

Holmøy, Erling; Hægeland, Torbjørn

Working Paper

Aggregate Productivity and Heterogeneous Firms

Discussion Papers, No. 266

Provided in Cooperation with:

Research Department, Statistics Norway, Oslo

Suggested Citation: Holmøy, Erling; Hægeland, Torbjørn (2000) : Aggregate Productivity and Heterogeneous Firms, Discussion Papers, No. 266, Statistics Norway, Research Department, Oslo

This Version is available at:

<https://hdl.handle.net/10419/192248>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Erling Holmøy and Torbjørn Hægeland

Aggregate Productivity and
Heterogeneous Firms

Abstract:

It is an established fact that firms, even within narrowly defined industries, differ with respect to productivity. In this paper we analyse how observed heterogeneity in productivity is affected by endogenous producer behaviour, and to what extent shifts in firm specific productivity parameters will affect aggregate industry productivity. We find that endogenous producer behaviour and equilibrium adjustments may strongly affect observed productivity of firms and aggregate industry productivity. This makes it problematic to interpret them as structural parameters. The main lesson from the paper is that identification of such parameters should rely on structural models, in which the equilibrium determinants of observable productivity for individual firms, the distribution of output shares over firms, and the number of firms are taken into account. One may otherwise draw very misleading conclusions about changes in structural parameters from observed productivity variations, either between firms or for an industry over time.

Keywords: Productivity, Heterogeneity, Aggregation

JEL classification: D24, L11

Address: Erling Holmøy (corresponding author), Statistics Norway, Research Department. E-mail: [erling.holmoy]@ssb.no

Torbjørn Hægeland, Statistics Norway, Research Department. E-mail: torbjorn.haegeland@ssb.no

Discussion Papers

comprise research papers intended for international journals or books. As a preprint a Discussion Paper can be longer and more elaborate than a standard journal article by including intermediate calculation and background material etc.

Abstracts with downloadable PDF files of
Discussion Papers are available on the Internet: <http://www.ssb.no>

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
N-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1. Introduction

Firms differ, even within narrowly defined industries. They differ with respect to size, performance and productivity. This observation would neither be sensational nor challenging if inter-firm heterogeneity was modest and temporary. However, there is strong empirical evidence that such differences are both substantial and persistent, see e.g. Sutton (1997), Baily, Hulten and Campbell (1992), Klette and Mathiassen (1995, 1996) and Klette (1999).

The evidence of intra-industry productivity differentials raises a number of questions, and we address two of them in this paper. First, to what extent is observed heterogeneity affected by endogenous producer behaviour? For example, we examine whether or not differences in exogenous productivity parameters are magnified or diminished by profit maximisation. The second question is: How, and to what extent, will shifts in firm specific productivity parameters affect aggregate industry productivity, defined as an output weighted average over the analogous observable firm productivities? Endogenous variation in firm productivity, output weights and the number of firms deteriorates the relationship between micro productivity and aggregate industry productivity.

Obviously, any analysis of these questions calls for a model framework that incorporates productivity heterogeneity of firms within the same industry. Houthakker (1955-56), Johansen (1959, 1972) and Salter (1960) represent pioneering work on the correspondence between micro and macro production functions. However, productivity heterogeneity has typically been ruled out by assumption in popular models of aggregate industry behaviour in e.g. the "new" literature of international trade and economic growth, see Helpman and Krugman (1985) and Helpman and Grossman (1992). On the other hand, heterogeneity in the form of product differentiation has been given considerable attention¹. Montagna (1995) stands out as one exception from the tradition in which aggregate industry behaviour is explained within the model of (symmetric) monopolistic competition between a large number of firms with identical technologies.

The analysis in this paper is based on a very simple structure of productivity heterogeneity, described in Section 2. Still this structure allows us to address the two questions stated above. Our ambition is that the analysis required for answering these questions also sheds light on more general issues. On the

¹ The models of *symmetric* monopolistic competition do not take the consequences of Chamberlin (1933) and Stigler (1949), who argued that product differentiation is unlikely to exist without non-uniform costs.

general level, we hope to increase the consciousness when interpreting observed productivity variations, either between firms, or for an industry over time. The main lesson from the paper is that such variations may reflect equilibrium mechanisms rather than changes in "deep" structural productivity parameters.

Mainly in order to alert the reader's understanding we present, in Section 3, a rather extreme illustration of the potential importance of equilibrium mechanisms. Here, the two questions stated above are analysed within a model where price taking firms use different technologies to produce a homogenous product. In equilibrium, firms have the same observable productivity, an implication highly inconsistent with the empirical evidence referred to above. We therefore proceed in Section 4 by analysing the same questions within a richer model, in which products from different firms are regarded as close but imperfect substitutes. Now, competition among firms is monopolistic, and endogenous producer behaviour will influence, but not *eliminate*, heterogeneity with respect to observable productivity.

We derive explicit analytical expressions for the equilibrium elasticities of aggregate productivity with respect to exogenous shifts in firm specific productivity parameters. These expressions enable us to identify how equilibrium adjustments of observable firm productivities, output weights and the number of firms contribute to the change in aggregate productivity. In the differentiated products model, we find that, if the industry is sufficiently heterogeneous with respect to productivity, the aggregate productivity response to a uniform productivity shift in all firms is approximately independent of the degree of market power of firms. Moreover, a high degree of initial productivity heterogeneity implies that a marginal change in this heterogeneity will only affect aggregate industry productivity if the firm technologies exhibit decreasing returns to scale.

Compared to the related literature, our model framework has much in common with Montagna (1995). It differs, however, by imposing more structure on the inter-firm productivity differentials, yielding a return in terms of analytical tractability. Also, our focus on aggregate productivity differs from the issues analysed by Montagna. By considering markets where both demand and supply simultaneously determine equilibrium prices and quantities, our approach also differs drastically from the putty-clay/vintage model tradition pioneered by Houthakker (1955-56) and Johansen (1972).

2. Technology assumptions

Consider an industry consisting of n active firms. The cost function of firm i is

$$(1) \quad C_i^T = C_i + F = c_i(X_i)^{1/s} + F,$$

where $i \in [0, n]$, C_i is variable costs in firm i . X_i is output, c_i is a firm-specific cost parameter, $0 < s \leq 1$ is the scale elasticity and F is a fixed cost. Our productivity measure is variable unit costs $\bar{C}_i \equiv C_i/X_i$. The marginal cost function is $C'_i(X_i) = c_i(X_i)^\lambda/s$, where $\lambda = 1/s - 1$ is the elasticity of marginal costs with respect to output. Since the scale elasticity is constant, unit costs are related to marginal costs by

$$(2) \quad \bar{C}_i(X_i) = sC'_i(X_i).$$

For analytical convenience, the set of firms is treated as a continuum. Aggregate productivity is measured by

$$(3) \quad \bar{C} = \frac{\int_0^n C_i di}{\int_0^n X_i di} = \int_0^n \left(\frac{X_i}{\int_0^n X_j dj} \right) \bar{C}_i di.$$

Whereas fixed costs are assumed to be identical for all firms, productivity differentials cause variable costs to differ between firms. Ranking firms according to productivity, so that firm 0 is the most efficient firm, we assume a constant relative productivity differential between any two adjacent firms, i.e.

$$(4) \quad \frac{dc_i}{di} = tc_i \Leftrightarrow c_i = ce^{ti}, t > 0,$$

where $(-t)$ is proportional to the relative productivity differential between firms. $c_0 = c$ is exogenous.

3. Aggregate productivity when heterogeneous firms produce a homogenous good

In this section we establish a particularly transparent case illustrating how, and to what extent, *observable* productivity at both the firm and at the aggregate level is determined by other variables than the productivity parameter characterising the micro production function, i.e. c_i above. This case highlights the problem of identifying productivity shifts at the firm level by looking at aggregate measures.

Assume that all firms are price takers and produce a homogenous good. Let P be the product price facing all firms. Firms maximise profits, defined by

$$(5) \quad \pi_i = PX_i - c_i(X_i)^{1/s}$$

with respect to X_i . Using (2), the first order condition can be written

$$(6) \quad \bar{C}_i(X_i) = \bar{C} = sP.$$

Result 1:

When different price taking firms produce a homogenous good with a technology characterised by a common constant scale elasticity s , where $0 < s < 1$, individual unit costs are identical across firms and equal sP . This result holds for any structure of the productivity heterogeneity. The observable productivity measure for the individual firm, \bar{C}_i , is independent of the productivity parameter c_i .

The intuition is straightforward: A reduction of c_i , implies *cet. par.* that output is increased to the extent that the scale effect on \bar{C}_i exactly neutralises the shift in the cost function.

However, micro productivity parameters may affect \bar{C}_i and \bar{C} through the equilibrium price P . In order to explore this possibility within a partial equilibrium framework, we assume that demand is derived from maximisation of a quasi-linear and separable utility function

$$U = \frac{A^{1/\varepsilon} D^{1-1/\varepsilon}}{1-1/\varepsilon} + y,$$

where A and ε are positive parameters, y is utility from consumption of products not produced by the industry, subject to the budget constraint $PD + y = E$. The demand for D becomes $D = AP^{-\varepsilon}$. In equilibrium, aggregate output $X = \int_0^n X_i di$ equals demand, i.e.

$$(7) \quad X = AP^{-\varepsilon}.$$

By combining (4) and (6), we see that output is decreasing exponentially in the firm's position on the productivity ranking list:

$$(8a) \quad X_i = \left(\frac{sP}{ce^{ti}} \right)^{1/\lambda} = X_0 e^{-ti/\lambda},$$

$$(8b) \quad X_0 = \left(\frac{sP}{c} \right)^{1/\lambda}.$$

Define $N = e^{-nt/(m/s-1)}$ where $m \geq 1$ is a fixed parameter, which will be interpreted as a ratio between the output price and marginal costs, i.e. a mark-up factor. In the present model with price taking firms, $m=1$, and N takes the value $N' = e^{-nt/\lambda}$. Aggregate supply, contingent on n , can then be written:

$$(9) \quad X = X_0 \frac{\lambda}{t} (1 - N') = \left(\frac{sP}{c} \right)^{1/\lambda} \frac{\lambda}{t} (1 - N').$$

The equilibrium price is found by combining (7) and (9):

$$(10) \quad P = \left[\frac{At}{\lambda} \left(\frac{c}{s} \right)^{1/\lambda} \left(\frac{1}{1 - N'} \right) \right]^{\frac{\lambda}{1 + \varepsilon \lambda}}.$$

Inserting (10) into (6) yields aggregate productivity:

$$(11) \quad \bar{C}_i = \bar{C} = sP = \left[\frac{Ats^\varepsilon c^{1/\lambda}}{\lambda} \left(\frac{1}{1-N'} \right) \right]^{\frac{\lambda}{1+\varepsilon\lambda}}.$$

A particularly simple case is the one where $\varepsilon = \infty$, so that P is exogenous, cf. Result 1. (11) is the closed form solution for aggregate productivity when the number of firms is exogenous so that $(1 - N')^{-1} > 1$ is a constant. We will consider the case where n is endogenous, determined by the standard assumption that profits equal fixed costs in the marginal firm, i.e.

$$(12) \quad \pi_n = F.$$

Since equilibrium profits are decreasing monotonically in i , the solution for n is unique. In the appendix we derive the following:

Result 2:

When firms produce a homogeneous product, the partial elasticities of \bar{C} with respect to c and t equal,

$$(13) \quad \hat{\bar{C}}_c = \frac{1}{1+\varepsilon\lambda} + \frac{(\varepsilon-1)\lambda}{(1+\varepsilon\lambda)^2} \left(\frac{N'}{1-N'} \right) \left[1 + \left(\frac{1+\lambda}{1+\varepsilon\lambda} \right) \left(\frac{N'}{1-N'} \right) \right]^{-1},$$

and

$$(14) \quad \hat{\bar{C}}_t = \frac{\lambda}{1+\varepsilon\lambda} - \frac{\lambda(1+\lambda)}{(1+\varepsilon\lambda)^2} \left(\frac{N'}{1-N'} \right) \left[1 + \left(\frac{1+\lambda}{1+\varepsilon\lambda} \right) \left(\frac{N'}{1-N'} \right) \right]^{-1}.$$

Observe that the first (positive) terms on the right hand side of (13) and (14) equal the corresponding partial elasticities in the case where n is fixed. The intuition behind these terms is analogous to familiar incidence analysis in partial equilibrium models. Consider first (13). If output, hypothetically, is held constant, raising c by one percent implies a one percent increase in \bar{C} . However, the positive shift in the individual marginal cost functions implies a negative shift in the aggregate supply curve. In the new equilibrium, P is higher and output lower compared to the initial equilibrium. All firms reduce their output, and decreasing returns to scale ($\lambda > 0$) cause a negative feedback on marginal costs in all firms. In equilibrium, the contribution to the change rate of \bar{C}_0 from the reduction of X_0 caused by a 1 percent increase in c , equals $-\varepsilon\lambda/(1 + \varepsilon\lambda)$. Adding the 1 percent direct shift in the cost function yields

the first term on the right hand side of (13). The incidence from c to P and \bar{C} is stronger the less elastic is demand, and the closer s is to unity.

The intuition behind the first term on the right hand side of (14) may be explained as follows: The increase in t raises costs in all firms except firm 0 . Their output reduction is larger the closer s is to unity, but is modified by the upward equilibrium adjustment of P . Finally, the effect on P is modified by the increased output of firm 0 . The net effect on price and unit costs is decreasing in ϵ , and increasing in s .

The second term on the right hand side of (13) and (14) captures the effects of changes in n on \bar{C} . When $\epsilon=1$, there will be no changes in n and the solution for \hat{C}_c degenerates to $\hat{C}_c \Big|_{\epsilon=1} = s$. The reason why n remains constant in this case may easily be explained from the profit expression for the marginal firm *ex ante* the change in c . Using the first order condition, profits equal $\pi_n = (1-s)PX_n$. With unitary price elasticity, the product PX_n will stay constant since the relative output adjustments are identical in all existing firms.

Inspection of (13) reveals that the equilibrium changes in n implies a positive (negative) contribution to \hat{C}_c when $\epsilon > 1$ ($\epsilon < 1$)². When $\epsilon > 1$, $\pi_n = (1-s)PX_n$ falls and firms exit. Exit of firms implies, *cet. par.*, another negative shift in the aggregate supply curve. P will rise further to clear the market, which makes it profitable for the remaining firms to expand their output along their increasing marginal cost curves. This output effect explains why \hat{C}_c is higher (lower) when $\epsilon > 1$ ($\epsilon < 1$). This result may appear somewhat paradoxical in the present context of heterogeneous firms, since the additional positive impact on \hat{C}_c is associated with exit of the *least* productive firms. As demonstrated, however, changes in productivity follows P in this model. What matters for the changes in P , is that entry or exit of firms shifts the aggregate supply function. Compared to a model of identical firms, the (model) fact that marginal firms will have a smaller output share than intra-marginal firms, reduces the effect on P caused by entry or exit.

² In (13) the term $\frac{\lambda}{(1+\epsilon\lambda)^2} \left(\frac{N'}{1-N'} \right) \left[1 + \left(\frac{1+\lambda}{1+\epsilon\lambda} \right) \left(\frac{N'}{1-N'} \right) \right]^{-1} > 0$.

Turning to the effect of changes in n on \hat{C}_t , (14) shows that this influence is negative. As shown in the appendix, the rise in t brings about a decrease in N' , and inspection of (11) shows that \bar{C} is increasing in N' . This can be seen from (9). Reducing N' implies, *cet. par.*, a positive shift in aggregate supply, which forces P , and thereby \bar{C} , down. Why does N' fall when t rises? From the definition of N' , $\hat{N}'_t < 0$ implies that $\hat{n}_t > -1$. If, hypothetically, $\hat{n}_t = -1$, N' would have been constant, and profits in the *ex post* marginal firm would balance the fixed cost for the *ex ante* equilibrium price. But P goes up for a given N' due to the contraction of supplies from all remaining firms but the most efficient one. The increase in P raises profits, and explains why $\hat{n}_t > -1$.

It follows that when an industry becomes more heterogeneous, measured by an increase in nt , N' tends to zero. Consequently, when nt is large, \bar{C} can be approximated by neglecting the factor $1/(1-N')$ in (12). The corresponding approximate elasticities are equal to the first term on the right hand side of (13) and (14). Technically, this approximation is obtained by including all firms instead of only the n most productive when calculating aggregate variables. The error resulting from including the firms with profits less than F , is small when the output share of these firms are small.

4. Productivity in an industry where heterogeneous firms produce a differentiated product

In the model with homogeneous products in the previous section, productivity heterogeneity is eliminated because firms adjust their output to the level where marginal costs equal a common output price. This result contrasts the empirical findings of large and persistent within-industry productivity differentials, which indicates that productivity heterogeneity is also an equilibrium phenomenon. This motivates a richer model specification. In this section we relax the assumption of a homogenous industry product, and consider the case where firms produce different varieties of a product. A convenient framework is the popular "large group case" model of monopolistic competition (LGMC). Below we extend this model by allowing firms to differ in productivity.

4.1. The asymmetric LGMC model

The demand for the industry product is derived from the quasi-linear utility function introduced in the previous section. However, D is now the utility from consuming a composite of the different varieties. Following the standard approach in models of monopolistic competition, the sub-utility function associated with D takes the symmetric CES-form (cf. Dixit and Stiglitz, 1977):

$$(15) \quad D = \left[\int_0^n (X_i)^{1-1/\sigma} di \right]^{\frac{1}{1-1/\sigma}},$$

where $\sigma > 1$ is the elasticity of substitution between the varieties. We assume that $\sigma > \varepsilon$. Let P_i denote the price of variety i and E the given total expenditure. Demand functions for all varieties are derived from utility maximisation subject to the budget constraint $\int_0^n P_i X_i di + y = E$. Equilibrium in each variety market requires

$$(16) \quad X_i = \left(\frac{P_i}{P} \right)^{-\sigma} A P^{-\varepsilon},$$

where P is the price index for the composite industry product. P is the minimum cost of obtaining one unit of the composite industry product. The CES form of D implies

$$(17) \quad P = \left(\int_0^n (P_i)^{1-\sigma} di \right)^{1/(1-\sigma)}.$$

Firms maximise profits, defined by (5), with respect to P_i . Although the market share will differ between firms, we stick to the standard LGMC assumption that each firm neglects the influence of its own price on the price index P . Thus, the perceived own price elasticity facing all firms equals $-\sigma$. Optimal price setting gives the familiar mark-up rule

$$(18) \quad P_i = m \frac{c_i}{s} (X_i)^\lambda,$$

where $m = \sigma/(\sigma - 1)$ is the mark-up factor. The equations (4), (5), (14), (15) and (16) determine c_i , π_i , P_i , P and X_i . As shown in detail in Holmøy and Hægeland (1997), the model can be written

$$(19a) \quad X_i = X_0 e^{\left(\frac{-\sigma}{1+\sigma\lambda} \right) i} = X_0 e^{\left(-\frac{mt}{m/s-1} \right) i},$$

$$(19b) \quad X_0 = \left[A \left(\frac{mc}{s} \right)^{-\sigma} P^{\sigma-\varepsilon} \right]^{\frac{1}{1+\sigma\lambda}}.$$

$$(20a) \quad P_i = P_0 e^{\left(\frac{t}{1+\sigma\lambda} \right) i},$$

$$(20b) \quad P_0 = \frac{mc}{s} (X_0)^\lambda,$$

$$(21) \quad P = \left\{ A^\lambda \left(\frac{mc}{s} \right) \left[\left(\frac{m/s-1}{t} \right) (1-N) \right]^{-(m/s-1)} \right\}^{\frac{1}{1+\varepsilon\lambda}},$$

The second equality in (17a) follows from $m = \sigma/(\sigma-1)$. In order to calculate $\bar{C} = C/X$ we need expressions for aggregate output and costs. Given n , we have

$$(22) \quad X = \left(\frac{m/s-1}{mt} \right) X_0 (1-N^m).$$

$$(23) \quad C = \left(\frac{m/s-1}{t} \right) c (X_0)^{1/s} (1-N).$$

Dividing (23) by (21) yields aggregate unit costs

$$(24) \quad \bar{C} = mg(nt) \bar{C}_0,$$

where $\bar{C}_0 = c(X_0)^\lambda$ is unit costs in the most efficient firm, and

$$(25) \quad g(nt) = \frac{1-N}{1-N^m} = \frac{1-e^{-nt/(m/s-1)}}{1-e^{-mnt/(m/s-1)}}.$$

4.2. Bounds for variations in aggregate productivity

So far, the solution for \bar{C} has been contingent on n . Rather than considering a market structure with a fixed number of firms, n may be determined through entry or exit according to (12). Before considering \bar{C} in the free entry case, we analyse how n affects productivity in more detail.

(24) shows that the influence of n on \bar{C} can be decomposed into changes in the multiplier $g(nt)$ and changes in \bar{C}_0 . Moreover, the ratio of \bar{C} to \bar{C}_0 equals $mg(nt)$. This ratio may be interpreted as a measure of the scope for entry and exit to generate aggregate scale diseconomies. We now show that variations in n will have only limited influence on \bar{C}/\bar{C}_0 . To see this, note that $g(nt)$ is strictly increasing and concave in nt , $\lim_{nt \rightarrow 0} g(nt) = 1/m$, $\lim_{nt \rightarrow 0} g'(nt) = \infty$, $\lim_{nt \rightarrow \infty} g(nt) = 1$, and $\lim_{nt \rightarrow \infty} g'(nt) = 0$. The latter property implies that $g(nt)$ converges fast towards unity when n increases from 0 into the range where n is large enough to make the LGMC model appropriate. (24) and the properties of the $g(nt)$ function, implies the following sharp and general conclusion about the scope for heterogeneity to cause \bar{C} to deviate from \bar{C}_0 :

Result 3:

When the industry produces a differentiated product, aggregate productivity, measured by \bar{C} , is bounded according to

$$(26) \quad \bar{C}_0 < \bar{C} < m\bar{C}_0 = \lim_{n \rightarrow \infty} \bar{C}.$$

The relative width of this interval is given by $m = \sigma/(\sigma - 1) > 0$, which is smaller the higher is the substitutability among the differentiated products.

Note that the mark-up factor itself is not a primary determinant of \bar{C} . However, under the price setting assumptions in the LGMC model, it compactly summarises the effects on \bar{C}/\bar{C}_0 caused by equilibrium adjustments of demand to price differentials between varieties. The shape of the $g(nt)$ function allows an even sharper conclusion:

Result 4:

When nt is large in the sense that even the largest firm has an insignificant market share and the LGMC model is appropriate, the variation in aggregate productivity will be close to the upper boundary of the interval, i.e. $m\bar{C}_0$.

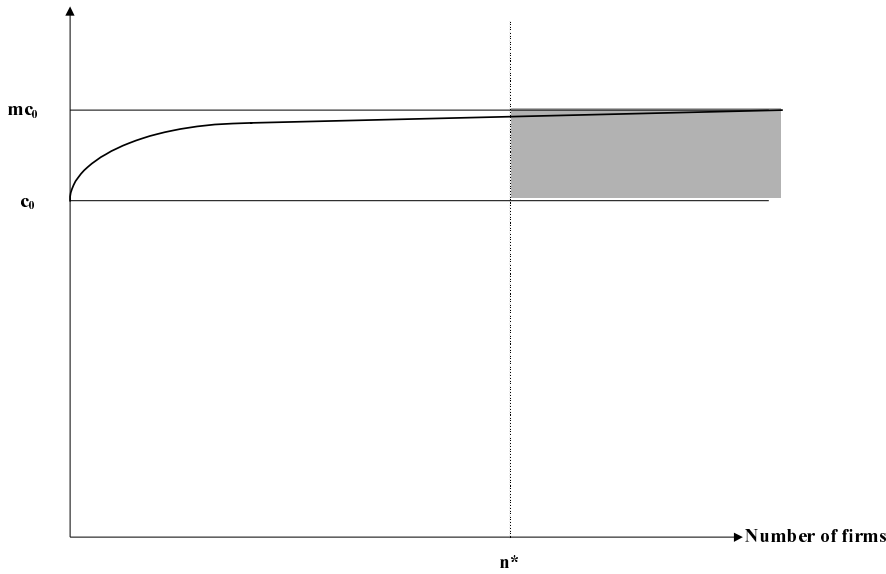
Consequently, when the LGMC model is relevant, endogenous mechanisms cause $m\bar{C}_0$ to be a good approximation of average productivity. Relative *changes* in aggregate productivity will be approximately equal to relative changes in \bar{C}_0 .

The non-trivial statement in (26) is of course the existence of an upper asymptotic bound for \bar{C} even entering firms are successively less productive. The principal reason for this result is that C_i is decreasing in i , which causes the integral defining C to converge when n approaches infinity. When n grows beyond all limits, \bar{C}/\bar{C}_0 is determined by the ratio between the growth rate of X_i and the growth rate of C_i with respect to i . The growth rate of X_i equals $-\sigma/(1+\sigma\lambda) = -mt/(m/s-1)$, and the growth rate of C_i equals $t - \sigma/[s(1 + \sigma\lambda)] = -t/(m/s-1)$. The ratio between these growth rates equals m .

With constant returns to scale at the firm level, $\bar{C}_0 = c$. In this case, not only the range of potential variation in \bar{C}/\bar{C}_0 is limited - also the range of potential variation in the *level* of \bar{C} is limited. This case is illustrated in Figure 1. Here, n^* is the equilibrium number of firms in the case when the LGMC model is relevant.

Figure 1: The relationship between aggregate productivity (unit cost) and the number of firms.

Aggregate unit cost



With decreasing returns to scale at the firm level, \bar{C}_0 depends on n through modifications in X_0 . An increase in nt raises $(1 - N)$, which reduces X_0 and \bar{C}_0 . The intuition behind this effect is associated with the valuation of variety inherent in the utility function, which shows up as a decline in P when n increases³. Such a fall in P implies two effects on X_0 . First, the demand for all varieties of the differentiated product increases. Second, demand is redirected from the pre-existing varieties to the varieties produced by the new marginal firms. Since $\sigma > \varepsilon$, the latter "internal" substitution effect dominates.

Consequently, the two effects on \bar{C} induced by increasing n pull in opposite directions; $g(nt)$ rises whereas \bar{C}_0 falls. In Figure 1 these opposing effects can be illustrated by a simultaneous downward shift of the \bar{C} -curve, and a movement to the right along the shifted \bar{C} -curve. In the appendix, we derive the premises for the following result:

Result 5:

If firm technologies exhibit constant returns to scale, a given partial increase in n raises \bar{C} .

If firm technologies exhibit decreasing returns to scale, the net effect of a given partial

increase in n on \bar{C} is theoretically ambiguous. The partial elasticity of \bar{C} with respect to n , is

$$(27) \quad \hat{\bar{C}}_n = \left[\left(\frac{mN^m}{1-N^m} \right) - d \left(\frac{N}{1-N} \right) \right] \hat{N},$$

where $\hat{N} = -nt/(m/s - 1) < 0$ is the elasticity of N with respect to nt , and $0 < d \equiv [1 + m\lambda(\varepsilon - 1)]/(1 + \varepsilon\lambda) \leq 1$.

d captures the negative scale effect on unit costs within pre-existing firms due to redistribution of demand from pre-existing to less productive entrants. From the properties of the $g(nt)$ function, it can be shown that $\hat{g}_N = mN^m/(1-N^m) - N/(1-N) < 0$. However, since $d < 1$, the sign of the bracketed term, and thereby the sign of $\hat{\bar{C}}_n$, is ambiguous.

It follows that if firm technologies exhibit constant returns to scale, an exogenous expansion of industry output through entry of firms implies aggregate decreasing returns to scale. This conclusion may, however, be reversed if there is a strong degree of decreasing returns to scale within each firm. The positive contribution to aggregate productivity from redistributing the most costly units of output from pre-existing firms to entrants, may then be large enough to dominate the negative contribution caused by the lower overall efficiency of these entrants.

4.3. The dependency of aggregate productivity on productivity at the micro level

We now consider exogenous changes in c and t , in the situation where n is determined by the entry/exit condition (12). In the appendix we derive the following result:

Result 6:

The partial elasticity of \bar{C} with respect to c , $\hat{\bar{C}}_c$, is

³ Compared to the standard symmetric LGMC model, the love of variety effect of n on P is somewhat modified in our asymmetric model. But although the index P includes higher prices as n grows, the valuation of increased variety still causes P to be negatively related to n .

$$(28) \quad \hat{\bar{C}}_c = \frac{1}{1+\varepsilon\lambda} + \left[\left(\frac{mN^m}{1-N^m} \right) - d \left(\frac{N}{1-N} \right) \right] \left(\frac{\varepsilon-1}{1+\varepsilon\lambda} \right) \left(1 + \frac{YN}{1-N} \right)^{-1},$$

$$\text{where } Y \equiv \left(\frac{\sigma-\varepsilon}{\sigma-1} \right) \left(\frac{1}{s(1+\varepsilon\lambda)} \right) > 0.$$

The first term on the right hand side of (28) is equal to $\hat{\bar{C}}_c$ when n is constant, $\hat{\bar{C}}_c \Big|_n$. Observe that $\hat{\bar{C}}_c \Big|_n$ is identical to the corresponding partial elasticity in the homogeneous good model, which was analysed in Section 3. The intuition behind this particular elasticity is also the same in the two models. In both models, and for the same reason, n is invariant with respect to c when $\varepsilon = 1$. Thus, $\hat{\bar{C}}_c = \hat{\bar{C}}_c \Big|_n$ when $\varepsilon = 1$. Even when $\varepsilon \neq 1$, $\hat{\bar{C}}_c \Big|_n$ may serve as a good approximation to the exact elasticity, when the degree of heterogeneity is strong, i.e. when nt is large so that N is small.

In the second term in (28), which accounts for the influence of equilibrium adjustments in n , we know that $[1+YN/(1-N)]^{-1} > 0$. However, the sign of the bracketed term, which equals $\hat{\bar{C}}_N$, is ambiguous as shown in the explanation of (27). In the special case where $\lambda = 0$, there is no feedback on $\bar{C}(X_0)$ from X_0 , and $d=1$. Therefore, when $\varepsilon > 1$ ($\varepsilon < 1$), the contribution to $\hat{\bar{C}}_c$ from changes in n is negative (positive). When $\varepsilon > 1$, the logic goes as follows: The rise in c reduces profits in the *ex ante* marginal firm, π_n , and cause exit of the least efficient firms. P increases when n goes down and this triggers substitution of demand within the industry in favour of the more efficient ones, which accounts for the positive productivity effect.

When $\lambda > 0$, it is not necessarily true that exit (entry) of firms has a positive (negative) impact on aggregate productivity. The reason is that the most efficient and profitable firms may produce at a scale that makes their marginal costs higher than in the less efficient firms. Thus, when output previously produced by the least efficient firms are reallocated to the most efficient ones, aggregate unit costs may rise.

We now turn to the equilibrium solution for the aggregate productivity effect induced by a change in t . In the appendix we derive:

Result 7:

The partial elasticity of \bar{C} with respect to t , $\hat{\bar{C}}_t$, is

$$(29) \quad \hat{\bar{C}}_t = \left(\frac{\sigma - \varepsilon}{\sigma - 1} \right) \left(\frac{\lambda}{1 + \varepsilon \lambda} \right) + \left[\left(\frac{mN^m}{1 - N^m} \right) - d \left(\frac{N}{1 - N} \right) \right] \left(\frac{Y}{1 + \frac{YN}{1 - N}} \right).$$

The first term in (29) accounts for the relative impact on \bar{C} caused by a 1 percent increase in t contingent on N being constant, $\hat{\bar{C}}_t|_N$. The reason why $\hat{\bar{C}}_t|_N > 0$ is that a rise in t induces an increase in the marginal costs and prices of all but the most efficient firm. The resulting increase in P implies a reduction of the demand for the differentiated product, as well as an internal substitution of demand in favour of the most efficient firm. As noted above, the latter effect dominates since $\sigma > \varepsilon$. Consequently, X_0 goes up, which raises \bar{C} when $\lambda > 0$. Again, $\hat{\bar{C}}_t|_N$ will be a better approximation to $\hat{\bar{C}}_t$ the stronger is the initial degree of heterogeneity measured by nt .

The last term in (29) accounts for the contribution to $\hat{\bar{C}}_t$ from changes in N , which are due to both the exogenous change in t as well as the endogenous adjustment of n . As pointed out above, the sign of the bracketed term, which equals $\hat{\bar{C}}_N$, was found to be indeterminate. The last parenthesis, which is positive, equals $-(1 + \hat{n}_t)$. $(1 + \hat{n}_t)$ is equal to the equilibrium change rate of nt when $\hat{t} = 1$ (percent). In the appendix, equation (A19), we show that $1 + \hat{n}_t > 0$, which implies that the equilibrium value of $c_n = ce^{tn}$ is higher after the increase in t than prior to the shift. Thus, while we can not say whether the industry equilibrium *ex post* the increase in t includes more or fewer firms compared to the *ex ante* equilibrium, it is unambiguously true that the *ex post* industry equilibrium includes active firms with lower observable productivity compared to the initial equilibrium.

In the special case when $\lambda = 0$, there is no contribution to $\hat{\bar{C}}_t$ from scale effects at the firm level. Then the first term in (29) vanishes. Moreover, $d = 1$ so $\hat{\bar{C}}_N$ in the brackets becomes negative, which confirms that $\hat{\bar{C}}_n > 0$ when $\lambda = 0$. We can therefore conclude

Result 8:

If firm technologies exhibit constant returns to scale, a given partial increase in t raises \bar{C} .

The content of Result 8 can be explained intuitively as follows: The rise in t raises unit and marginal costs in all but the most efficient firm. Computing the new value of \bar{C} with the initial output shares and the initial number of firms will produce a higher aggregate unit cost. It turns out that this positive contribution to the equilibrium value of $\hat{\bar{C}}_t$ is the dominating one. There will, however, be two kinds of modifications. First, as shown in (A22) and (A23) in the appendix, the output shares will change in favour of the most efficient firm. Second, the number of firms may go down through exit of the least efficient firms. However, since $1 + \hat{n}_t > 0$, this effect is not strong enough to generate a reduction in nt .

5. Conclusions

In this paper we have analysed the questions: i) to what extent is observed heterogeneity in productivity affected by endogenous producer behaviour? ii) How, and to what extent, do shifts in firm specific productivity parameters affect aggregate industry productivity? Within market models based on perfect and monopolistic competition, the analysis has taken into account the endogeneity of observable productivity of individual firms, the output shares of firms used to compute average industry productivity, and the number of firms. Based on analytical expressions for the equilibrium elasticities of aggregate industry productivity with respect to firm specific productivity parameters, we have derived 8 specific results that are relevant when attempting to answer the two questions stated above.

A very transparent and extreme illustration of the potential importance of equilibrium adjustments in the determination of observable productivity is obtained in the model assuming perfect competition. Here firms, using different technologies to produce a homogenous product, end up with the same *observable* productivity level as long as their elasticity of scale is common and constant. This result contrasts massive empirical evidence, and motivates richer models, where productivity heterogeneity remains as an equilibrium phenomenon. We show that the model based on monopolistic competition

may be qualitatively consistent with data, but the quantitative fitness remains to be checked. When trying to estimate "deep" structural parameters characterising the market structure and productivity heterogeneity for a specific industry, one will probably have to relax several of the simplifying assumptions made in this paper. However, such research should take into account the main lesson from this paper: Identification of such parameters should rely on structural models, which account for the equilibrium determinants of observable productivity for individual firms, the distribution of output shares over firms, and the number of firms. Our paper shows that one may otherwise draw very misleading conclusions about changes in "deep" structural parameters from observed productivity variations, either between firms or for an industry over time.

Appendix

Derivation of Result 2

Profits in the marginal firm equal

$$(A1) \quad \pi_n = (1-s)PX_n = (1-s) \left(\frac{sP}{c} \right)^{1/\lambda} N'.$$

Inserting (10) into (A1.1) yields

$$(A2) \quad \pi_n = (1-s)N' \left(\frac{s}{c} \right)^{1/\lambda} \left[\frac{(At/\lambda)(c/s)^{1/\lambda}}{1-N'} \right]^{(1+\lambda)/(1+\varepsilon\lambda)}$$

From the equilibrium condition

$$(12) \quad \pi_n = F,$$

we obtain the following implicit solution for N' :

$$(A3) \quad N'(1-N')^{-(1+\lambda)/(1+\varepsilon\lambda)} = K,$$

where

$$K \equiv \left(\frac{F}{1-s} \right) \left[\left(\frac{c}{s} \right)^{\varepsilon-1} \left(\frac{At}{\lambda} \right)^{1+\lambda} \right]^{\frac{1}{1+\varepsilon\lambda}}$$

Let $\hat{N}' = dN'/N'$ and similarly for the relative change of other variables. From the definition of N' we have

$$(A4) \quad \hat{N}' = - \left(\frac{nt}{\lambda} \right) (\hat{n} + \hat{t}).$$

Logarithmic differentiation of (A3) gives

$$(A5) \quad \hat{N}' = \frac{\hat{K}}{1 + \left(\frac{1+\lambda}{1+\varepsilon\lambda} \right) \left(\frac{N'}{1-N'} \right)}$$

Combining (A4) and (A5) yields

$$(A6) \quad \hat{n} = - \left(\frac{\lambda}{nt} \left[1 + \left(\frac{1+\lambda}{1+\varepsilon\lambda} \right) \left(\frac{N'}{1-N'} \right) \right] \right)^{-1} \hat{K} - \hat{t},$$

where

$$(A7) \quad \hat{K} = \left(\frac{\varepsilon-1}{1+\varepsilon\lambda} \right) \hat{c} - \left(\frac{1+\lambda}{1+\varepsilon\lambda} \right) \hat{t}$$

when all variables but c and t are constant. Logarithmic differentiation of (11) with respect to c , t and n yields

$$(A8) \quad \frac{\hat{c}}{\hat{C}} = \left(\frac{\lambda}{1+\varepsilon\lambda} \left[\frac{\hat{c}}{\lambda} + \hat{t} + \left(\frac{N'}{1-N'} \right) \hat{N}' \right] \right)$$

Inserting (A4), (A6) and finally (A7) yields

$$(A9) \quad \frac{\hat{c}}{\hat{C}} = \left(\frac{\lambda}{1+\varepsilon\lambda} \left[\frac{\hat{c}}{\lambda} + \hat{t} + \left(\frac{N'}{1-N'} \right) \left[1 + \left(\frac{1+\lambda}{1+\varepsilon\lambda} \right) \left(\frac{N'}{1-N'} \right) \right]^{-1} \hat{K} \right] \right)$$

Substituting the expression for \hat{K} implies

$$(A10) \quad \frac{\hat{c}}{\hat{C}_c} = \frac{1}{1+\varepsilon\lambda} + \frac{(\varepsilon-1)\lambda}{(1+\varepsilon\lambda)^2} \left(\frac{N'}{1-N'} \right) \left[1 + \left(\frac{1+\lambda}{1+\varepsilon\lambda} \right) \left(\frac{N'}{1-N'} \right) \right]^{-1},$$

$$(A11) \quad \hat{\bar{C}}_t = \frac{\lambda}{1+\varepsilon\lambda} - \frac{\lambda(1+\lambda)}{(1+\varepsilon\lambda)^2} \left(\frac{N'}{1-N'} \right) \left[1 + \left(\frac{1+\lambda}{1+\varepsilon\lambda} \right) \left(\frac{N'}{1-N'} \right) \right]^{-1},$$

where $\hat{\bar{C}}_c$ and $\hat{\bar{C}}_t$ are the partial elasticities of \bar{C} with respect to c and t respectively. In the special case where $\varepsilon = 1$, the elasticities simplifies to

$$\hat{\bar{C}}_c \Big|_{\varepsilon=1} = \frac{1}{1+\lambda} = s,$$

$$\hat{\bar{C}}_t \Big|_{\varepsilon=1} = \left(\frac{\lambda}{1+\lambda} \right) \left[1 - \frac{\frac{N'}{1-N'}}{1 + \frac{N'}{1-N'}} \right] = \left(\frac{\lambda}{1+\varepsilon\lambda} \right) (1-N').$$

Derivation of Result 5

By deriving the analytical solution for the partial elasticity of \bar{C} with respect to n , we can verify that the net effect of a partial increase in n on \bar{C} is theoretically ambiguous. To this end, we need the elasticities of $g(nt)$ with respect to n and t ,

$$(A12) \quad \hat{g}_n = \frac{\partial g(nt)}{\partial n} \frac{n}{g(nt)} = \hat{g}_t = \left(\frac{mN^m}{1-N^m} - \frac{N}{1-N} \right) \hat{N} > 0.$$

The positive sign of $\hat{g}_n = \hat{g}_t$ follows from the fact that $\hat{N} < 0$ and $mN^m/(1-N^m) - N/(1-N) < 0$. The sign of the latter term follows from the properties of the function $g(nt)$, cf. Holmøy (1999). Utilising (19b), (21) and (25), logarithmic differentiation of (24) with respect to n implies

$$(A13) \quad \hat{\bar{C}}_n = \left[\left(\frac{mN^m}{1-N^m} \right) - d \left(\frac{N}{1-N} \right) \right] \hat{N},$$

The bracketed term in (A13) equals the elasticity of \bar{C} with respect to N , i.e. $\hat{\bar{C}}_N$. Since $mN^m/(1-N^m) - N/(1-N) < 0$, $\hat{\bar{C}}_N < 0$ if d were equal to unity. However, since $d \leq 1$, the sign of $\hat{\bar{C}}_N$ is ambiguous.

Derivation of Result 6

Logarithmic differentiation of (24) implies

$$(A14) \quad \hat{C} = \hat{g}_n(\hat{n}_c \hat{c} + \hat{n}_t \hat{t}) + \hat{g}_t \hat{t} + \hat{c} + \lambda \left[(\hat{X}_{0c} + \hat{X}_{0n} \hat{n}_c) \hat{c} + (\hat{X}_{0t} + \hat{X}_{0n} \hat{n}_t) \hat{t} \right].$$

The partial logarithmic derivatives of X_0 are given by

$$(A15) \quad \hat{X}_{0c} = \frac{-\varepsilon}{1 + \lambda \varepsilon} < 0,$$

$$(A16) \quad \hat{X}_{0n} = \frac{(\sigma - \varepsilon)(m/s - 1)}{(1 + \sigma \lambda)(1 + \varepsilon \lambda)} \left(\frac{N}{1 - N} \right) \hat{N} = sY \left(\frac{N}{1 - N} \right) \hat{N} < 0,$$

$$(A17) \quad \hat{X}_{0t} = \frac{(\sigma - \varepsilon)(m/s - 1)}{(1 + \sigma \lambda)(1 + \varepsilon \lambda)} \left(1 + \frac{N\hat{N}}{1 - N} \right) = sY \left(1 + \frac{N\hat{N}}{1 - N} \right) = sY + \hat{X}_{0n} > 0,$$

where we have defined $Y \equiv \left(\frac{\sigma - \varepsilon}{\sigma - 1} \right) \left(\frac{1}{s(1 + \varepsilon \lambda)} \right) > 0$, which is the elasticity of C_0 with respect to t when the variation in N is neglected. That $\hat{X}_{0t} > 0$ follows from $1 + N\hat{N}/(1 - N) = \left[(\hat{N} - 1)N + 1 \right] / (1 - N) > 0$. To prove the latter result, consider the numerator $\varphi(\hat{N}) = (\hat{N} - 1)N + 1 = (\hat{N} - 1)e^{\hat{N}} + 1$ defined for $\hat{N} < 0$. It is straightforward to verify that $\varphi'(\hat{N}) < 0$, that $\lim_{\hat{N} \rightarrow -\infty} \varphi(\hat{N}) = 1$ and that $\varphi(0) = 0$. Thus, $0 < \varphi(\hat{N}) < 1$, which implies that $1 + N\hat{N}/(1 - N) = \varphi(\hat{N}) / (1 - e^{\hat{N}}) > 0$.

In order to derive \hat{n}_c and \hat{n}_t , we use the entry/exit equilibrium condition $\pi_n(c, t, n) = F$. Logarithmic differentiation implies

$$\hat{n}_c = -\frac{\hat{\pi}_{nc}}{\hat{\pi}_{nn}}, \quad \hat{n}_t = -\frac{\hat{\pi}_{nt}}{\hat{\pi}_{nn}}.$$

Rewriting the profit function as $\pi_n(c, t, n) = \left(\frac{m}{s} - 1 \right) c X_0^{1/s} N$, recognising (A15), (A16) and (A17) as well as the definition of N , the partial elasticities of $\pi_n(c, t, n)$ becomes

$$\hat{\pi}_{nc} = \frac{1-\varepsilon}{1+\varepsilon\lambda}$$

$$\hat{\pi}_{nn} = \frac{1}{s}\hat{X}_{0n} + \hat{N}_n = \frac{1}{s}\left(\frac{\sigma-\varepsilon}{1+\sigma\lambda}\right)\hat{P}_n + \hat{N} = \left(1 + \frac{YN}{1-N}\right)\hat{N} < 0$$

$$\hat{\pi}_{ni} = \frac{1}{s}\hat{X}_{0i} + \hat{N}_i = \frac{\sigma-\varepsilon}{s(1+\sigma\lambda)}\hat{P}_i + \hat{N} = Y + \left(1 + \frac{YN}{1-N}\right)\hat{N}$$

We then obtain

$$(A18) \quad \hat{n}_c = -\frac{\hat{\pi}_{nc}}{\hat{\pi}_{nn}} = \frac{1-\varepsilon}{\left(1 + \frac{YN}{1-N}\right)(1+\varepsilon\lambda)\hat{N}}$$

$$(A19) \quad \hat{n}_i = -\frac{\hat{\pi}_{ni}}{\hat{\pi}_{nn}} = -\frac{Y + \left(1 + \frac{YN}{1-N}\right)\hat{N}}{\left(1 + \frac{YN}{1-N}\right)\hat{N}} = -1 - \frac{Y}{\left(1 + \frac{YN}{1-N}\right)\hat{N}} > -1.$$

Inserting (A12) and (A15) - (A19) into (A14) yields the following solution for $\hat{\bar{C}}_c$:

$$\begin{aligned} (A20) \quad \hat{\bar{C}}_c &= 1 + \hat{g}_n \hat{n}_c + \lambda(\hat{X}_{0c} + \hat{X}_{0n} \hat{n}_c) \\ &= 1 - \frac{\lambda\varepsilon}{1+\lambda\varepsilon} + \left[\left(\frac{mN^m}{1-N^m} - \frac{N}{1-N} \right) + \lambda s Y \left(\frac{N}{1-N} \right) \right] \hat{N} \hat{n}_c \\ &= \frac{1}{1+\lambda\varepsilon} + \left[\left(\frac{mN^m}{1-N^m} - \frac{N}{1-N} \right) + (1-s) \left(\frac{YN}{1-N} \right) \right] \left(\frac{\varepsilon-1}{1+\lambda\varepsilon} \right) \left(1 + \frac{YN}{1-N} \right)^{-1} \\ &= \frac{1}{1+\varepsilon\lambda} + \left[\left(\frac{mN^m}{1-N^m} \right) - d \left(\frac{N}{1-N} \right) \right] \left(\frac{\varepsilon-1}{1+\varepsilon\lambda} \right) \left(1 + \frac{YN}{1-N} \right)^{-1}, \end{aligned}$$

where the last expression was obtained by using the definition of $0 < d \equiv [1 + m\lambda(\varepsilon - 1)] / (1 + \varepsilon\lambda) \leq 1$ introduced above.

Derivation of Result 7

Recall that $\hat{g}_n = \hat{g}_t$, $\hat{X}_{0t} = sY + \hat{X}_{0n}$, and that (A19) implies that $1 + \hat{n}_t = -Y \left[\left(1 + \frac{YN}{1-N} \right) \hat{N} \right]^{-1} > 0$.

The solution for \hat{C}_t can then be written

$$\begin{aligned}
 \text{(A21) } \hat{C}_t &= \hat{g}_n \hat{n}_t + \hat{g}_t + \lambda (\hat{X}_{0t} + \hat{X}_{0n} \hat{n}_t) \\
 &= \lambda sY + (\hat{g}_n + \lambda \hat{X}_{0n}) (1 + \hat{n}_t) \\
 &= (1-s)Y - \left[\left(\frac{mN^m}{1-N^m} - \frac{N}{1-N} \right) \hat{N} + \lambda sY \left(\frac{N}{1-N} \right) \hat{N} \right] \left[\frac{Y}{\hat{N} \left(1 + \frac{YN}{1-N} \right)} \right] \\
 &= (1-s)Y - \left[\left(\frac{mN^m}{1-N^m} \right) - d \left(\frac{N}{1-N} \right) \right] \left[\frac{Y}{1 + \frac{YN}{1-N}} \right],
 \end{aligned}$$

$$\text{where } (1-s)Y = \left(\frac{\lambda}{1+\varepsilon\lambda} \right) \left(\frac{\sigma - \varepsilon}{\sigma - 1} \right).$$

Changes in outputs

Cost increments are transmitted to output prices and to the price index P . The increase in P reduces demand for the differentiated industry product. At the same time, the increased cost and price heterogeneity also induces substitution of demand within the industry in favour of the most efficient firm. For the most efficient firm the latter substitution effect dominates, so the output from the most efficient firm will increase. On the other hand, output from the *ex post* marginal firm goes down. The equilibrium relative change will be

$$\begin{aligned}
\text{(A22)} \quad \hat{X}_{0t} &= \left(\frac{m/s - 1}{1 + \varepsilon\lambda} \right) \left(\frac{\sigma - \varepsilon}{1 + \sigma\lambda} \right) \left(1 + \frac{N\hat{N}_t}{1 - N} \right) \\
&= sY + \left(\frac{sYN}{1 - N} \right) \hat{N}(-1) \left(\frac{Y}{(1 + YN/(1 - N))\hat{N}} \right) = sY \left[1 - \frac{YN/(1 - N)}{1 + YN/(1 - N)} \right] = \frac{sY}{1 + YN/(1 - N)} > 0
\end{aligned}$$

$$\begin{aligned}
\text{(A23)} \quad \hat{X}_m &= \hat{X}_{0t} + m\hat{N}_t \\
&= sY + \left(\frac{sYN}{1 - N} + m \right) \hat{N}(-1) \left(\frac{Y}{\left(1 + \frac{YN}{1 - N} \right) \hat{N}} \right) = sY \left[1 - \frac{m/s + YN/(1 - N)}{1 + YN/(1 - N)} \right] < 0
\end{aligned}$$

Recall, however, that output changes at the firm level have no cost effect when $\lambda=0$. On the other hand, when $\lambda > 0$, so that $0 < d < 1$, the redirection of output within the industry implies a positive contribution to \hat{C}_t from scale effects at the individual firm level. In this case changes in N may have a negative impact on \hat{C}_t . However, when $\lambda > 0$ the term in (A21) becomes positive, and will dominate effects caused by changes in N as long as the market share of the least efficient firm is sufficiently small.

References

- Baily, N. M., C. Hulten and D. Campbell (1992): Productivity Dynamics in Manufacturing Plants, *Brookings Papers on Economic Activity: Microeconomics*, 187-268.
- Chamberlin, E. H. (1933): *The Theory of Monopolistic Competition*, Cambridge: Harvard University Press.
- Dixit, A. and J. E. Stiglitz (1977): Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, **67**, 297-308.
- Helpman, E. and G. M. Grossman (1992): *Innovation and Growth in the Global Economy*, Cambridge: MIT Press.
- Helpman, E. and P. Krugman (1985): *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition and the International Economy*, Cambridge: MIT Press.
- Holmøy, E. (1999): Aggregate Industry Behaviour in a Model of Asymmetric Monopolistic Competition, unpublished manuscript, Statistics Norway.
- Holmøy, E. and T. Hægeland (1997): Aggregate Productivity Effects of Technology Shocks in a Model of Heterogeneous Firms: The Importance of Equilibrium Adjustments, Discussion Papers 198, Statistics Norway.
- Houthakker, H. (1955/56): The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis, *Review of Economic Studies* **23**, 27-31.
- Johansen, L. (1959): Substitution versus fixed coefficients in the theory of economic growth, *Econometrica* **27**, 157-176.
- Johansen, L. (1972): *Production Functions*, Amsterdam: North-Holland.
- Klette, T. J. (1999): Market Power, Scale Economies and Productivity: Estimates from a Panel of Establishment Data, *Journal of Industrial Economics*, **47**, 451-476.

Klette, T. J. and A. Mathiassen (1995): Job Creation, Job Destruction and Plant Turnover in Norwegian Manufacturing, *Annales d'Economie et de Statistique*, **41/42**, 97-125.

Klette, T.J. and A. Mathiassen (1996): *Growth and Turnover among Norwegian Manufacturing Plants*. Social and Economic Studies 95, Statistics Norway. (In Norwegian).

Montagna, C. (1995): Monopolistic competition with Firm-Specific Costs, *Oxford Economic Papers* **47**, 318-328.

Salter, W. (1960): *Productivity and Technical Change*, London: Cambridge University Press.

Stigler, G. J. (1949): Monopolistic Competition in Retrospect, in *Five Lectures on Economic Problems*, London: London School of Economics.

Sutton, J. (1997): Gibrat's Legacy, *Journal of Economic Literature* **XXXV**, 40-59.

Recent publications in the series Discussion Papers

- | | | | |
|-----|---|-----|---|
| 172 | J.K. Dagsvik (1996): Dynamic Choice, Multistate Duration Models and Stochastic Structure | 196 | R. Aaberge and A. Langørgen (1997): Fiscal and Spending Behavior of Local Governments: An Empirical Analysis Based on Norwegian Data |
| 173 | J.K. Dagsvik (1996): Aggregation in Matching Markets | 197 | A.C. Hansen and H.K. Selte (1997): Air Pollution and Sick-leaves - is there a Connection? A Case Study using Air Pollution Data from Oslo |
| 174 | H.C. Bjørnland (1996): The Dynamic Effects of Aggregate Demand, Supply and Oil Price Shocks | 198 | E. Holmøy and T. Hægeland (1997): Aggregate Productivity Effects of Technology Shocks in a Model of Heterogeneous Firms: The Importance of Equilibrium Adjustments |
| 175 | A. Bruvoll and K. Ibenholt (1996): Future Waste Generation. Forecasts Based on a Macroeconomic Model | 199 | E. Berg, P. Boug and S. Kverndokk (1997): Norwegian Gas Sales and the Impacts on European CO ₂ Emissions |
| 176 | T. Fæhn and L. A. Grünfeld (1996) Recent Leaps Towards Free Trade. The Impact on Norwegian Industry and Trade Patterns | 200 | H.C. Bjørnland (1997): Estimating Core Inflation – The Role of Oil Price Shocks and Imported Inflation |
| 177 | R. Barrell and K. A. Magnussen (1996): Counterfactual Analyses of Oil price Shocks using a World Model | 201 | R. Aaberge, A. Bjørklund, M. Jäntti, P.J. Pedersen, N. Smith and T. Wennemo (1997): Unemployment Shocks and Income Distribution. How Did the Nordic Countries Fare During their Crises? |
| 178 | E. Bowitz and S. I. Hove (1996): Business cycles and fiscal policy: Norway 1973-93 | 202 | L. Brubakk (1997): Estimation of Price Elasticities from Norwegian Household Survey Data |
| 179 | H.C. Bjørnland (1996): Sources of Business Cycles in Energy Producing Economies: The case of Norway and United Kingdom | 203 | J. Aasness and L. Belsby (1997): Estimation of Time Series of Latent Variables in an Accounting System: Petrol Consumption of Norwegian Households 1973-1995 |
| 180 | K. Nyborg (1996): The Political Man and Contingent Valuation: Motives Do Count | 204 | A. Rygh Swensen (1997): Change in Regime and Markov Models |
| 181 | E. Berg, S. Kverndokk and K.E. Rosendahl (1996): Gains from Cartelisation in the Oil Market | 205 | K. Nyborg and I. Spangen (1997): Cost-Benefit Analysis and the Democratic Ideal |
| 182 | R. Aaberge and I. Aslaksen (1996): Decomposition of the Gini Coefficient by Income Components: Various Types of Applications and Interpretations | 206 | L. Belsby and J.F. Bjørnstad (1997): Modelling and Estimation Methods for Household Size in the Presence of Nonresponse: Applied to The Norwegian Consumer Expenditure Survey |
| 183 | B. Bye (1996): Taxation, Unemployment and Growth: Dynamic Welfare Effects of "Green" Policies | 207 | K.O. Aarbu and T.O. Thoresen (1997): The Norwegian Tax Reform; Distributional Effects and the High-Income Response |
| 184 | T.J. Klette and F. Johansen (1996): Accumulation of R&D Capital and Dynamic Firm Performance: A Not-so-fixed Effect Model | 208 | T. Hægeland and T.J. Klette (1997): Do Higher Wages Reflect Higher Productivity? Education, Gender and Experience Premiums in a Matched Plant-Worker Data Set |
| 185 | B. Bye (1996): Environmental Tax Reform and Producer Foresight: An Intertemporal Computable General Equilibrium Analysis | 209 | J. Gjerde, S. Grepperud and S. Kverndokk (1998): Optimate Climate Policy under the Possibility of a Catastrophe |
| 186 | S. Grepperud (1997): Soil Depletion Choices under Production and Price Uncertainty | 210 | T. Eika and K.A. Magnussen (1998): Did Norway Gain from the 1979-85 Oil Price Shock? |
| 187 | N.M. Stølen and T. Åvitsland (1997): Has Growth in Supply of Educated Persons Been Important for the Composition of Employment? | 211 | K.O. Aarbu and J.K. MacKie-Mason (1998): Why Some Corporations Pay More Tax than Necessary |
| 188 | T.J. Klette and Z. Griliches (1997): Empirical Patterns of Firm Growth and R&D Investment: A Quality Ladder Model Interpretation | 212 | R. Aaberge (1998): UMP Unbiased Tests for Multiparameter Testing Problems with Restricted Alternatives |
| 189 | J. Aune, S. Glomsrød, V. Iversen and H. Wiig (1997): Structural Adjustment and Soil Degradation in Tanzania. A CGE-model Approach with Endogenous Soil Productivity | 213 | M. Søberg (1998): "EPA's New Emissions Trading Mechanism: A Laboratory Evaluation" – A Comment |
| 190 | E. Biørn and T.J. Klette (1997): Panel Data with Errors-in-Variables: A Note on Essential and Redundant Orthogonality Conditions in GMM-estimation | 214 | K. Nyborg (1998): Non-Verifiable Emissions, Voluntary Agreements, and Emission Taxes |
| 191 | L. Belsby and B.K. Wold (1997): Primary Schooling in Zambia Squeezed at Community and Household Level | 215 | H. C. Bjørnland (1998): Economic Fluctuations in a Small Open Economy - Real versus Nominal Stocks |
| 192 | E. Bowitz and Å. Cappelen (1997): Incomes Policies and the Norwegian Economy 1973-93 | 216 | L.C. Zhang (1998): Post-Stratification and Calibration – A Synthesis |
| 193 | S. Glomsrød, M.D. Monge A. and H. Vennemo (1997): Structural Adjustment and Deforestation in Nicaragua | 217 | R. Aaberge and Y. Zhu: The Pattern of Household Savings during a Hyperinflation. The Case of Urban China in the Late 1980s |
| 194 | F. Johansen and T.J. Klette (1997): Wage and Employment Effects of Payroll Taxes and Investment Subsidies | | |
| 195 | T. Fæhn (1997): Non-Tariff Barriers - the Achilles' Heel of Trade Policy Analysis | | |

- 218 L.C. Zhang (1998): A Note on Post-stratification when Analyzing Binary Survey Data Subject to Non-response
- 219 T.J. Klette and J. Møen (1998): From Growth Theory to Technology Policy – Coordination Problems in Theory and Practice
- 220 L.C. Zhang (1998): Synthetic Graphical Modelling of Binary Panel Data Subject to Non-response. A Case Study
- 221 J.K. Dagsvik (1998): Choice among Lotteries when Preferences are Stochastic
- 222 J.K. Dagsvik (1998): Nonparametric Identification of Discrete Choice Models
- 223 L. Brubakk and J.K. Dagsvik (1998): Consumer Demand and Unobservable Product Attributes
- 224 E. Holmøy (1998): Potential Aggregate Gains from Improved Sectoral Efficiency: General Equilibrium Estimates for Norway
- 225 B.H. Vatne and J.K. Dagsvik (1998): Estimation of Generalized Extreme Value Models by a Max-spectral Representation
- 226 I. Svendsen (1998): Rational Expectations in Price Setting – Tests Based on Norwegian Export Prices
- 227 K.A. Brekke and R.B. Howarth (1998): The Social Contingency of Wants: Implications for Growth and the Environment
- 228 K.A. Brekke and E. Moxnes (1998): Do Models Improve Fishery Management? Empirical Evidence from a Experimental Study
- 229 J.K. Dagsvik, Yu Zhu and R. Aaberge (1998): A Framework for Empirical Modelling of Consumer Demand with Latent Quality Attributes
- 230 R. Aaberge, U. Colombino and S. Strøm (1998): Social Evaluation of Individual Welfare Effects from Income Taxation: Empirical Evidence Based on Italian Data for Married Couples
- 231 R. Nesbakken (1998): Residential Energy Consumption for Space Heating in Norwegian Households. A Discrete-Continuous Choice Approach
- 232 R. Nesbakken (1998): Price Sensitivity of Residential Energy Consumption in Norway
- 233 M. Sjøberg (1998): Uncertainty and International Negotiations on Tradable Quota Treaties
- 234 J.K. Dagsvik and L. Brubakk: Price Indexes for Elementary Aggregates Derived from Behavioral Assumptions
- 235 E. Biørn, K.-G. Lindquist and T. Skjerpen (1998): Random Coefficients and Unbalanced Panels: An Application on Data from Norwegian Chemical Plants
- 236 K. Ibenholt (1998): Material Accounting in a Macroeconomic Framework. Forecast of waste generated in manufacturing industries in Norway
- 237 K.-G. Lindquist (1998): The Response by the Norwegian Aluminium Industry to Changing Market Structure
- 238 J.K. Dagsvik, A.S. Flaatten and H. Brunborg: A Behavioral Two-Sex Model
- 239 K.A. Brekke, R.B. Howarth and K. Nyborg (1998): Are there Social Limits to Growth?
- 240 R. B. Howarth and K. A. Brekke (1998): Status Preferences and Economic Growth
- 241 H. Medin, K. Nyborg and I. Bateman (1998): The Assumption of Equal Marginal Utility of Income: How Much Does it Matter?
- 242 B. Bye (1998): Labour Market Rigidities and Environmental Tax Reforms: Welfare Effects of Different Regimes
- 243 B.E. Naug (1999): Modelling the Demand for Imports and Domestic Output
- 244 J. Sexton and A. R. Swensen (1999): ECM-algorithms that converge at the rate of EM
- 245 E. Berg, S. Kverndokk and K.E. Rosendahl (1999): Optimal Oil Exploration under Climate Treaties
- 246 J.K. Dagsvik and B.H. Vatne (1999): Is the Distribution of Income Compatible with a Stable Distribution?
- 247 R. Johansen and J.K. Dagsvik (1999): The Dynamics of a Behavioral Two-Sex Demographic Model
- 248 M. Sjøberg (1999): Asymmetric information and international tradable quota treaties. An experimental evaluation
- 249 S. Grepperud, H. Wiig and F.A. Aune (1999): Maize Trade Liberalization vs. Fertilizer Subsidies in Tanzania: A CGE Model Analysis with Endogenous Soil Fertility
- 250 K.A. Brekke and Nils Chr. Stenseth (1999): A Bio-Economic Approach to the study of Pastoralism, Famine and Cycles. Changes in ecological dynamics resulting from changes in socio-political factors
- 251 T. Fæhn and E. Holmøy (1999): Welfare Effects of Trade Liberalisation in Distorted Economies. A Dynamic General Equilibrium Assessment for Norway
- 252 R. Aaberge (1999): Sampling Errors and Cross-Country Comparisons of Income Inequality
- 253 I. Svendsen (1999): Female labour participation rates in Norway – trends and cycles
- 254 A. Langørgen and R. Aaberge: A Structural Approach for Measuring Fiscal Disparities
- 255 B. Halvorsen and B.M. Larsen (1999): Changes in the Pattern of Household Electricity Demand over Time
- 256 P. Boug (1999): The Demand for Labour and the Lucas Critique. Evidence from Norwegian Manufacturing
- 257 M. Rege (1999): Social Norms and Private Provision of Public Goods: Endogenous Peer Groups
- 258 L. Lindholt (1999): Beyond Kyoto: CO₂ permit prices and the markets for fossil fuels
- 259 R. Bjørnstad and R. Nymoen (1999): Wage and Profitability: Norwegian Manufacturing 1967-1998
- 260 T.O. Thoresen and K.O. Aarbu (1999): Income Responses to Tax Changes – Evidence from the Norwegian Tax Reform
- 261 B. Bye and K. Nyborg (1999): The Welfare Effects of Carbon Policies: Grandfathered Quotas versus Differentiated Taxes
- 262 T. Kornstad and T.O. Thoresen (1999): Means-testing the Child Benefit
- 263 M. Rønsen and M. Sundström (1999): Public Policies and the Employment Dynamics among new Mothers – A Comparison of Finland, Norway and Sweden
- 264 J.K. Dagsvik: Multinomial Choice and Selectivity
- 265 Y. Li: Modeling the Choice of Working when the Set of Job Opportunities is Latent
- 266 E. Holmøy and T. Hægeland: Aggregate Productivity and Heterogeneous Firms