}{a}}$$

We can rewrite the price index as a function of income and the industry output index as follows:

$$P_t = E_t H_t^{-1}$$

The assumption of CES preferences is critical. Under a CES demand system there is no demand-driven market power and hence no demand-driven markup heterogeneity. The only demand-side market power stems from imperfect substitutability which is equal across varieties. As a result, unlike models like Melitz and Ottaviano (2008), we can attribute any markup heterogeneity to supply-side determinants, namely productivity and strategic market power. Controlling for productivity, any residual markup inequality will be driven by differences in strategic market power. This assumption becomes particularly important in our analysis of an open economy. Since CES is the standard preference structure in international trade, our results are straightforwardly comparable to the majority of trade papers with firm heterogeneity.

Assuming that there are no means available to the consumer to transfer wealth from one period to the other, we can ignore time discounting[7] and focus on the per-period optimisation problem, which, using the expressions presented above, we can re-write as follows:

$$\max_{\{x_{it}\}_{i=1}^{n_{xt}}, \{y_{it}\}_{i=1}^{n_{yt}}, O_t} \left\{ U_t = \ln \left[\left(\sum_{i=1}^{n_{xt}} x_{it}^a \right) + \left(\sum_{i=1}^{n_{yt}} y_{it}^a \right) \right]^{\frac{1}{a}} + \beta \ln O_t \right\}$$

$$\text{s.t. } O_t + \sum_{i=1}^{n_{xt}} p_{xit} x_{it} + \sum_{i=1}^{n_{yt}} p_{yit} y_{it} = I_t$$

The problem can be decomposed in three choices: between H and O, between X and Y and the allocation of expenditure within X and Y. Utility maximisation leads to the following optimal decisions:

$$O_t = \beta E_t$$

$$X_t = H_t \left(\frac{P_{xt}}{P_t} \right)^{-\frac{1}{1-a}}$$

[7]Including a time discounting parameter would be a straight-forward generalisation.

$$Y_t = H_t \left(\frac{P_{yt}}{P_t}\right)^{-\frac{1}{1-a}}$$

$$x_{it} = H_t \left(\frac{p_{xit}}{P_t}\right)^{-\frac{1}{1-a}}$$

$$p_{xit} = x_{it}^{a-1} H_t^{1-a} P_t$$

$$y_{it} = H_t \left(\frac{p_{yit}}{P_t}\right)^{-\frac{1}{1-a}}$$

$$p_{yit} = y_{it}^{a-1} H_t^{1-a} P_t$$

According to the first equation, the demand for the homogeneous good is β times the total spending on the differentiated good, which is a standard result of log preferences. The following two equations give the total demand from the oligopolistic and the MC-firms subsector respectively, whereas the final four equations correspond to the demand and the inverse demand for each differentiated variety, produced by either an oligopolist or by an MC firm. The aggregate price index of each subsector is defined as follows:

$$P_x = \left(\sum_{i=1}^{n_{xt}} p_{xit}^{\frac{-a}{1-a}}\right)^{-\frac{1-a}{a}}$$

$$P_y = \left(\sum_{i=1}^{n_{yt}} p_{yit}^{\frac{-a}{1-a}}\right)^{-\frac{1-a}{a}}$$

Note that although income I is endogenous the income *share* spent on the differentiated good is constant:

$$E_t = \frac{1}{\beta+1} I_t$$

2.2 Production

We focus on a two-period game. The timing of the game is presented below.

Period 1 At the beginning of the first period N firms enter the market. Entry only occurs at the beginning of the game and there are no exogenous shocks that could force a firm to exit before the end of the game. With costless differentiation, each firm will choose to produce a different variety. To enter, firms must make an initial investment in the form of a fixed sunk entry cost $f_e > 0$ (measured in units of the homogeneous good). N is such that the total expected profits from both periods of the game are equal to the entry cost. Prior to entry firms are identical and they are considered symmetrically non-strategic. Once they enter the market, firms draw an initial productivity parameter $z > 0$ from a common distribution, where z is defined as the inverse of a firm's marginal cost. The fixed entry cost usually corresponds to the number of days needed to create a new firm (Djankov et al., 2002) and one can think of a high versus a low random productivity draw as a "good", as opposed to a "bad", business idea. The first period of our model is similar to Melitz (2003) with two key differences. Firstly, instead of a continuum of firms producing a continuum of varieties, we follow the original Dixit and Stiglitz (1977) paper and assume discrete varieties. Without this assumption, in order for a firm to "grow" it should switch from being of zero to positive mass and such a transition would be mathematically inconsistent. Secondly, to keep the model tractable, we start off with only two possible levels of productivity. Relaxing this assumption by incorporating a continuous productivity distribution à la Melitz would severely affect the model's tractability without adding to the key insights of the model. We assume that z can only take two discrete values, each with a known probability:

$$z \in \{z_l, z_h\} \text{ where } z_l \ll z_h \text{ and } P(z = z_h) = \gamma \ll 1$$

This simplification is in line with the intuition of the empirical literature like, for example, Freund and Pierola (2015) who argue that the well-documented skewness in the firm distribution "could be achieved in a heterogeneous firm framework with firms having a small probability of getting a very high productivity draw".

Production only requires labour which, at each period, is inelastically supplied at the aggregate level L. The technology of each firm is represented by its productivity z. Production also requires a fixed cost $f > 0$ which is the same across all firms. Thus, labour used for the production of output q equals:

$$l(q; z) = \frac{1}{z}q + f$$

Given that the common (across all sectors) wage w is set equal to 1, the above equation is also the cost function of a productivity-z firm.

In the first period all firms operate under monopolistic competition and hence only the Y-subsector is active. This is equivalent to the following notation:

$$n_{x1} = 0 \text{ and } n_{y1} = N$$

A firm with productivity z will face the following profit maximisation subject to the inverse demand function:

$$\max_{y_{i1}} \left\{ \pi_{yi1} = p_{yi1} y_{i1} - \frac{1}{z} y_{i1} - f \right\}$$

$$\text{s.t. } p_{yi1} = y_{i1}^{a-1} H_1^{1-a} P_1$$

Since $a \in (0, 1)$, the above profit function is strictly concave in the level of output. Therefore, the f.o.c. will give the unique optimal solution:

$$y_{i1}(E_1, P_1; z) = \left(\frac{1}{az}\right)^{\frac{-1}{1-a}} E_1 P_1^{\frac{a}{1-a}}$$

which is equivalent to the following optimal pricing rule:

$$p_{yi1}(z) = \frac{1}{az}$$

The maximised period-one profit of a firm with productivity z will be given by:

$$\pi_{yi1}(E_1, P_1; z) = \left(\frac{1}{az}\right)^{\frac{a}{a-1}} (1-a) E_1 P_1^{\frac{a}{1-a}} - f$$

which is strictly increasing in the level of productivity z.

To sum up, in the first period, the more productive firms (with productivity equal to z_h) produce the same higher output: $y_1(E_1, P_1; z_h)$, charge the same lower price $p_{y1}(z_h)$ and earn the same higher profits $\pi_{y1}(E_1, P_1; z_h) \equiv \pi_{y1}(z_h)$, compared to the less productive firms. Similarly, all low-productivity firms (with $z = z_l$) choose the same output $y_1(E_1, P_1; z_l)$, which corresponds to the same price $p_{y1}(z_l)$ and the same profits $\pi_{y1}(E_1, P_1; z_l) \equiv \pi_{y1}(z_l)$. All firms, regardless of their productivity, will choose the same profit maximising markup of price over marginal cost equal to $\frac{1}{a}$.

Aggregating over the total number of entrants (or, equivalently, the total number of varieties)N, and given the distribution of productivities we get the total period-one production of the differentiated good:

$$H_1(N) = \left[N\gamma \left(\left(\frac{1}{az_h}\right)^{\frac{1}{a-1}} E_1 P_1^{\frac{a}{1-a}}\right)^a + N(1-\gamma) \left(\left(\frac{1}{az_l}\right)^{\frac{1}{a-1}} E_1 P_1^{\frac{a}{1-a}}\right)^a\right]^{\frac{1}{a}}$$

We can also compute the equilibrium aggregate price level of the first period which only depends on the number of firms $P_1(N)$ and, using the definition for the period-one income:

$$E_1 = (\beta + 1)^{-1} \left[L + N\gamma\pi_{y1} + N(1-\gamma)\pi_{y1}\right]$$

we can write $E_1(N)$ as a function of N. Using $E_1(N)$ and $P_1(N)$ we can express the optimal profits $\pi_{y1}(z_l)$ and $\pi_{y1}(z_h)$ in terms of N only.

Period 2 In the beginning of the second period each firm has the option to invest in efficiency-enhancing innovation. This innovation works as follows. Incurring a fixed sunk cost $f_k \geq 0$ increases a firm's productivity by $k \geq 0$ and, more importantly, it enables the firm to act as a Cournot oligopolist. We define the post-innovation level of productivity as follows:

$$z_k \equiv z_h + k$$

To simplify the analysis this innovation is successful with probability 1. A non-deterministic innovation would result in a heterogeneous monopolistically competitive fringe with two different levels of productivity and fixed costs. Given the extended research on monopolistic competition with heterogeneous firms, relaxing the assumption of deterministic innovation would unnecessarily complicate the analysis.

This innovation could be rationalised in a number of ways. The obvious intuition is that it acts as a reduced form representation of a technological adoption cost in line with the endogenous growth literature[8]. A high productivity draw in period 1 enables firms to accumulate profit which they may choose to invest to further increase their cost efficiency. Having increased the productivity gap from their competitors, innovators are now responsible for a disproportionately large market share and, as a result, have a significant influence on market aggregates, namely aggregate output H_2. Under general assumptions that will be presented below, it is optimal for these firms to internalise this effect which is equivalent to adopting an oligopolistic behaviour. Alternatively, borrowing from Igami (2015) our innovation could also be explained as the cost of offshoring, since offshoring alters market structure along with its cost structure and takes the form of a discrete investment in order to reduce future costs. Finally, the lobbying literature like, for example, Bombardini (2008) incorporates such fixed costs of making political contributions that will create a level of protection that is equivalent to a decrease in costs.

Assuming that, in the second period, all z_h-productivity firms choose to incur f_k and compete as oligopolists and all z_l-productivity firms compete monopolistically we get that now both the X- and the Y- subsector are now active:

$$n_{x2} = \gamma N \text{ and } n_{y2} = (1-\gamma)N$$

Notice that in the following section of the paper we will derive the conditions under which we get the equilibrium result that the decision to become an oligopolist is optimal for a firm if and only if this firm has initial productivity equal to z_h.

Large firms (Cournot oligopolists) Large firms maximise their period-two profits subject to the inverse demand function:

$$\max_{x_{i2}} \left\{ \pi_{xi2} = p_{xi2} x_{i2} - \frac{1}{z_k} x_{i2} - f - f_k \right\}$$

$$\text{s.t. } p_{xi2} = x_{i2}^{a-1} H_2^{1-a} P_2$$

and taking as given that the MC firms maximise their profits subject to the inverse demand and that every other oligopolist j (where $j = 1, 2, \ldots, n_{x2}$ and $j \neq i$) produces according to their reaction function.

Substituting the inverse demand, the profit function of a large firm can be re-written as follows:

$$\pi_{xi2} = x_{i2}^{a} E_2^{1-a} P_2^{a} - \frac{1}{z_k} x_{i2} - f - f_k$$

Notice that every large firm maximises π_{xi2} with respect to x_{i2} where π_{xi2} is a function of the firm's own output x_{i2}, the industry price index P_2 and the total industry output H_2 (produced by both small and large firms). The intuition behind this is that any large firm understands that:

[8] See for example Impullitti and Licandro (2016).

1. Their input choice affects the industry price index and is therefore involved in a game-theoretic environment.

2. The industry price index is influenced by the *aggregate* behaviour of the MC firms (Shimomura and Thisse, 2012). Therefore, we represent the MC firms by the pseudo-player $i = 0$ producing a total optimal output equal to x_{02}.

3. Since the income share spent on the differentiated product is constant we get that the income level influences firms' demands and hence their profits. As a result, all firms must correctly anticipate what the total income will be. Because of their market power, large firms could in theory manipulate the income level and hence their demands through their choices. To avoid the non-existence of an equilibrium that is a usual result of accounting for these feedback effects we will take the approach followed by Neary (2015) that each firm that is large within the sector is small in the economy as a whole[9]. This *income-taking* assumption means that no large firm seeks to manipulate its demand through the income level.

Let $x_{-i2} = (x_{12}, \ldots, x_{i-12}, x_{i+12}, \ldots, x_{n_x 2})$ be the vector of all outputs of all large firms but that of firm i. Because $\frac{\partial \pi_{xi2}}{\partial x_{i2}}$ is strictly decreasing in x_{i2} we have that:

Lemma 1 *For any oligopolist $i = 1, 2, \ldots, n_{x2}$ and for any given output vector of all other oligopolists x_{-i2} and any aggregate behaviour of the MC firms x_{02}, the profit function π_{xi2} is strictly concave with respect to own output x_{i2}. Hence, the best response of firm i, defined as $x_{i2}(x_{-i2}, x_{02}; E_2)$, is the unique solution of:*

$$\frac{\partial \pi_{xi2}}{\partial x_{i2}} = \frac{a \sum_{j \neq i} x_{j2}^a}{x_{i2}^{1-a} \left(x_{i2}^a + \sum_{j \neq i} x_{j2}^a \right)^2} E_2 - \frac{1}{z_k} = 0$$

Small firms (MC firms) Every small firm treats the period-two prices and income as given parameters. Therefore the optimisation problem of a small firm is the following:

$$\max_{y_{i2}} \left\{ \pi_{yi2} = p_{yi2} y_{i2} - \frac{1}{z_l} y_{i2} - f \right\}$$

$$\text{s.t. } p_{yi2} = y_{i2}^{a-1} H_2^{1-a} P_2$$

Exactly as in the period-one case, π_{yi2} is strictly concave in y_{i2}. Applying the f.o.c. yields the equilibrium output of a small firm:

$$y_2(E_2, P_2; z_l) = \left(\frac{1}{a z_l} \right)^{\frac{-1}{1-a}} E_2 P_2^{\frac{a}{1-a}}$$

which yields the equilibrium price:

[9] This assumption is equivalent to selecting a sufficiently high β.

$$p_{y2}(z_l) = \frac{1}{az_l}$$

We have omitted the i-index since, in equilibrium, all MC firms produce the same output, sold at the same price, and, hence, make the same period-2 profits. Given that period-1 profits are also the same across all productivity-z_l firms, the total profits will also be equal within the MC subsector. Notice that the equilibrium price is the same as the one under pure monopolistic competition whereas the equilibrium output varies with the quantities chosen by the large firms through the price index P_2 and the income share E_2. From the equilibrium output and the definition of Y_2 we get that:

$$Y_2(E_2, P_2; z_l) = x_{02} = \left(\sum_{i=1}^{n_y} y_{i2}^a\right)^{\frac{1}{a}} = (1-\gamma)^{\frac{1}{a}} N^{\frac{1}{a}} \left(\frac{1}{az_l}\right)^{\frac{-1}{1-a}} E_2 P_2^{\frac{a}{1-a}}$$

2.3 Closed economy equilibrium

We consider a non-cooperative game in which big and small firms choose their output simultaneously. We define the market equilibrium as the state in which the following conditions hold:

1. The representative consumer maximises the utility function subject to the budget constraint.
2. Both high- and low-productivity firms maximise their per-period profits with respect to output.
3. The free entry condition holds[10], where the free entry condition is that the total number of firms N that enter the differentiated sector is adjusted up to the point where the total expected profits (prior to entry) become equal to zero.
4. All markets clear.

We can characterise a mixed market equilibrium using the profit maximisation conditions of the small and large firms as well as the free entry condition. We consider a *symmetric* equilibrium where, in the second period, all large firms choose the same output (i.e. $x_{i2} \equiv x_2$ for every $i = 1, 2, \ldots, n_{x2}$) sold at the same price (i.e. $p_{xi2} \equiv p_{x2}$ for every $i = 1, 2, \ldots, n_{x2}$) and, hence, earn the same profits (i.e. $\pi_{xi2} \equiv \pi_{x2}(z_k)$ for every $i = 1, 2, \ldots, n_{x2}$).

The first step is to express the equilibrium conditions in terms of N only. Using the symmetry assumption we can express the period-two aggregate price index in terms of the output level of an oligopolist:

$$P_2 = E_2(Y_2^a + \gamma N x_2^a)^{\frac{-1}{a}}$$

Substituting the above expression in the f.o.c. for the period-two profit maximisation of a large firm we get:

$$E_2^{1-a} = \frac{1}{az_k} P_2^{-a} x_2^{1-a} + E_2^{1-2a} P_2^a x_2^a$$

[10] Oligopoly with free entry raises an integer problem. For a discussion see for example Friedman (1977).

Our equilibrium conditions (for given N) form the following system of 9 equations in 9 unknowns: the per-period profits ($\pi_{y1}(z_h)$, $\pi_{y1}(z_l)$, $\pi_{x2}(z_k)$ and $\pi_{y2}(z_l)$), the price indices (P_1 and P_2), the levels of expenditure on the differentiated good ($E_1$and E_2) and the level of output produced by an oligopolist in period 2 (x_2).
First-period profits:

$$\pi_{y1}(z_h) = \left(\frac{1}{az_h}\right)^{\frac{a}{a-1}} (1-a)E_1 P_1^{\frac{a}{1-a}} - f = \pi_{y1}(N; z_h) \tag{1}$$

$$\pi_{y1}(z_l) = \left(\frac{1}{az_l}\right)^{\frac{a}{a-1}} (1-a)E_1 P_1^{\frac{a}{1-a}} - f = \pi_{y1}(N; z_l) \tag{2}$$

First-period market aggregates:

$$P_1 = \left[\gamma N \left(\frac{1}{az_h}\right)^{\frac{a}{a-1}} + (1-\gamma)N\left(\frac{1}{az_l}\right)^{\frac{a}{a-1}}\right]^{\frac{a-1}{a}} = P_1(N) \tag{3}$$

$$E_1 = \frac{L - Nf}{\beta + 1 - \left[\gamma\left(\frac{1}{az_h}\right)^{\frac{-a}{1-a}} + (1-\gamma)\left(\frac{1}{az_l}\right)^{\frac{-a}{1-a}}\right]^{-1}\left[(1-\gamma)\left(\frac{1-a}{(az_l)^{\frac{a}{a-1}}}\right) + \gamma\left(\frac{1-a}{(az_h)^{\frac{a}{a-1}}}\right)\right]} = E_1(N) \tag{4}$$

Second-period profits:

$$\pi_{x2}(z_k) = x_2^a E_2^{1-a} P_2^a - \frac{1}{z_k}x_2 - f - f_k = \pi_{x2}(E_2, P_2, x_2; z_k) \tag{5}$$

$$\pi_{y2}(z_l) = \left(\frac{1}{az_l}\right)^{\frac{a}{a-1}} (1-a)E_2 P_2^{\frac{a}{1-a}} - f = \pi_{y2}(E_2, P_2; z_l) \tag{6}$$

F.o.c. for maximisation of an oligopolist's period-2 profits linking E_2, P_2 and x_2:

$$E_2^{1-a} = \frac{1}{az_k} P_2^{-a} x_2^{1-a} + E_2^{1-2a} P_2^a x_2^a \tag{7}$$

Second-period market aggregates:

$$P_2 = E_2\left[\left(\left(\frac{1}{az_l}\right)^{\frac{-1}{1-a}} (1-\gamma)^{\frac{1}{a}} N^{\frac{1}{a}} E_2 P_2^{\frac{a}{1-a}}\right)^a + \gamma N x_2^a\right]^{\frac{-1}{a}} = P_2(E_2, x_2, N) \tag{8}$$

$$E_2 = (\beta+1)^{-1}\left(L + \gamma N \pi_{x2}(z_k) + (1-\gamma)N\pi_{y2}(z_l)\right) = E_2(P_2, N, x_2) \tag{9}$$

The 3 equations (7), (8) and (9) yield the equilibrium values of x_2, E_2 and P_2 as a function of N. Plugging $x_2(N)$, $E_2(N)$ and $P_2(N)$ into $\pi_{x2}(E_2, P_2, x_2; z_k)$ and $\pi_{y2}(E_2, P_2; z_l)$ we obtain the profit functions in terms of N, i.e. $\pi_{x2}(N; z_k)$ and $\pi_{y2}(N; z_l)$. From $\pi_{x2}(N; z_k)$, $\pi_{y2}(N; z_l)$ and using $\pi_{y1}(N; z_h)$ from (1) and $\pi_{y1}(N; z_l)$ from (2) we can express the total expected profits prior to entry as a function of N only. We use the following expression to simplify notation Proofs are presented in the appendix:

$$M \equiv (az_l P_2)^{\frac{a}{1-a}} (1-\gamma)N$$

We can express the aggregate period-two price index and the pricing rule and markup (defined as θ_{x2}) charged by a large firm in terms of M:

$$P_2 = \frac{1}{az_l}\left(\frac{M}{(1-\gamma)N}\right)^{\frac{1-a}{a}}$$

$$p_{x2} = \frac{1}{az_l}\left(\frac{\gamma M}{(1-\gamma)(1-M)}\right)^{\frac{1-a}{a}}$$

$$\theta_{x2} \equiv \frac{1}{a}\frac{z_k}{z_l}\left(\frac{\gamma M}{(1-\gamma)(1-M)}\right)^{\frac{1-a}{a}}$$

Using the above expressions and because M is a decreasing function of N we can show the following result:

Proposition 1 *Assume that the size of the market for the differentiated good is exogenous (i.e. N is given). Then, the industry price index as well as the price and markup charged by a large firm decrease when the total number of firms increases.*

The combination of CES and monopolistic competition imply that the price and the profit-maximising markup for a small firm (defined as θ_{y2}) will be the following:

$$p_{y2} = \frac{1}{az_l}$$

$$\theta_{y2} \equiv \frac{1}{a}$$

We get that large firms will charge a higher markup than small firms:

$$\frac{z_k}{z_l}\left(\frac{\gamma M}{(1-\gamma)(1-M)}\right)^{\frac{1-a}{a}} > 1$$

And that, for a sufficiently high productivity gap between large and small firms, large firms will charge a lower price:

$$\left(\frac{\gamma M}{(1-\gamma)(1-M)}\right)^{\frac{1-a}{a}} < 1$$

Notice that if all firms have the same marginal cost, large firms will choose a higher markup but also a higher price. However, in the presence of a productivity advantage, oligopolistic firms will select a lower price compared to their non-strategic competitors.

Having expressed everything in terms of N, we add the free entry condition. We show that there exists a value of N that solves our system of equations. This N is the solution to the following free entry condition:

$$E\Pi(N^*) = \gamma\left[\pi_{y1}(z_h) + \pi_{x2}(z_k)\right] + (1-\gamma)\left[\pi_{y1}(z_l) + \pi_{y2}(z_l)\right] - f_e = 0 \tag{10}$$

We can show the following result[11]:

Proposition 2 *There exists at least one $N > 0$ such that the free entry condition holds. This N will be unique for sufficiently low levels of a and γ.*

To compute aggregate welfare we calculate the indirect utility of the representative consumer by substituting:

$$\tilde{x}_t = E_t \left(\frac{p_{xt}}{P_t^a}\right)^{-\frac{1}{1-a}}$$

$$\tilde{y}_t = E_t \left(\frac{p_{yt}}{P_t^a}\right)^{-\frac{1}{1-a}}$$

$$O_t = \beta E_t$$

in the utility function. Given homothetic preferences, the indirect utility function will describe the social welfare.

3 Discussion of the closed economy

3.1 Subgame perfection

Having computed the mixed market equilibrium, we go on to specify the conditions under which this equilibrium is indeed subgame perfect in pure strategies[12].

Proposition 3 *Investing in acquiring market power is subgame perfect if and only if:*

1. *z_h is sufficiently higher than z_l so that $f_k \in (\pi_{y1}(z_l), \pi_{y1}(z_h))$ is unaffordable for z_l-productivity firms.*

2. *Given f_k, z_k could be higher or even equal to z_h, as long as f_k, γ and a are sufficiently low.*

Intuitively, if there is only one possible initial productivity draw and all firms are equally productive then at the end of period 1 they will all have accumulated the same profits. As a result, there is no reason why there will be different abilities to innovate in period 2. In a perfectly homogeneous setup, mixed markets can only arise as an equilibrium in mixed strategies. The only possible equilibria in

[11] As shown in the appendix, uniqueness holds for the parameter values for which:

$$\frac{z_l}{z_k}\left(\frac{\gamma M}{(1-\gamma)(1-M)}\right)^{-\frac{1-a}{a}}\left(\frac{1-a}{a}M^{-1}+1\right) > 1$$

[12] Conditions are presented in the appendix.

pure strategies are pure monopolistic competition and pure oligopoly. Which of the two will emerge depends on the demand elasticity and the equilibrium number of firms. It is also worth noting that, for our results, the assumption of simultaneous choice of output is crucial. In sequential models, if large firms act first, they choose to mimic the behaviour of the fringe (Kokovin et al., 2011).

With (at least) two possible productivity draws, firms accumulate different levels of period-one profits. Therefore productivity heterogeneity that is commonly assumed in trade models emerges as a necessary condition for the existence of mixed markets. This result is similar to Parenti (2013). Parenti develops a partial equilibrium mixed market model with an exogenously given number of firms which are born with the ability to be oligopolistic and multi-product and compete with a monopolistically competitive fringe of single-product firms, the number of which adjusts to a free-entry condition. Although in his model the exogenous cost differences stem from the choice of multi-product firms to benefit from economies of scope, the key implication is the same: without an efficiency advantage all firms will choose to neglect their impact on market outcomes and compete monopolistically. The reason is that, by neglecting their influence on the market, firms compete more aggressively, which turns out to be an advantage among equally productive firms.

However, in our model, a high initial productivity advantage can be endogenously magnified via the decision of a firm to engage in innovation. Innovation plays a dual role: it could be efficiency-enhancing if $z_h < z_k$ and it always shields successful innovators from competition (even in the limit where $z_h = z_k$). Deciding whether to innovate, a firm compares the losses (fixed cost f_k) with the productivity gains and the gains from entering an elite subsector. This case where $z_h < z_k$ mimics a fully dynamic setup where firms will engage in innovation if we allow enough time for the innovation to pay back. Here everything happens within the same period but the idea is the same. This intuition is in line with Peters (2013) who argues that market power affects markets not only statically but also dynamically, since it influences the innovation decisions of firms.

3.2 Numerical comparative statics

Parameter values We do not have a full calibration since this is a very simple model. We normalise $L = 1$ and set our key parameters as follows:

Parameters	Value	Interpretation	Source
a	0.5	Varieties substitutability	hypothetical
β	0.5	Share of homogeneous good	Rauch, 1999
γ	0.01	Percentage of oligopolists	Freund and Pierola, 2015

The rest of our benchmark parameter values are the following: $z_k = 7$ $z_h = 3$, $z_l = 0.1$, $f = 0.001$, $f_k = 0.015$ and $f_e = 0.03$. The selection of the three productivity levels z_l, z_h and z_k is arbitrary. However, as long as the ranking is preserved, a change in the parameter values only rescales the numerical analysis. Our results are robust to changes in the parameters provided subgame perfection is guaranteed. The same holds for the fixed cost of production f and the cost of entry f_e. Given z_l, z_h, z_k, f and f_e, the cost of innovation f_k is selected so that innovation becomes only just unaffordable for a small firm.

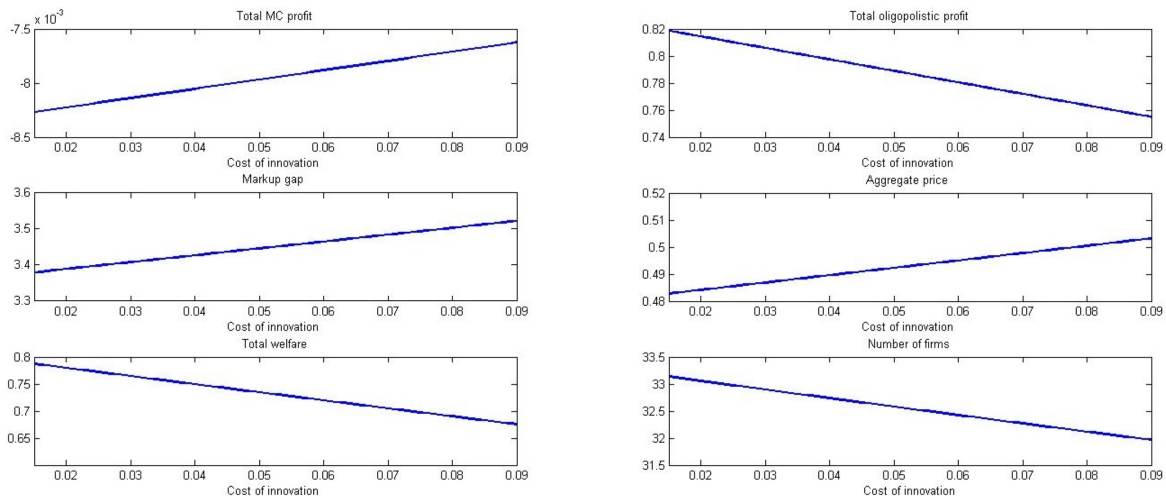

Figure 1: Cost of innovation f_k

Fixed cost of innovation In Figure 1 we present the numerical comparative statics with respect to the fixed cost of innovation f_k. We start from a very low level of f_k that, as mentioned above, makes it marginally unaffordable for MC firms to innovate and then we increase it but only up to a point where it is still optimal to innovate (so as not to violate subgame perfection), given the productivity boost due to innovation. We find that for a fixed boost from z_h to z_k, as innovation becomes more costly the profitability of large firms decreases and this creates a market share reallocation from large towards small firms, since the latter are not harmed by the increase in f_k. In expectation, the decrease of expected oligopolistic profits dominates the increase of expected MC profits driving down the total expected profits prior to entry. As a result, the number of firms in the market decreases. The decrease of the number of firms decreases the competitive pressures within the oligopolistic subsector increasing their markup difference from the small firms. This increase in markups drives an increase in the aggregate price index which causes aggregate welfare to decrease.

Post-innovation productivity Fixing z_h and f_k, an increase in z_k is equivalent to a more efficiency-enhancing innovation. In Figure 2 we present the impact of a change in z_k. We start from $z_k = z_h$ and increase z_k. When $z_k = z_h$ innovation is not cost-reducing but has the sole purpose of allowing the firm to act as a Cournot oligopolist. This is the approach followed by Cellini et al. (2015) who argue that "realising" one's impact on the market implies costly information acquisition and processing. We conclude that as the productivity gap between large and small firms increases, their profitability gap increases accordingly and there is a market share reallocation from the less productive MC firms towards the more productive oligopolists. Again the effect on the oligopolistic profits dominates the effect on monopolistically competitive profits causing expected profits prior to entry to rise. This creates entry which increases the equilibrium number of firms. The increase of N increases competition within the oligopolistic subsector and this is why markups of price over marginal cost do not increase as much as productivity. This decrease in the cost that is only partly incorporated in the markups drives prices down, thus increasing social welfare.

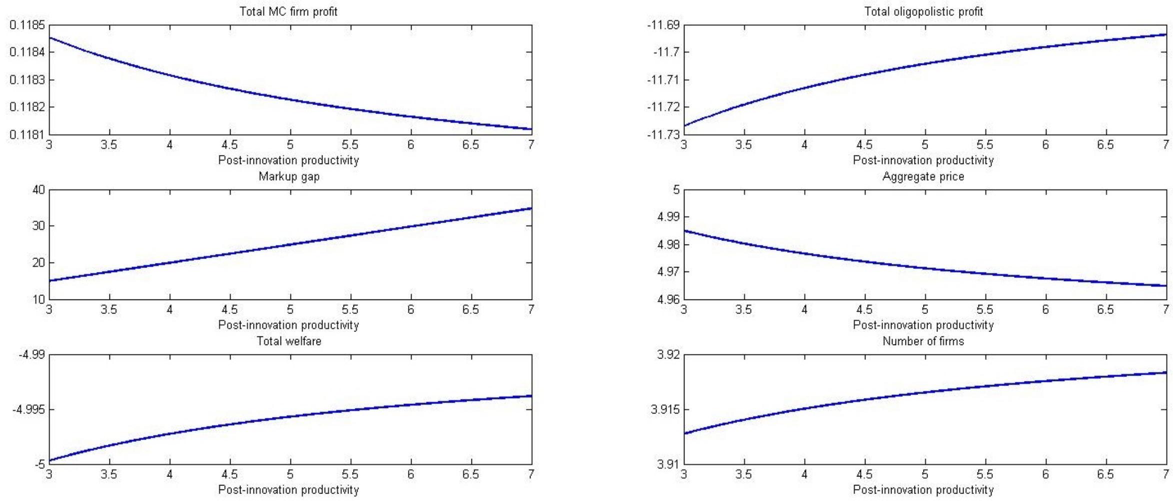

Figure 2: Post-innovation productivity z_k

Percentage of oligopolists Our standard value for the probability of an initial high draw γ, which is essentially the percentage of large firms, is $\gamma = 0.01$. This is motivated by the empirical literature that emphasises the role of the superstar firms, formally defined as the top 1%. However, our model is insensitive to changes in the whole range of possible $\gamma \in (0, 1)$. As $\gamma \to 1$ our model tends to pure oligopoly and as $\gamma \to 0$ it tends to pure monopolistic competition. As γ increases competitive pressures within the oligopolistic subsector increase and the competition among MC firms becomes lower. However, the non-strategic type of the MC firms implies that the competitiveness of the market is driven by the competitiveness of the oligopolistic subsector. Hence, competition increases as γ increases.

In Figure 3 we illustrate how our model behaves as we change γ in the $(0, 1)$ interval. As γ increases markups decrease but at a decreasing rate. This is because from a level of γ on the two subsectors are almost of comparable magnitude and behave very similarly. As $\gamma \to 1$ markups in the market tend to be equalised. They never are perfectly equalised, though, because by construction oligopolists behave differently than the MC firms. The decrease of markups suppresses prices which has a positive effect on the welfare. Profits tend to 0 as each one of the two subsectors tends to disappear. In the oligopolistic subsector, as γ increases competition among large firms, profits decrease as γ increases. However, the profits of the MC firms also decrease as γ increases but this result is different. A higher γ by definition means a smaller fringe since the fringe is proportionately equal to $(1 - \gamma)$. Therefore, an increase in γ decreases the market share of the fringe by suppressing its magnitude. Since there is no exit this will be equivalent to a decrease in the profits.

Finally, we present the relationship between the number of firms and the level of γ. Given that the equilibrium total number of firms adjusts according to the free entry condition, we need to examine the impact of increasing γ on the expected profits. MC firms make positive total operational profits but, when the entry cost is subtracted, their total profits are negative. The reason they do not exit is that they have to repay the sunk fixed entry cost. On the contrary, even after the deduction of the entry cost, oligopolists make positive profits. This means that entry is created from the prospect

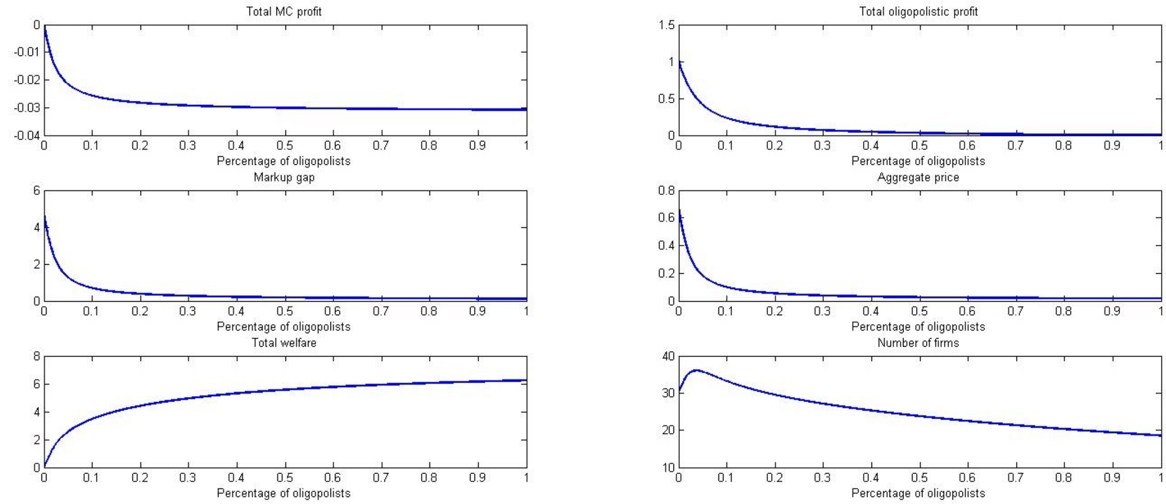

Figure 3: Percentage of oligopolists γ

of the oligopolistic profit. More firms will enter when the oligopolistic profit is higher and when the probability of being an oligopolist and making the high profit is higher. As we increase γ these two effects move in the opposite direction. The high oligopolistic profit decreases but the probability of being the high type increases. We show that for low levels of γ where the competition among large firms is still low, the increase in the probability of being a profitable firm dominates and N increases. However, beyond a level of γ, the oligopolistic subsector is already very competitive, oligopolistic profits have decreased enough and, hence, any further increase in γ will drive the number of firms down. The result is an inverted-U curve which expresses the relationship between the level of competition (which increases with γ) and the level of Romer (1990)-type innovation, as it is reflected in the number of differentiated varieties produced in the economy (N). This inverted-U pattern that peaks at relatively low levels of competition is in line with the endogenous growth literature following Aghion et al. (2002).

4 Open economy

4.1 Production and exports

We now turn to a global economy consisting of two symmetric countries with the same preference and production structure. We assume that the countries open up to trade only in the second period of the game presented above, i.e. only after both economies are "mature" enough to have both large and small firms operating in the differentiated sector. The reason we introduce trade at the second period only is that we want to focus on trade between countries characterised by the empirically supported market competition with a handful of big players and a myriad of small businesses. Trade in the first period would involve two monopolistically competitive industries as in Melitz (2003) and we have nothing new to say here about the gains from this kind of trade. Going back to the analysis of polarised markets, we also assume that all firms of one country compete with all firms of the other, regardless of their size. However closed-economy oligopolists compete oligopolistically when they trade in the global market and, similarly, closed-economy MC firms compete monopolistically when they open up to trade. Trade costs are of the iceberg type: $\tau > 1$ units of goods must be shipped abroad in order for 1 unit to be consumed. We abstract from the presence of a fixed entry cost in the export market and therefore all firms will choose to export. Since the two countries are perfectly symmetric, we can focus on the home one.

Large firms (Cournot oligopolists) Large domestic firm i chooses the quantity sold in the domestic and the foreign market, denoted by x_i^d and x_i^f respectively, to maximise its profits subject to the inverse demand function of each country:

$$\max_{x_i^d, x_i^f} \left\{ \pi_{x\tau i} = \left(p_{xi}^d - \frac{1}{z_k} \right) x_i^d + \left(p_{xi}^f - \frac{\tau}{z_k} \right) x_i^f - f - f_k \right\}$$

$$\text{s.t. } p_{xi}^d = (x_i^d)^{a-1} (H^d)^{1-a} P^d$$

$$\text{and } p_{xi}^f = (x_i^f)^{a-1} (H^f)^{1-a} P^f$$

and taking as given that both domestic and foreign MC firms maximise their profits subject to the inverse demands at home and abroad and that every other oligopolist (foreign and domestic) produces according to their reaction functions.
Substituting the inverse demand, the profit function of a large firm can be re-written as follows:

$$\pi_{x\tau i} = (x_i^d)^a (H^d)^{1-a} P^d - \frac{1}{z_k} x_i^d + (x_i^f)^a (H^f)^{1-a} P^f - \frac{\tau}{z_k} x_i^f - f - f_k$$

Notice that the symmetry assumption implies that the two countries have the same price level ($P^d = P^f = P_\tau$), the same number of oligopolists ($n_x^d = n_x^f = n_{x\tau}$) and also the same level of industry output:

$$H^d = \left[(x_i^d)^a + \sum_{j \neq i} (x_j^d)^a + \sum_{j=0}^{n_{x\tau}} (x_j^f)^a \right]^{\frac{1}{a}} = H^f = \left[(x_i^f)^a + \sum_{j \neq i} (x_j^f)^a + \sum_{j=0}^{n_{x\tau}} (x_j^d)^a \right]^{\frac{1}{a}} = H_\tau$$

where x_0^d and x_0^f represent the aggregate behaviour of the monopolistically competitive subsector.
As in the closed-economy analysis, because $\frac{\partial \pi_{x\tau i}}{\partial x_i^d}$ is strictly decreasing in x_i^d and $\frac{\partial \pi_{x\tau i}}{\partial x_i^f}$ is strictly decreasing in x_i^f we have that, from Lemma 1, the best response functions of domestic oligopolist i are the unique solution of each one of the following first order conditions:

$$\frac{\partial \pi_{x\tau i}}{\partial x_i^d} = \frac{a \left(\sum_{j \neq i} (x_j^d)^a + \sum_{j=0}^{n_{x\tau}} (x_j^f)^a \right)}{(x_i^d)^{1-a} \left((x_i^d)^a + \sum_{j \neq i} (x_j^d)^a + \sum_{j=0}^{n_{x\tau}} (x_j^f)^a \right)^2} E_\tau - \frac{1}{z_k} = 0$$

$$\frac{\partial \pi_{x\tau i}}{\partial x_i^f} = \frac{a \left(\sum_{j \neq i} (x_j^f)^a + \sum_{j=0}^{n_{x\tau}} (x_j^d)^a \right)}{(x_i^f)^{1-a} \left((x_i^f)^a + \sum_{j \neq i} (x_j^f)^a + \sum_{j=0}^{n_{x\tau}} (x_j^d)^a \right)^2} E_\tau - \frac{\tau}{z_k} = 0$$

Small firms (MC firms) Small domestic firm i chooses the level of output supplied in the domestic market y_i^d the following:

$$\max_{y_i^d, y_i^f} \left\{ \pi_{yi\tau} = \left(p_{yi}^d - \frac{1}{z_l} \right) y_i^d + \left(p_{yi}^f - \frac{\tau}{z_l} \right) y_i^f - f \right\}$$

$$\text{s.t. } p_{yi}^d = (y_i^d)^{a-1} (H^d)^{1-a} P^d$$

$$\text{and } p_{yi}^f = (y_i^f)^{a-1} (H^f)^{1-a} P^f$$

Exactly as in the analysis of the oligopolistic behaviour, $P^d = P^f = P_\tau$, $n_y^d = n_y^f = n_{y\tau}$ and $H^d = H^f = H_\tau$. Omitting the i-index, since in equilibrium all MC firms make the same optimal decisions, and applying the f.o.c. yields the equilibrium output sold domestically and abroad:

$$y^d(E_\tau, P_\tau; \tau, z_l) = \left(\frac{1}{az_l} \right)^{\frac{-1}{1-a}} E_\tau P_\tau^{\frac{a}{1-a}}$$

$$y^f(E_\tau, P_\tau; \tau, z_l) = \left(\frac{\tau}{az_l} \right)^{\frac{-1}{1-a}} E_\tau P_\tau^{\frac{a}{1-a}}$$

which yield the equilibrium prices charged in the domestic and the export market:

$$p_y^d(z_l) = \frac{1}{az_l}$$

$$p_y^f(z_l) = \frac{\tau}{az_l} = \tau p_y^d$$

Note each domestic small firm sets a domestic price equal to the one in the closed-economy setup but a higher price is set in the foreign market, reflecting the increased marginal cost of supplying to this market.

The equilibrium profit of every small firm is given by the following expression :

$$\pi_{y\tau}(E_\tau, P_\tau; \tau, z_l) = \left(\frac{1}{az_l} \right)^{\frac{a}{a-1}} (1-a) E_\tau P_\tau^{\frac{a}{1-a}} + \left(\frac{\tau}{az_l} \right)^{\frac{a}{a-1}} (1-a) E_\tau P_\tau^{\frac{a}{1-a}} - f$$

4.2 Open economy equilibrium

As in the closed-economy case, we consider a non-cooperative game in which big and small firms choose their output simultaneously. We define the market equilibrium exactly as specified above. We characterise a mixed market equilibrium using the profit maximisation conditions of the small and large firms as well as the free entry condition.

We consider a *symmetric* equilibrium where all large firms choose the same output, foreign and domestic (i.e. $x_i^d \equiv x^d$ and $x_i^f \equiv x^f$ for every $i = 1, 2, \ldots, n_{x\tau}$), sold at the same corresponding price (i.e. $p_{xi}^d \equiv p_x^d$ and $p_{xi}^f \equiv p_x^f$ for every $i = 1, 2, \ldots, n_{x\tau}$) and, hence, earn the same profits (i.e. $\pi_{xi\tau} \equiv \pi_{x\tau}(\tau, z_k)$ for every $i = 1, 2, \ldots, n_{x\tau}$). The equilibrium conditions for the open economy (for given N_τ) along with equations (1)-(3) for each one of the two countries form the following system of 10 equations in 10 unknowns: the per-period profits ($\pi_{y1}(z_h)$, $\pi_{y1}(z_l)$, $\pi_{x\tau}(\tau, z_k)$ and $\pi_{y\tau}(\tau, z_l)$), the price indices (P_1 and P_τ), the levels of expenditure on the differentiated good ($E_1$and E_τ) and the levels of output sold by an oligopolist in the domestic (x^d) and the foreign market (x^f) once costly trade is allowed. The total first-period expenditure in the global economy is exactly the same as in the closed-economy setup, given that in the first period there is no international trade and in the second period there is free trade for the homogeneous good:

$$E_1 = \frac{L - N_\tau f}{\beta + 1 - \left[\gamma \left(\frac{1}{az_h}\right)^{\frac{-a}{1-a}} + (1-\gamma)\left(\frac{1}{az_l}\right)^{\frac{-a}{1-a}}\right]^{-1} \left[(1-\gamma)\left(\frac{1-a}{(az_l)^{\frac{a}{a-1}}}\right) + \gamma\left(\frac{1-a}{(az_h)^{\frac{a}{a-1}}}\right)\right]} \tag{11}$$

Open-economy profits:

$$\pi_{x\tau}(\tau, z_k) = (x^d)^a E_\tau^{1-a} P_\tau^a - \frac{1}{z_k} x^d + (x^f)^a E_\tau^{1-a} P_\tau^a - \frac{\tau}{z_k} x^f - f - f_k \tag{12}$$

$$\pi_{y\tau}(\tau, z_l) = \left(\frac{1}{az_l}\right)^{\frac{a}{a-1}} (1-a) E_\tau P_\tau^{\frac{a}{1-a}} + \left(\frac{\tau}{az_l}\right)^{\frac{a}{a-1}} (1-a) E_\tau P_\tau^{\frac{a}{1-a}} - f \tag{13}$$

From the two optimality conditions for the open-economy profit maximisation of a large firm we get the conditions linking E_τ, P_τ, x^d and x^f:

$$E_\tau^{1-a} = \frac{1}{az_k} P_\tau^{-a} (x^d)^{1-a} + E_\tau^{1-2a} P_\tau^a (x^d)^a \tag{14}$$

$$E_\tau^{1-a} = \frac{\tau}{az_k} P_\tau^{-a} (x^f)^{1-a} + E_\tau^{1-2a} P_\tau^a (x^f)^a \tag{15}$$

Open-economy market aggregates:

$$P_\tau = E_\tau \left\{ (1-\gamma) N_\tau \left[\left(\frac{1}{az_l}\right)^{\frac{-1}{1-a}} E_\tau P_\tau^{\frac{a}{1-a}}\right]^a + (1-\gamma) N_\tau \left[\left(\frac{\tau}{az_l}\right)^{\frac{-1}{1-a}} E_\tau P_\tau^{\frac{a}{1-a}}\right]^a + \gamma N_\tau (x^d)^a + \gamma N_\tau (x^f)^a \right\}^{\frac{-1}{a}} \tag{16}$$

$$E_\tau = (\beta + 1)^{-1} \left(L + \gamma N_\tau \pi_{x\tau}(\tau, z_k) + (1-\gamma) N_\tau \pi_{y\tau}(\tau, z_l)\right) \tag{17}$$

Using P_τ we get that:

$$(x^d)^a + (x^f)^a = E_\tau^a \left[\frac{1 - (az_l P_\tau)^{\frac{a}{1-a}} (1-\gamma) N_\tau (1 + \tau^{\frac{-\alpha}{1-\alpha}})}{\gamma N_\tau P_\tau^a}\right]$$

Equations (14), (15) and (17) along with the above expression yield the equilibrium values of x^d, x^f, E_τ and P_τ as functions of N_τ only. Using (11) and the free entry condition:

$$E\Pi(N_\tau) = 0 \tag{18}$$

we can compute the open-economy equilibrium number of firms operating in each country.

For the welfare analysis we follow the same steps as in the closed economy. We calculate the indirect utility of the representative consumer which, given homothetic preferences, will describe the social welfare. For the evaluation of welfare gains from trade we use the compensating variation ω of going from the trade equilibrium Ω_1 to the autarky equilibrium Ω_2. It can be expressed as follows:

$$\ln \omega = U(\Omega_1) - U(\Omega_2)$$

where U is the welfare function described above. For the evaluation of a trade policy regime we use the exact same approach where Ω_1 is the equilibrium under the policy regime and Ω_2 is the equilibrium without the policy.

5 Discussion of the open economy

5.1 Iceberg trade cost

We consider a scenario where there are no fixed costs to export. In the presence of fixed costs we would get one of two possible cases. Either the fixed cost of export would be too high for the small firms, in which case we would have trade only among oligopolists or the fixed costs would be low enough for all firms to export, in which case the no-fixed cost analysis is exactly applicable, only scaled down. The case where only oligopolists trade can be embedded as the extreme version of our setup.

In Figures 4 - 6 we illustrate the effects of trade liberalisation. We start from free trade where $\tau = 1$ and increase τ. Trade liberalisation is captured as a decrease of the iceberg cost τ and so our figures should be read from right to left. We set our parameters as follows: $z_k = 7$ $z_h = 3$, $z_l = 0.1$, $f = 0.001$, $f_k = 0.015$ and $f_e = 0.03$. Due to its increased complexity, the open-economy simulations are more sensitive compared to the closed economy model. However, our results are robust within a wide range of parameter values.

As τ decreases and trade liberalisation proceeds, the export penetration from both small and large foreign firms increases. This decreases the market shares of all types of domestic firms in the home market and increases the shares of all types of foreign firms. With CES, MC firms cannot adjust their markups, as opposed to the oligopolists who can. As in the standard oligopolistic models, oligopolists will engage in dumping by charging a lower markup abroad than at home (see Figure 5). More importantly, as τ decreases, large firms lower their markups in the domestic market because of increased competition and increase them in the foreign because of decreased export costs. Differential abilities to adjust markups give rise to a composition effect of the two traditional sources of gains, namely the variety gains predicted by monopolistically competitive models and the pro-competitive gains which are expected to materialise under oligopoly. Our predictions from a model with mixed markets differ from the predictions of both the Krugman (1980)- and the Brander and Krugman (1983)-type models. In the presence of market power heterogeneity, there is a new channel through which trade liberalisation affects welfare. This channel works through the differential abilities of firms to price-to-market and materialises a market share reallocation from small towards large firms.

In Figures 4 and 5 we focus on the markup adjustment that takes place as trade liberalisation proceeds. Notice that variable markups imply that the change in trade costs is only partly being passed on to the prices. Prices decrease but less than the decrease in trade costs. The decrease in prices creates an increase in aggregate welfare and the welfare gains from trade (in comparison to the autarky equilibrium). Finally average markups decrease but only marginally. The increase in foreign markups is slightly lower that than the decrease in domestic markups. This implies that some pro-competitive gains from trade survive but they are severely diminished due to market power asymmetries. Average markup and markup dispersion both decrease as τ decreases. Overall, as τ decreases, MC firms become less profitable whereas oligopolistic firms increase their profits. The pro-competitive effect of trade liberalisation is dampened. This is because, in the presence of a competitive fringe, the fringe largely absorbs the competitive pressures. The increase in the oligopolistic profits dominates the decrease in the monopolistically competitive profits causing expected profits prior to entry to increase and so a higher number of firms enters the market. Even though the number of firms increases, intensifying competition among large firms, this is not sufficient to overturn the increase in oligopolistic profits due to trade liberalisation. Comparing the number of available varieties (N versus $2N_\tau$), variety gains from trade survive, although the adjustment of N as τ changes is again because of the existence of market power[13].

The idea of incomplete pass-through has been explored in a number of theoretical and empirical works. The pattern is generated with a variety of assumptions on the demand and market structure. For example, Atkeson and Burnstein (2008) assume CES preferences and Cournot oligopoly, Melitz and Ottaviano (2008) assume linear demand and monopolistic competition, Goldberg and Verboven (2005) use nested logit and Bertrand oligopoly and Goldberg and Hellerstein (2013) use Bertrand with random coefficients[14]. In all the aforementioned papers all firms have the ability to charge variable markups and any differences in markups across firms are a result of productivity heterogeneity. In our case, the existence of a productivity gap alone leads to zero markup dispersion. The co-existence of fixed- and variable-markup firms is crucial. This result becomes obvious when we compare the levels of welfare for different levels of γ (Figure 6). We compare welfare gains for different values of γ and conclude that as concentration increases (γ decreases) and large firms are more protected from competition, welfare gains from trade decrease, although trade liberalisation always increases welfare. These results are in line with Mayer et al. (2011) and Edmond et al. (2012) who find that markups decrease due to pro-competitive gains from trade. However, contrary to the findings of Edmond et al. (2012), pro-competitive gains are lower in the presence of extensive misallocation which is due to large inefficiencies associated with markups. In fact, in the presence of a very low γ, welfare gains from trade are so low that we could even speculate that in a richer setup that would allow for the analysis of inputs as well as outputs markets trade liberalisation could even lead to markup increase. De Loecker et al. (2015) find that markups actually increase as a result of trade liberalisation because trade liberalisation not only reduces a firm's cost of exporting its output but it also reduces the cost of imported inputs. Given that superstars tend to make more intensive use of imported inputs (see for example Bernard et al. (2015)), allowing for input considerations, would suppress the already limited pro-competitive gains even further possibly to the point of even overturning the effect on markups in

[13]See Krugman (1980).

[14]For a survey see De Loecker and Goldberg (2014). Other papers include Feenstra and Weinstein (2010), Mayer et al. (2011) and Arkolakis et al. (2015).

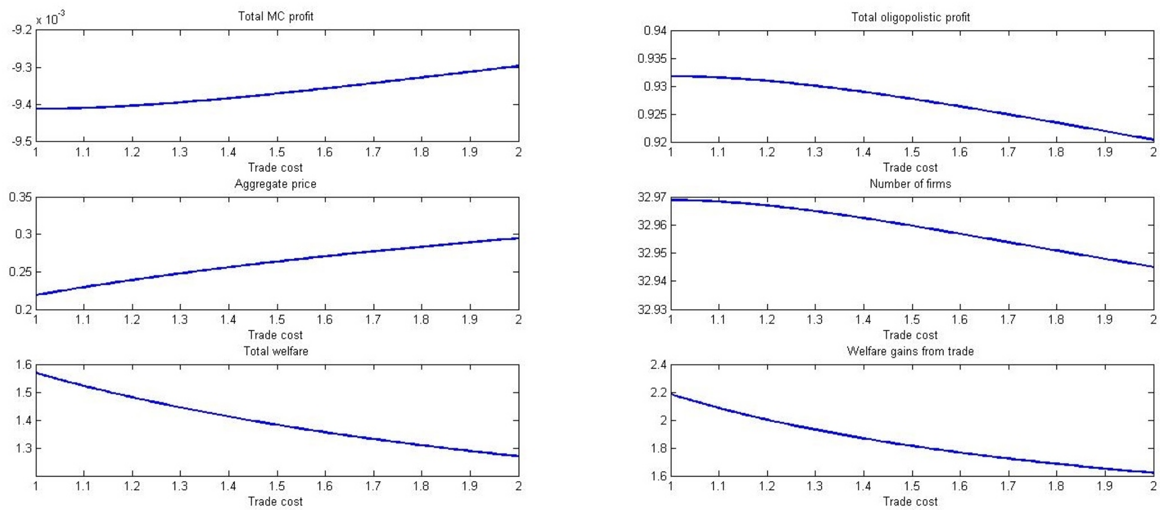

Figure 4: Trade liberalisation

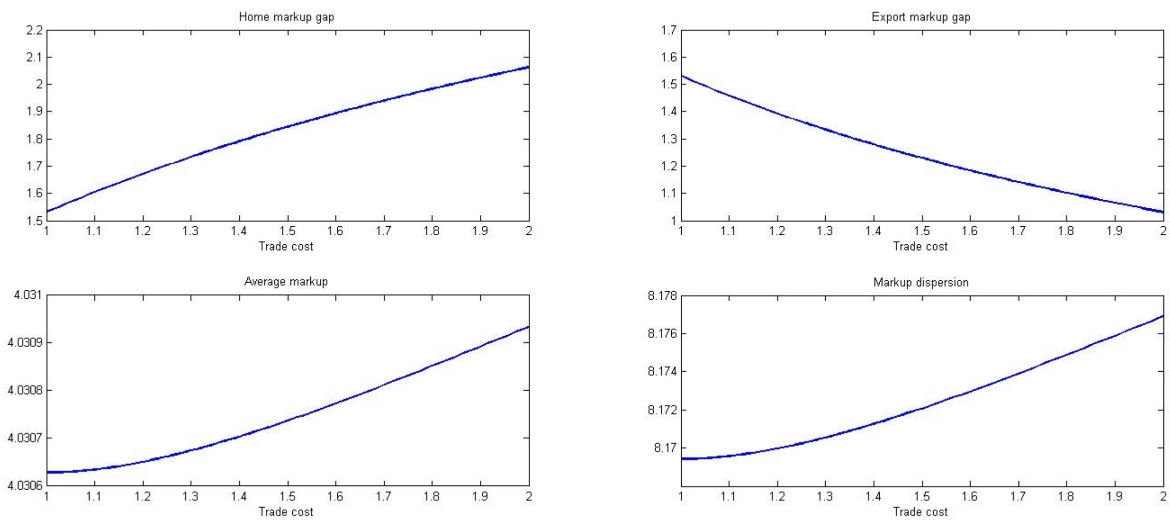

Figure 5: Trade liberalisation - Markup analysis

line with De Loecker et al. (2015). However, a rigorous input market analysis lies beyond the scope of this paper.

Finally, our model shares the same intuition as Parenti (2013). However, without a free entry condition in the oligopolistic subsector, in Parenti trade liberalisation only works through the adjustment of the competitive fringe. In that extreme case, small firms completely absorb all the pro-competitive effects driving consumer surplus down[15]. Our paper relaxes this result: welfare increases with trade liberalisation but the pro-competitive gains are reduced due to a new channel of market share reallocation which becomes active only when we explicitly allow for differences in market power.

[15]There is also a number of oligopolistic models arguing that, in the presence of firm heterogeneity, trade liberalisation could decrease welfare. See for example Bekkers and Francois (2013).

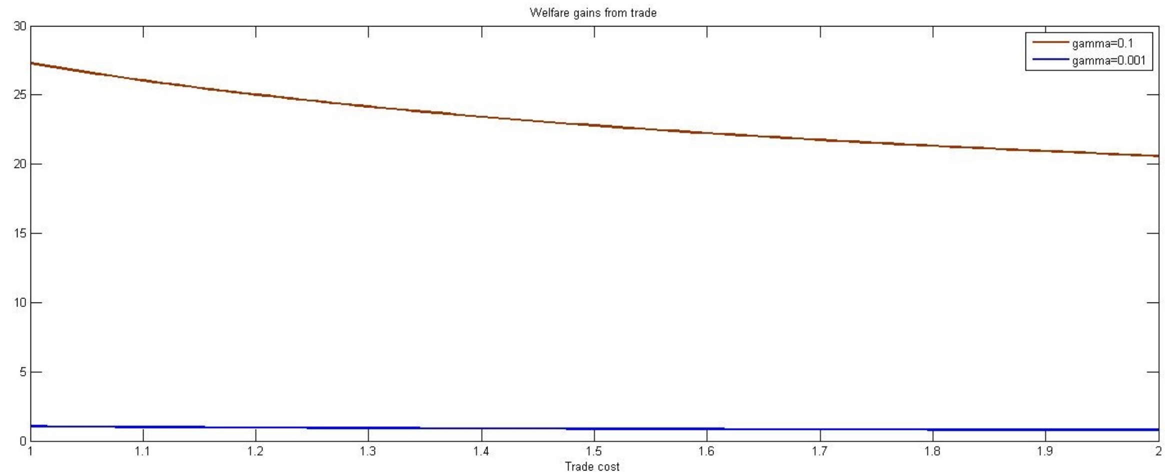

Figure 6: Gains from trade and market concentration

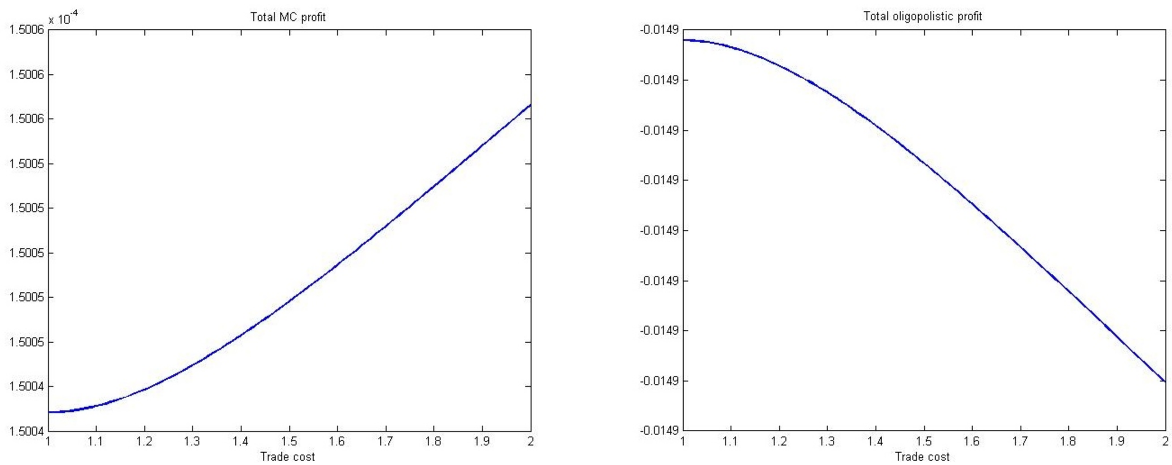

Figure 7: Inefficiency

5.2 Policy implications

Going back to the markup analysis under trade liberalisation, the occurring market share reallocation is partly due to an exit-less Melitz (2003)-type selection. Oligopolistic firms are more efficient and therefore they can expand more as trade becomes less costly. Selection-related reallocation is welfare-enhancing and any policy intervention raising obstacles to this mechanism is, without doubt, a bad idea. However, selection is not the whole story. In Figure 7 we present an exercise where all firms in the market are equally productive. For example we set $z_l = z_h = z_k = 3$. In this scenario (which is not subgame perfect since oligopolistic firms are not behaving optimally when they choose to incur f_k despite their lack of productivity advantage) market share reallocation still occurs. However this is an inefficiency triggered by trade liberalisation. Even without a cost advantage, strategic firms will always set a higher markup and will benefit from pricing-to-market. As a result, our model could potentially serve as a justification for size-dependent policies.

Our market structure highlights the complementarities between trade and competition policy that are yet to be exploited. In the presence of the inefficiency presented above we go on to investigate whether size-dependent trade policies could magnify the welfare gains from trade. This type of government policies that restrict large establishments and/or promote small ones are widespread across countries. They take different forms including trade restrictions and there is a large literature, for example Restuccia and Rogerson (2008) and Guner et al. (2008) aiming at evaluating costs and benefits from such policies as well as their impact on the size distribution of firms. Following the neoclassical growth model with heterogeneous productivities the findings from these works tend to agree that size-dependent policies are harmful both in terms of aggregate output and average productivity. In this section we show that, in the presence of strategic market power heterogeneity, size-dependent policies might actually raise social welfare.

Our policy is simple and abstract taking the form of an export subsidy. The level of the subsidy is revealed once each firm has draw an initial productivity. Government budget balances through a lump-sum tax T to the consumers, whose lifetime income drops by T. We compare the results from a subsidy to the low-productivity firms to the situation where all firms face the same trade cost, for example $\tau = 2$. In Figures 8 and 9 we present the effect of an export subsidy to the monopolistically competitive firms where the subsidy takes the form of a decreased trade cost faced by these firms. In other words, oligopolistic firms face the real trade cost $\tau = 2$ whereas the MC firms face a trade cost $\tilde{\tau} \in [1, 2]$. The difference between the two costs $\tau - \tilde{\tau}$ is equal to the subsidy per unit of exported output. The lump sum tax imposed on consumers is equal to:

$$T = (1-\gamma) N_\tau \left(\frac{\tilde{\tau}}{a z_l} \right)^{\frac{-1}{1-a}} E_\tau P_\tau (\tau - \tilde{\tau})$$

We find that contrary to the homogeneous market power case, a size-depend policy like the one described above leads to an increase in output which drives aggregate prices down. Reading Figures 8 and 9 from right to left reveals that substituting small firms suppresses the average markup and the markup dispersion by decreasing both the home and the foreign markup gap between small and large firms. As a result, a market share reallocation of the opposite direction from the one that takes place during trade liberalisation increases the monopolistically competitive profit in the expense of the oligopolistic subsector. As a result, aggregate welfare increases and the welfare gains from the policy increase with the magnitude of the subsidy. Finally, subsidising MC firms leads to a decrease in the expected profits prior to entry which are, again, driven by the decrease in oligopolistic profits. Hence, although they are found to be socially beneficial, size-dependent policies in our setup will discourage entry.

To conclude, we should note that, although in our simple setup it is beneficial to promote size-dependent policies supporting SMEs, we should be careful in our evaluation of such practices. On the one hand lower concentration increases welfare but, on the other, in a dynamic setup, size-dependent policies could harm innovation incentives and there is an important although nuanced trade-off between efficiency gains from superstar's innovation and welfare losses because of decreased competition and markup-related misallocation. This concern emphasises the need to design and pursue trade reforms jointly with competition and innovation policy.

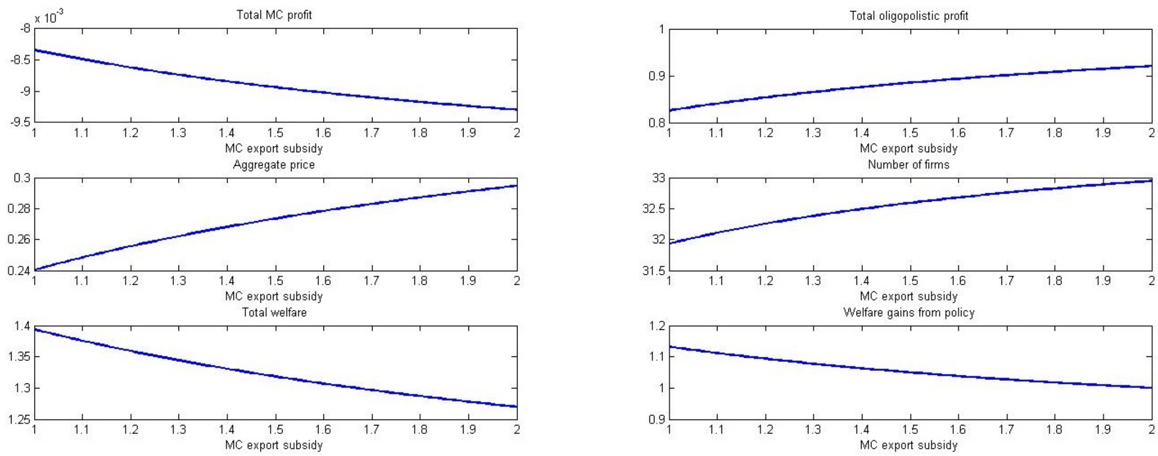

Figure 8: Size-dependent export subsidy

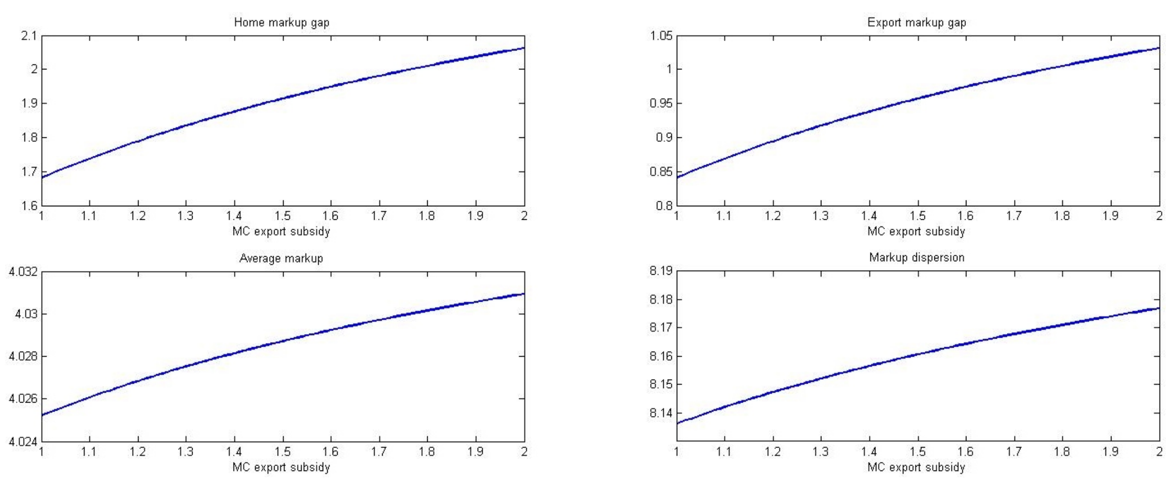

Figure 9: Size-dependent export subsidy - Markup analysis

6 Conclusion

Based on the assumption that firms grow because of luck and successful innovation, we construct a two-period general equilibrium model with free entry where some firms choose to acquire market power and behave as Cournot oligopolists whereas others remain non-strategic and compete monopolistically. Conditions are derived for the this asymmetry to emerge as the subgame perfect equilibrium: exogenous efficiency differences are endogenously magnified in a dynamic setup leading to differences in strategic behaviour. This setup reproduces the empirically confirmed residual markup inequality which cannot be attributed to differences in productivity. We then use this model to investigate the impact of bilateral trade liberalisation on a polarised industry with both significant and negligible firms. We show that the result from liberalising a market with a monopolistically competitive and an oligopolistic subsector is not a simple combination of the variety gains predicted by monopolistically competitive models with the pro-competitive gains which are expected to materialise under pure oligopoly. In fact, the presence of firms with heterogeneous abilities to adjust their markups opens up a new channel of market share reallocation from the monopolistic competitors to the oligopolists that tends to significantly suppresses the pro-competitive welfare gains from trade. As trade becomes less costly, oligopolists benefit more from exporting to the foreign market than they are harmed by the increase in domestic competition because most of the competitive pressures are absorbed by the monopolistically competitive fringe. As a result, as long as trade involves an equally polarised economy, superstars are not threatened enough by the competitive pressures driven by trade liberalisation.

The intuition goes in the direction of Stiglitz and Greenwald (2012) in that trade liberalisation alone would not ensure output growth. Liberalising a concentrated industry is only marginally beneficial in the aggregate while superstars are the big winners. This result is in line with Epifani and Garcia (2011) who find that heterogeneity among firms increases overtime and trade is partly responsible. In that sense, domestic competition policy and trade policy might have a role to play. We show that size-dependent trade policies could be welfare-enhancing leading to a decrease in the markup dispersion.

Further research could exploit the possibilities of competition policy. In our model, this could work through manipulations of the fixed cost of entering the oligopolistic subsector (f_k) which could be interpreted as a reduced-form representation of the policy instruments that could potentially affect the cost of acquiring market power. As Peters (2013) observes, there is much more to gain from reducing entry barriers than from reducing marginal trade costs.

References

[1] Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2002). *Competition and innovation: An inverted U relationship* (No. w9269). National Bureau of Economic Research.

[2] Aggarwal, R. K., & Samwick, A. A. (1999). Executive compensation, strategic competition, and relative performance evaluation: Theory and evidence. *The Journal of Finance*, 54(6).

[3] Anderson, S.P., De Palma, A. & Thisse, J.F. (1992) *Discrete Choice Theory of Product Differentiation.* Cambridge, Mass.: MIT Press.

[4] Arkolakis, C., Costinot, A., Donaldson, D., Rodríguez-Clare, A. (2015) *The elusive pro-competitive effects of trade* (No. w21370). National Bureau of Economic Research.

[5] Atkeson, A., & Burstein, A. (2008). Pricing-to-market, trade costs, and international relative prices, *The American Economic Review*, 98(5), 1998-2031.

[6] Bekkers, E., & Francois, J. (2013). Trade and industrial structure with large firms and heterogeneity. *European Economic Review*, 60, 69-90.

[7] Bernard, A. B., Jensen, J. B., Redding, S. J., & Schott, P. K. (2007). *Firms in international trade* (No. w13054), National Bureau of Economic Research.

[8] Bernard, A. B., Jensen, J. B., Redding, S. J., & Schott, P. K. (2015). Global firms.

[9] Bombardini, M. (2008). Firm heterogeneity and lobby participation. *Journal of International Economics*, 75(2), 329-348.

[10] Brander, J., & Krugman, P. (1983). A 'reciprocal dumping'model of international trade, *Journal of International Economics*, 15(3), 313-321.

[11] Cellini, R., Lambertini, L., & Ottaviano, G. I. (2015). To Know or Not To Know: Strategic Inattention and Endogenous Market Structure.

[12] De Loecker, J., & Goldberg, P. K. (2014). Firm Performance in a Global Market. *Annu. Rev. Econ.*, 6(1), 201-227.

[13] De Loecker, J., Goldberg, P. K., Khandelwal, A. K., & Pavcnik, N. (2015). Prices, markups and trade reform, *Econometrica*, Forth.

[14] De Loecker, J., & Warzynski, F. (2012). "Markups and firm-level export status," *The American Economic Review*, 102(6), 2437-2471.

[15] Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2002). The regulation of entry. *Quarterly journal of Economics,* 1-37.

[16] Dixit, A. K., & Stiglitz, J. E. (1977). "Monopolistic competition and optimum product diversity," *The American Economic Review*, 297-308.

[17] Edmond, C., Midrigan, V., & Xu, D. Y. (2012). *Competition, markups, and the gains from international trade* (No. w18041). National Bureau of Economic Research.

[18] Epifani, P., & Gancia, G. (2011). "Trade, markup heterogeneity and misallocations," *Journal of International Economic*s, 83(1), 1-13.

[19] Feenstra, R. C., & Weinstein, D. E. (2010). *Globalization, markups, and the US price level* (No. w15749). National Bureau of Economic Research.

[20] Freund, C., & Pierola, M. D. (2015). Export superstars. *Review of Economics and Statistics*, 97(5), 1023-1032.

[21] Friedman, J. W. (1977). *Oligopoly and the Theory of Games* (Vol. 8). North-Holland.

[22] Goldberg, P. K., & Verboven, F. (2005). Market integration and convergence to the Law of One Price: evidence from the European car market. *Journal of international Economics*, 65(1), 49-73.

[23] Goldberg, P., & Hellerstein, R. (2013). A Structural Approach to Identifying the Sources of Local Currency Price Stability. *The Review of economic studies*, 80(1), 175-210.

[24] Greenwald, B., & Stiglitz, J. E. (2012). *Creating a learning society: a new paradigm for development and social progress*, First Annual Arrow Lecture, Columbia University. New York: Columbia University Press.

[25] Guner, N., Ventura, G., & Xu, Y. (2008). Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics*, 11(4), 721-744.

[26] Hall, R. E. (1986). "Chronic excess capacity in US industry," *NBER Working Paper*, (w1973).

[27] Igami, M. (2015). Offshoring under oligopoly. Available at SSRN 1726910.

[28] Impullitti, G., & Licandro, O. (2016). Trade, firm selection, and innovation: the competition channel.

[29] Kokovin, S., Parenti, M., Thisse, J. F., & Zhelobodko, E. (2011). "Monopolistic competition with big firms," *Working Paper*.

[30] Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. The American Economic Review, 70(5), 950-959.

[31] Mayer, T., Melitz, M. J., & Ottaviano, G. I. (2011). *Market size, competition, and the product mix of exporters* (No. w16959). National Bureau of Economic Research.

[32] Mayer, T., & Ottaviano, G. I. (2008). "The happy few: The internationalisation of European firms," *Intereconomics*, 43(3), 135-148.

[33] Melitz, M. J. (2003). "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, 71(6), 1695-1725.

[34] Melitz, M. J., & Ottaviano, G. I. (2008)." Market size, trade, and productivity," *The Review of Economic Studies*, 75(1), 295-316.

[35] Melitz, M. J., & Redding, S. J. (2014). Heterogeneous Firms and Trade. *Handbook of International Economics*, 4, 1.

[36] Neary, P. (2010). "Two and a half theories of trade," *The World Economy*, 33(1), 1-19.

[37] Neary, P. (2015). International Trade in General Oligopolistic Equilibrium (No. 769).

[38] Pan, L., & Hanazono, M. (2015). Competition among the Big and the Small with Different Product Substitution.

[39] Parenti, M. (2013). *Large and small firms in a global market: David vs. Goliath* (No. 2013058), Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

[40] Peters, M. (2013). *Heterogeneous mark-ups, growth and endogenous misallocation.* London School of Economics and Political Science, LSE Library.

[41] Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of international Economics*, 48(1), 7-35.

[42] Restuccia, D., & Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, 11(4), 707-720.

[43] Romer, P. M. (1990). Endogenous Technological Change. Journal of Political Economy, 98(5 pt 2).

[44] Shapiro, M. D. (1987). *Measuring market power in US industry* (No. 2212). National Bureau of Economic Research, Inc.

[45] Shimomura, K. I., & Thisse, J. F. (2012). "Competition among the big and the small," *The RAND Journal of Economics*, 43(2), 329-347.

[46] Tamminen, S., & Chang, H. H. (2013). "Firm and sectoral heterogeneity in markup variability," *The Journal of International Trade & Economic Development*, 22(1), 157-178.

[47] Gordon, S. (2016, March 3). *The British businesses that relish prospect of life outside EU.* Retrieved from http://www.ft.com

Appendix

A Proofs of equilibrium properties

Existence (Propositions 1 and 2) Using P_2 we get that:

$$x_2 = E_2 \left[\frac{1 - (az_l P_2)^{\frac{a}{1-a}} (1-\gamma) N}{\gamma N P_2^a} \right]^{\frac{1}{a}} = x_2(E_2, P_2, N)$$

To simplify notation we define:

$$M \equiv (az_l P_2)^{\frac{a}{1-a}} (1-\gamma) N$$

Therefore we can re-write x_2 as follows:

$$x_2 = E_2 P_2^{-1} \left[\frac{1-M}{\gamma N} \right]^{\frac{1}{a}}$$

Substituting the above in equation (7) we get:

$$1 = \frac{1}{az_k} P_2^{-1} \left(\frac{1-M}{\gamma N} \right)^{\frac{1-a}{a}} + \frac{1-M}{\gamma N}$$

Solving the definition of M with respect to P_2:

$$P_2 = \frac{1}{az_l} \left(\frac{M}{(1-\gamma)N} \right)^{\frac{1-a}{a}}$$

Substituting P_2 in the above expression yields:

$$\frac{1 - \frac{z_l}{z_k} \left(\frac{1-\gamma}{\gamma} \frac{1-M}{M} \right)^{\frac{1-a}{a}}}{1-M} = \frac{1}{\gamma N}$$

Define:

$$G(M) \equiv \frac{1 - \frac{z_l}{z_k} \left(\frac{1-\gamma}{\gamma} \frac{1-M}{M} \right)^{\frac{1-a}{a}}}{1-M}$$

The function G increases with M and is such that $\lim_{M \to 0^+} G(M) = -\infty$ and $\lim_{M \to 1^-} G(M) = \infty$. Therefore, for any given N the equation:

$$G(M) = \frac{1}{\gamma N}$$

has a unique solution $M(N) \in (0,1)$ which decreases with N. It then follows from the definition of M that $P_2(N)$ decreases with N. The inverse demand function faced by an oligopolist is the following:

$$p_{x2} = x_2^{a-1} E_2^{1-a} P_2^a$$

Substituting x_2 and using the definition of M, we get that the equilibrium price $p_{x2}(N)$ set by a large firm is also a decreasing function of N since it can be written as:

$$p_{x2} = \frac{1}{az_l} \left(\frac{\gamma M}{(1-\gamma)(1-M)} \right)^{\frac{1-a}{a}}$$

We go on to express the second-period profits of each type of firms as a function of N only. Equation (5) gives the profit of a large firm, which can be writes as:

$$\pi_{x2}(z_k) = E_2 \left(\frac{1-M}{\gamma N} \right) - \frac{1}{z_k} E_2 P_2^{-1} \left(\frac{1-M}{\gamma N} \right)^{\frac{1}{a}} - f - f_k$$

Substituting the above and (6) into (9) gives the equilibrium income spent on the differentiated good:

$$E_2(N) = \frac{L - \gamma N f_k - N f}{\beta + R(N)}$$

where:

$$R(N) \equiv a\frac{z_l}{z_k}\left(\frac{\gamma}{1-\gamma}\right)^{\frac{a-1}{a}}(1-M(N))^{\frac{1}{a}}M(N)^{\frac{a-1}{a}}+aM(N)$$

Having expressed P_2 and E_2 in terms of N, we can use (6) and (5) to obtain $\pi_{y2}(z_l)$ and $\pi_{x2}(z_k)$ as a function of N only:

$$\pi_{y2}(N;z_l)=\frac{(1-a)M(N)}{(1-\gamma)N}E_2(N)-f$$

$$\pi_{x2}(N;z_k)=\left[\frac{1-M(N)}{\gamma N}-\frac{1}{z_k}P_2(N)^{-1}\left(\frac{1-M(N)}{\gamma N}\right)^{\frac{1}{a}}\right]E_2(N)-f-f_k$$

Using the above results we can write the total expected profits, prior to entry, as a function of N:

$$E\Pi(N)=\frac{1}{N}\left[(1-\gamma)N\left[\pi_{y1}(N;z_l)+\pi_{y2}(N;z_l)\right]+\gamma N\left[\pi_{x1}(N;z_h)+\pi_{x2}(N;z_k)\right]-Nf_e\right]$$

which can be written as:

$$E\Pi(N)=\frac{1}{N}\left[(\beta+1)\left[E_1(N)+E_2(N)\right]-2L-Nf_e\right]$$

which is equivalent to:

$$E\Pi(N)=\frac{1}{N}\left[-2L-Nf_e\right]+$$

$$\frac{(\beta+1)}{N}\left[\frac{L-Nf}{\beta+1-\left[\gamma\left(\frac{1}{az_h}\right)^{\frac{-a}{1-a}}+(1-\gamma)\left(\frac{1}{az_l}\right)^{\frac{-a}{1-a}}\right]^{-1}\left[(1-\gamma)\left(\frac{1-a}{(az_l)^{\frac{a}{a-1}}}\right)+\gamma\left(\frac{1-a}{(az_h)^{\frac{a}{a-1}}}\right)\right]}+\frac{L-\gamma Nf_k-Nf}{\beta+R(N)}\right]$$

Now we have to show that there exists a value of N, denoted by N^*, that solves our system of equations. This N^* is the solution of the free entry condition:

$$E\Pi(N^*)=0$$

Given that for any given N we have that $M(N)\in(0,1)$, we get that $\lim_{N\to\infty}E\Pi(N)=-\infty$ and since $E\Pi(0)>0$ we have that there exists at least one $N^*>0$ such that (10) holds.

Uniqueness (Proposition 2) For N^* to be unique it suffices that $R(N)$ increases with N. If $R(N)$ is indeed strictly increasing in N or, equivalently, strictly decreasing in M then the expected profits prior to entry $E\Pi(N)$ will be decreasing in N and therefore N^* will be the unique general equilibrium number of firms. We have that for a sufficiently high productivity gap between large and small firms, large firms charge a lower price which implies that:

$$\left(\frac{\gamma M}{(1-\gamma)(1-M)}\right)^{\frac{1-a}{a}}<1$$

We define:

$$F(M) \equiv \left(\frac{\gamma M}{(1-\gamma)(1-M)}\right)^{-\frac{1-a}{a}} > 1$$

$F(M)$ will also be decreasing in M (increasing in N). Using $F(M)$ we re-write $R(N)$ as a function of M,$R(M)$ and show that it is in fact decreasing in M. We have that:

$$R(M) = a\frac{z_l}{z_k}\left(\frac{\gamma}{1-\gamma}\right)^{\frac{a-1}{a}}(1-M)^{\frac{1}{a}}M^{\frac{a-1}{a}} + aM$$

To show that $R(M)$ decreases with M we need to show that the term:

$$\frac{z_l}{z_k}\left(\frac{\gamma}{1-\gamma}\right)^{\frac{a-1}{a}}(1-M)^{\frac{1}{a}}M^{\frac{a-1}{a}}$$

is decreasing in M and at a rate which is higher than 1. We show that the first term is indeed decreasing since:

$$\frac{\partial \bar{R}(M)}{\partial M} = -\frac{z_l}{z_k}F(M)\left(\frac{1-a}{a}M^{-1}+1\right) < 0$$

From the comparison of the large and the small firm markup we have that $\frac{z_l}{z_k}F(M) < 1$. Therefore, we can guarantee uniqueness for the parameter values for which the term inside the brackets is above 1 so that:

$$\frac{z_l}{z_k}F(M)\left(\frac{1-a}{a}M^{-1}+1\right) > 1$$

This is true for syfficiently low levels of a and γ.

Subgame perfection (Proposition 3) By construction we select the cost of innovation so that $f_k \in (\pi_{y1}(z_l), \pi_{y1}(z_h))$. Since the innovation decision must be made before the second period production and given that there is no borrowing in the model, the low-productivity firms will be credit constrained and therefore unable to invest. As a result all z_l-productivity firms will compete monopolistically in the second period and their efficiency will be pinned down by their initial productivity z_l. Having constrained the behaviour of inefficient firms, we go on to show under which conditions all z_h-productivity firms will choose to innovate and behave as Cournot oligopolists, resulting in the endogenous emergence of mixed market competition. The game we are considering is the following: in the first stage high-productivity firms choose whether to innovate and become Cournot oligopolists or not innovate and compete monopolistically. Given their decision, in the second stage firms choose their output simultaneously. Solving backwards, we start with the second stage and consider a generic partition of the high-productivity firms in which a percentage $g \in [0,1]$ of these firms act as Cournot oligopolists considering the effect of their choice on the market aggregates whereas a percentage $(1-g)$ of these firms choose to neglect their impact. We define ε, ϱ and η to be the expenditure, aggregate price index and total production of the differentiated good under this partition. We show that there is no SPNE where $g \in (0,1)$. Under this partition, we have that the total production of the differentiated industry equals to:

$$\eta = (Ng\gamma x^{\alpha} + N(1-g)\gamma y_h^a + N(1-\gamma)y_l^a)^{\frac{1}{a}}$$

where x is the production of z_k-productivity Cournot oligopolists, y_h is the production of z_h-productivity monopolistic competitors and y_l is the production of z_l-productivity monopolistic competitors. As a result, the inverse demand function faced by each firm can be expressed as:

$$p = q^{a-1}\eta^{1-a}\varrho$$

We assume symmetry across the $g\gamma N$ high-productivity firms that act as Cournot oligopolists and across the $(1-g)\gamma N$ high-productivity firms that act as monopolistic competitors and fixed N. The optimal decisions for each type of firms are presented below.

High-productivity MC firms The maximisation problem of z_h-productivity monopolistic competitors is the following:

$$\max_{y_h}\left\{\pi_{yh} = p_{yh}y_h - \frac{1}{z_h}y_h - f\right\}$$

$$\text{s.t. } p_{yh} = y_h^{a-1}\eta^{1-a}\varrho$$

where p_{yh} is the price of high-productivity monopolistic competitors and π_{yh} is their profit. The f.o.c. yields:

$$y_h = \left(\frac{1}{az_h}\right)^{\frac{-1}{1-a}}\varepsilon\varrho^{\frac{a}{1-a}}$$

which implies the following maximised profit for this type of firms:

$$\pi_{yh}(g\gamma N, N) = \left(\frac{1}{az_h}\right)^{\frac{a}{a-1}}(1-a)\varepsilon\varrho^{\frac{a}{1-a}} - f \tag{19}$$

Low-productivity MC firms The maximisation problem of z_l-productivity monopolistic competitors is the following:

$$\max_{y_l}\left\{\pi_{yl} = p_{yl}y_l - \frac{1}{z_l}y_l - f\right\}$$

$$\text{s.t. } p_{yl} = y_l^{a-1}\eta^{1-a}\varrho$$

where p_{yl} and π_{yl} is, respectively, the price and the profit of a low-productivity monopolistic competitor. The f.o.c. yields:

$$y_l = \left(\frac{1}{az_l}\right)^{\frac{-1}{1-a}}\varepsilon\varrho^{\frac{a}{1-a}}$$

which implies the following maximised profit for this type of firms:

$$\pi_{yl}(g\gamma N, N) = \left(\frac{1}{az_l}\right)^{\frac{a}{a-1}}(1-a)\varepsilon\varrho^{\frac{a}{1-a}} - f$$

Oligopolists The maximisation problem of z_k-productivity Cournot oligopolists is the following:

$$\max_x \left\{ \pi_x = p_x x - \frac{1}{z_k} x - f - f_k \right\}$$

$$\text{s.t. } p_x = x^{a-1}\eta^{1-a}\varrho$$

where p_x is the oligopolistic price and π_x is the oligopolistic profit. The f.o.c. yields:

$$\frac{\partial \pi_x}{\partial x} = \frac{a\sum_{j\neq i} x_j^a}{x^{1-a}\left(x^a + \sum_{j\neq i} x_j^a\right)^2}\varepsilon - \frac{1}{z_k} = 0$$

where the sum $\sum_{j\neq i} x_j^a$ contains the output of all other oligopolists and all monopolistic competitors, regardless of their productivity. Given the symmetry assumption we can supstitute $\sum_{j\neq i} x_j^a = \eta^a - x^a$ and using $\eta = \varepsilon\varrho^{-1}$ we get:

$$\varepsilon^{1-a} = \frac{1}{az_k}\varrho^{-a}(x)^{1-a} + \varepsilon^{1-2a}\varrho^a x^a$$

From the definition of the aggregate price index we get that:

$$x = \varepsilon\varrho^{-1}\left[\frac{1-S}{g\gamma N}\right]^{\frac{1}{a}}$$

where:

$$S \equiv (az_h\varrho)^{\frac{a}{1-a}} N\gamma(1-g) - (az_l\varrho)^{\frac{a}{1-a}} N(1-\gamma)$$

Using the above expression for x we can express the oligopolistic profit as follows:

$$\pi_x(g\gamma N, N) = \left[\frac{1-S}{g\gamma N} - \frac{1}{z_k}\varrho^{-1}\left(\frac{1-S}{g\gamma N}\right)^{\frac{1}{a}}\right]\varepsilon - f - f_k \tag{20}$$

Conditions for mixed market equilibrium Having computed the second-stage profit we move to the first stage where high-productivity firms simultaneously decide to innovate or not. Given a total number N of firms and $(1-\gamma)N$ of these firms having low productivity and behaving as monopolistic competitors there exists a SPNE partition $\{g\gamma N, (1-g)\gamma N\}$ in which $g\gamma N$ high-productivity Cournot oligopolists co-exist with $(1-g)\gamma N$ high-productivity monopolistic competitors iff:

1. No high-productivity monopolistic competitor has a unilateral incentive to deviate and become a Cournot oligopolist because:

$$\pi_x(g\gamma N + 1, N) - \pi_{yh}(g\gamma N, N) < 0 \tag{21}$$

2. And no Cournot oligopolist has a unilateral incentive to deviate and become a high-productivity

monopolistic competitor because

$$\pi_{yh}(g\gamma N - 1, N) - \pi_x(g\gamma N, N) < 0 \tag{22}$$

There is no SPNE in pure strategies in which some high-productivity firms incorporate the aggregate impact of their behaviour while others do not ($g \in (0,1)$). The proof follows Cellini et al. (2015) under the generalisation for non-linear functions as long as $\partial p_i(q_i, Q)/\partial q_i < 0$ and $\partial p_i(q_i, Q)/\partial Q < 0$ where Q is the total production, under Aggarwal and Samwick (1999). Consider any given partition $\{g\gamma N, (1-g)\gamma N\}$ with $g \in (0,1)$. Given the expressions for maximised profits (19) and (20), assuming that conditions (21) and (22) are simultaneously satisfied leads to a contradiction and, hence, either a high-productivity monopolistic competitor or a Cournot oligopolist has a unilateral incentive to deviate from $\{g\gamma N, (1-g)\gamma N\}$. Now consider the partition $\{g\gamma N, (1-g)\gamma N\} = \{\gamma N, 0\}$ in which all high productivity firms compete as Cournot oligopolists. We refer to this as the mixed market competition outcome, as opposed to $\{g\gamma N, (1-g)\gamma N\} = \{0, \gamma N\}$ which is the pure monopolistically competitive outcome. Unilateral deviation from mixed market competition is not profitable iff the profit of a single z_h-productivity monopolistic competitor is lower than any z_k-productivity firm's profit when all high productivity firms have chosen to act as oligopolists with productivity z_k. Formally:

$$\pi_{yh} - \pi_x < 0$$

which is equivalent to:

$$E\left[\frac{1-M(N)}{\gamma N} - \frac{1}{z_k}P^{-1}\left(\frac{1-M(N)}{\gamma N}\right)^{\frac{1}{a}} - (az_hP)^{\frac{a}{1-a}}(1-a)\right] > f_k \tag{23}$$

where P is the aggregate price index and E is the expenditure on the differentiated good. This is more likely to happen when the cost of innovation f_k is higher and the productivity increase due to the innovation (from z_h to z_k) is larger. Notice, however that (23) can hold even when $z_h = z_k$ as long as f_k, γ and a are sufficiently low.