ESSAYS IN INDUSTRIAL ECONOMICS

by

Marcus Asplund

A Dissertation for the Doctor's Degree in Economics

Department of Economics
Stockholm School of Economics
1996

STOCKHOLM SCHOOL OF ECONOMICS
THE ECONOMIC RESEARCH INSTITUTE

ISBN 91-7258-408-4

Stockholm 1995

ESSAYS IN INDUSTRIAL ECONOMICS

Marcus Asplund

STOCKHOLM SCHOOL
OF ECONOMICS

AKADEMISK AVHANDLING

som för avläggande av ekonomie doktorsexamen
vid Handelshögskolan i Stockholm
framlägges till offentlig granskning
fredagen den 12 januari 1996
kl 13.15 i sal 550
Handelshögskolan, Sveavägen 65

STOCKHOLM 1995

Cambridge, MA, has been of great value in the dissertation work, and will undoubtedly affect my future research too.

Financial support from the Swedish Competition Authority, Tom Hedelius' and Jan Wallander's Foundation for Economic Research, Sweden American Foundation, Borgrättsfonderna, Royal Swedish Academy of Science, and Bröderna af Jochnicks Donationsstipendium, are gratefully acknowledged. In conjunction with this I wish to thank Rune Castenäs for helping me with the practical side of economics, which is not taught in the regular courses in the programme.

Finally, I wish to express my warmest thanks to Ylva, for comfort at times when equations looked like a jungle and regression models refused to co-operate, and invaluable support at all times.

Stockholm, December 1995

Marcus Asplund

To the Reader

This dissertation consists of five separate and self-contained chapters. The notation in the chapters is not necessarily the same, and there is some overlap between the chapters on driving schools.

Contents

Chapter 3

What Fraction of a Capital Investment is Sunk Cost?

Chapter 4

Competition in Interrelated Markets: An Empirical Study

Chapter 5

Estimating the Number of Firms and Capacity in Small Markets

Oligopoly and Risk Aversion*

Marcus Asplund

Department of Economics

Stockholm School of Economics

Box 6501, S-113 83 Stockholm

e-mail: nema@hhs.se

Abstract

In this paper we consider how the degree of risk aversion, and demand/cost uncertainty, influence competition in oligopolistic markets. Under demand uncertainty, the best response strategies (both quantities and prices) are decreasing in the Arrow-Pratt measure of absolute risk aversion, but for cost uncertainty, quantities are decreasing while prices are increasing. If firms are risk averse, past profits and fixed costs are important because they affect net wealth. The paper also studies multi-stage games, where firms take into account that today's profit will determine the intensity of tomorrow's competition. Finally, we characterise the equilibrium number of risk averse firms in the market. Under demand uncertainty and price competition, concentration is never lower but prices may be higher or lower in equilibrium, compared to a risk neutral market, whereas with quantity competition the effects on both prices and equilibrium number of firms are ambiguous.

Keywords: Oligopoly; risk aversion; two stage game; free entry equilibrium.

JEL numbers: D43; D81; L13; L21.

* I wish to thank Tore Ellingsen, Torbjörn Becker, Lars Bergman, Richard Caves, Thomas Lindh, Rickard Sandin, participants at the Industrial organization seminar at Harvard University, and at the workshops in economic theory and financial economics at Stockholm School of Economics, for helpful comments and suggestions. Financial support from the Swedish Competition Authority is gratefully acknowledged.

1. Introduction

A standard assumption in oligopoly theory is that firms are risk neutral. This implies that under uncertainty firms are maximising expected profits, without any concern for risk. Arguments based on markets for corporate control, managerial labour markets, and survival in product market competition, have all been used to support this assumption. However, each of these have been challenged on both theoretical and empirical grounds.[1] Several arguments have also been used to defend the standpoint that firms act *as if* they were risk averse. Non diversified owners, separation of ownership and control, costly financial distress, and non linear tax systems are some which are frequently invoked.[2] Casual observation of firms reveals some behaviours which are not easily reconciled with risk neutrality, e.g. use of insurance, and hedging of firm specific risks. In spite of this, surprisingly little work have focused on the effects of risk aversion on competition.

Altering the assumption of risk neutrality gives several implications for the product market competition. Early work was mostly related to perfect competition or monopoly. Pioneering work by Baron (1970), and Sandmo (1971) shows how increased uncertainty about price lowers the quantity produced in perfectly competitive markets. One interesting conclusion is that fixed costs matters in the output decisions, and cause output to decrease in the presence of decreasing absolute risk aversion. In a monopoly framework Baron (1971), and Leland (1972) derive results similar to those obtained for competitive markets. Analysis of risk averse oligopolists are more recent. Holthausen (1979) shows that less risk averse firms are more likely to become price leaders, since they are more willing to gamble on price changes. Information sharing among risk averse oligopolists is studied in Hviid (1989), who find that it can be optimal for Cournot competitors to reveal private information about uncertain demand. Eldor and Zilcha (1990) considers how the existence of a futures market

[1]For discussion and references see e.g. Holmström and Tirole (1989), Scherer and Ross (1990), and Hay and Morris (1991).
[2]See e.g. Nance et al (1993) and Froot et al (1993).

affect the quantities set by risk averse oligopolists. Froot, et al (1993) discuss how hedging strategies can influence the possibilities for firms to undertake investments which influence the product market competition. Tessiotore (1994) shows that risk aversion in a differentiated product oligopoly may cause some firms to increase their production, compared to the case where all firms are risk neutral.

In this paper we present a simple, yet general, framework to study the behaviour of risk averse firms. Many of the previous results are derived as special cases, and we study a number of new questions. Under the assumption of normally distributed profits and marginal profits, it is shown that the first order conditions only depend on the covariance between these, the (possibly non constant) absolute risk aversion and expected marginal profits. This makes first order conditions easy to interpret, and useful for many applications where risk aversion may play a role. The intuition behind the strategic effects of risk aversion is that firms wish to be well adjusted in the bad realisations, where profits would otherwise be very low. For example, with demand uncertainty, profit is low when demand is low. When demand is low, it is better to have a low quantity or a low price, which leads to lower best response strategies than a risk neutral firm. With marginal cost uncertainty, a risk averse firm wishes to restrict output in realisations where costs are high and therefore sets relatively low quantities or high prices. In contrast to previous work on risk averse oligopolies, we allow for decreasing absolute risk aversion. If the objective function has this property, anything which influences the net wealth of the firm, e.g. fixed cost and previous profits, matters for the best response strategies. This is modelled in a two stage game where first period profits determine the intensity of the second period competition. A risk neutral firm can attempt to influence the first period profits of a risk averse rival, to induce softer competition in the second period. Finally, we examine how the predictions on concentration and pricing are affected by introducing risk aversion in two models of free entry where firms pay a fixed cost to enter. In a model of price competition, concentration is never lower under risk aversion, compared to the risk neutral case, but prices may be higher or lower. The results for quantity competition are ambiguous, concentration and quantities could be either higher or lower if firms are risk averse.

The structure of the paper is as follows. Section 2 derives first order conditions for risk averse firms in static oligopoly games, where the number of firms is taken as exogenous. Section 3 is an extension to a two stage game, where first period profits affect competition in the second period. Section 4 studies two models of free entry with fixed costs, to endogenise the number of firms. Section 5 concludes with some remarks on econometric specification in tests for risk aversion, and on the relation between risk aversion and limited liability.

2. Model

In this section we derive a framework to analyse the behaviour of risk averse oligopolists, under different assumptions about the nature of competition and types of uncertainty. Each firm has a twice continuously differentiable concave von Neuman - Morgenstern expected utility function, $EU[\pi, f]$, where π is (net) operating profit, and f is fixed (sunk) cost.[3] The partial derivatives are $U_{\pi}^{'} > 0$, $U_{f}^{'} < 0$, $U_{\pi\pi}^{''} \leq 0$, $U_{ff}^{''} \geq 0$. It is assumed that the expected utility in equilibrium is greater than the reservation utility.

To simplify the exposition we consider a duopoly with firms $i=1,2$, which simultaneously choose their strategies. The operating profits of firm i is a continuously differentiable function of its strategy, s_i, and the strategy of the other firm, s_j. Differentiating the objective function w.r.t. s_i yields the first order condition

$$EU_{\pi\pi_i}^{'}\pi_{is_i}^{'} = 0. \tag{1}$$

[3]This paper abstracts from capital market considerations, where the owner (or the manager) of the firm may have other risky assets. These risky assets can then be thought of as having some correlation with a market portfolio as in Harris (1986), and Tessiotore (1994). However, this could easily be incorporated into this framework by assuming that the utility is a function of the sum of firm profits and the return on a portfolio of other assets. Even if this is theoretically appealing, in practice it is likely to be difficult to assess the covariance between the payoffs of a strategy, and the return on a market portfolio.

Using the definition of covariance and rearranging, (1) is equivalent to

$$E\pi'_{is_i} + \frac{Cov(U'_{i\pi_i}, \pi'_{is_i})}{EU'_{i\pi_i}} = 0. \tag{2}$$

Some early work has analysed the first order conditions in the form of (2). In the second term the denominator is positive, and the numerator is negative if the covariance between profits and marginal profits is positive, (which gives a negative covariance between marginal profit and marginal utility). Under these conditions the expected marginal profit has to be positive to satisfy (2). However, it is difficult to analyse more complex problems and to explain the intuition behind the competitive effects of risk aversion based on (2). To simplify the exposition and arrive at tractable first order conditions, it is therefore assumed that π_i and π'_{is_i} are bivariate normally distributed

$$\pi_i, \pi'_{is_i} \sim BVN(E\pi_i, E\pi'_{is_i}, \sigma^2_{\pi_i}, \sigma^2_{\pi'_{is_i}}, \rho_i), \tag{A1}$$

where the correlation coefficient is

$$\rho_i = \frac{Cov(\pi_i, \pi'_{is_i})}{\sigma_{\pi_i}\sigma_{\pi'_{is_i}}}. \tag{3}$$

It is well known that the assumption of normality of payoffs gives a mean - variance model, since third and higher moments are zero.[4] However, under the assumption of normal distribution (and any other distribution with unbounded support) there is a positive probability that profits and marginal profits is either infinitely positive or negative. Still, even if the distribution is not exactly normal, it may for practical purposes be a good approximation to the true distribution over the relevant range.[5] Our conjecture is that the

[4]Mean - variance analysis can also be defended by assuming a quadratic utility function. However, it has the undesirable property of yielding increasing absolute risk aversion. In what follows we make use of decreasing absolute risk aversion which rule out quadratic utility functions.

[5]Normality of the density function is the standard defence of the mean - variance analysis in portfolio choice, even where stock prices are truncated at zero, see e.g. Ingersoll (1987 p. 95-97) or Huang and Litzenberger (1988 p. 61-62). To calculate the equilibrium strategies under the assumption that the distribution is approximately normal, it is necessary to truncate the distribution at some level. In Section 4 this problem is

results derived in this paper would be qualitatively unaffected if we would use other symmetric profit and marginal profit distributions, as explained after Proposition 1.

Under (A1) it is possible to use a theorem (Steins lemma) proved by Rubinstein (1976) by which[6]

$$Cov(U'_{i\pi_i}, \pi'_{is_i}) = EU''_{i\pi_i\pi_i} Cov(\pi_i, \pi'_{is_i}). \tag{4}$$

Define the Arrow-Pratt measure of global absolute risk aversion as

$$R_i \equiv -\frac{EU''_{i\pi_i\pi_i}}{EU'_{i\pi_i}}. \tag{5}$$

Combining (2), (4), and (5) and inserting the arguments yields the first order conditions

$$V'_{is_i} = E\pi'_{is_i}[s_i, s_j] - R_i\left[\pi_i[s_i, s_j], f_i\right] Cov(\pi_i[s_i, s_j], \pi'_{is_i}[s_i, s_j]) = 0. \tag{6}$$

The Arrow-Pratt measure of absolute risk aversion is positive under the assumption of risk aversion (and zero under risk neutrality). To be specific and to simplify some of the proofs we make the (overly strong) assumption that the utility function belongs to the HARA class with non increasing absolute risk aversion[7]

$$U_i[\pi_i, f_i] = \frac{1-\gamma}{\gamma}\left(\frac{\alpha(\pi_i - f_i)}{1-\gamma} + \beta\right)^{\gamma} \quad \gamma \leq 1, \beta > 0. \tag{A2}$$

The covariance between profits and marginal profits in (6) has an ambiguous sign in the general case. In Appendix 0 we show that with linear demand and cost functions it is

avoided by using a uniform distribution, but it is shown that the results in the propositions are qualitatively the same.

[6]See e.g. Huang and Litzenberger (1988 p. 101). The theorem states that if x and y are bivariate normally distributed, and *g[y]* is at least once differentiable then $Cov(x, g[y]) = Eg'[y]Cov(x,y)$.

[7]In what follows it is sufficient to assume a non negative and convex Arrow-Pratt measure, but to assure that the utility function is well behaved, HARA is convenient. The HARA class is the most widely used utility function in finance applications, and includes as special cases the negative exponential (with constant absolute risk aversion), quadratic, and power utility functions, Ingersoll (1987 p. 39-40).

possible to determine the sign of the covariance. Under demand uncertainty, it is positive both for strategic substitutes and complements, i.e. in realisations where the profits are high, marginal profits are high as well. Under marginal cost uncertainty the sign of the covariance will depend on the nature of competition. For strategic substitutes it is positive, but for strategic complements it is negative. The intuition is explained below.

Let s_i^*, s_j^* denote the unique pair of Nash strategies which satisfy the first order conditions for both firms. The second order conditions are $V''_{is_is_i} < 0$. Let the best response function be $BR_i\left[s_j\right]$, i.e. $s_i = BR_i\left[s_j\right]$ satisfies (6) for a given s_j. Differentiating (6) yields

$$BR'_{is_j}\left[s_j\right] = -\frac{V''_{is_is_j}}{V''_{is_is_i}}, \tag{7}$$

where

$$V''_{is_is_j} = E\pi''_{is_is_j} - R'_{i\pi_i}\pi'_{is_j}Cov(\pi_i, \pi'_{is_i}) - R_i\, Cov'_{s_j}(\pi_i, \pi'_{is_i}). \tag{8}$$

As in Bulow et al (1985), we say that s_i and s_j are strategic substitutes if $V''_{is_is_j} < 0$ for $i \neq j$, i.e. if best response functions are downward sloping. With strategic complements, the inequality is reversed. Informally, we denote the strategies "quantities" ("prices") for strategic substitutes (complements). To assure that a stable, unique equilibrium exists we assume that $V''_{is_is_i}V''_{js_js_j} - V''_{is_is_j}V''_{is_js_i} > 0$, (see Dixit (1986)). This condition can not be assumed to be met in general, but risk aversion seems to make the requirements for the existence of a stable equilibrium easier to satisfy, since it makes the objective functions more concave.

Under (A1) and (A2), (6) gives a simple and easily interpreted expression. It states that the expected marginal operating profit should equal the absolute risk aversion times the covariance of profits and marginal profits. Moreover, the expected marginal profit has to be strictly positive to satisfy (6) when $\rho_i > 0$, and $R_i > 0$, and strictly negative when the covariance is negative. To see the direct implication, consider first the case of a monopolist who has to set a single price or quantity under uncertainty. Let s^{0M} be the risk neutral monopolist's strategy and s^{*M} that of a risk averse monopolist.

Proposition 0 (Baron 1971)*:* If the monopolist's Arrow-Pratt measure of absolute risk aversion is positive and the distribution of profits and marginal profits is bivariate normal, then $s^{0M} > s^{*M}$ if $\rho > 0$, and $s^{0M} < s^{*M}$ if $\rho < 0$.

Proof: If $R = 0$ then the first order condition is $E\pi_s^{'}\left[s^{0M}\right] = 0$, and the second order condition is $E\pi_{ss}^{''}\left[s^{0M}\right] < 0$. If $R > 0$ and $\rho > 0$, then $E\pi_s^{'}\left[s^{*M}\right] > 0$, which implies that $s^{0M} > s^{*M}$, from the second order condition. For $R > 0$ and $\rho < 0$ the expected marginal profit has to be negative at the optimum and thus $s^{0M} < s^{*M}$. Q.E.D.

This is essentially the result first published in Baron (1971), but derived in a much simpler way. *If the covariance between profits and marginal profits is positive, then a quantity (price) setting risk averse monopolist sets lower quantities (prices), than a risk neutral counterpart.*

More generally, let θ_i be a parameter which influence firm i's absolute risk aversion (without changing the distribution of profits), and assume that it is increasing in θ_i, $R_{i\theta_i}^{'} > 0$.

Proposition 1: If firm i's Arrow-Pratt measure of absolute risk aversion is positive and the distribution of profits and marginal profits is bivariate normal, then changing firm i's absolute risk aversion have the following effects:

1) for $\rho_i > 0$ i) $ds_i^* / d\theta_i < 0$, ii) $ds_j^* / d\theta_i > 0$ if $V_{is_is_j}^{''} < 0$, and iii) $ds_j^* / d\theta_i < 0$ if $V_{is_is_j}^{''} > 0$.

2) for $\rho_i < 0$ i) $ds_i^* / d\theta_i > 0$, ii) $ds_j^* / d\theta_i < 0$ if $V_{is_is_j}^{''} < 0$, and iii) $ds_j^* / d\theta_i > 0$ if $V_{is_is_j}^{''} > 0$.

Proof: In Appendix 1.

With a positive covariance between profits and marginal profits, as in the demand uncertainty case in Appendix 0, firm i's Nash equilibrium strategy is decreasing in its absolute risk aversion, whereas the effect on the other firm's Nash strategy is dependent on the nature of competition. Moreover, for any s_j, the expected marginal profit of firm i has to be positive at the risk averse best response strategy, denoted $BR_i^{RA}[s_j]$. Analogous to Proposition 0, this implies a lower best response than the risk neutral best response strategy, $BR_i^{RN}[s_j]$. Similar results have been derived in other papers dealing with risk averse oligopolists e.g. Harris (1986), and Hviid (1989), but the simple intuition has not been stressed. *The risk averse firm gives relatively more weight to the bad realisations where profits are low. With positive covariance between profit and marginal profit, low quantities and prices are optimal in bad realisations. In a strategic context it implies that risk aversion shifts the best response function downward, because the firm wishes to be well adjusted in bad realisations.* However, with cost uncertainty the sign of the covariance (and thus the effect on the Nash equilibrium) was shown in Appendix 0 to be dependent on the nature of competition. The intuition here is that *high* prices or *low* quantities are better when marginal costs are high, since it restricts the quantity the firm has to sell in high cost states. This leads the risk averse firm's best response strategy to be lower under quantity competition, but higher under price competition.

This intuition motivates the conjecture that our results also hold for other symmetric distributions of payoffs. If strategies have symmetric distributions of profits, conditional on the rival's strategy, the risk averse firm will prefer strategies which do relatively well in the lower end of the distribution of realisations (and do relatively worse in the upper end). Then one must verify that for e.g. demand uncertainty a low price or quantity is the best response in the worst realisation, which is a plausible property.

Competition among risk averse firms will be "softer" or "tougher" compared to the risk neutral case, depending on whether they are assumed to set quantities or prices, and if uncertainty is primarily about the demand or cost conditions. Without information on the these variables, no prediction can be made on the expected profits in equilibrium, compared to the risk neutral case. However, the expected price-cost margins will be higher if firms are risk averse, except for the case with demand uncertainty and strategic complements. The difference between risk neutral and risk averse best response functions is illustrated in the positive covariance case in Figure 1A and 1B. The risk neutral best response functions are lines, and a risk averse best response function of firm i is dashed. Note that the iso profit function of firm i does not have an extreme point at the risk averse best response function, (which a iso utility function would).

FIGURE 1A AND 1B ABOUT HERE.

To continue with the comparative statics, let there be a change in the covariance of firm i, denoted Cov_i, which does not affect the expected profit and leave the covariance of firm j unaffected. That is, it is a mean preserving change in the idiosyncratic risk of firm i.

Proposition 2: If firm i's utility function is of the HARA class, with positive and non-increasing absolute risk aversion, and the distribution of profits and marginal profits is bivariate normal, then changing the risk of firm i have the following effects:
i) $ds_i^* / dCov_i < 0$, ii) $ds_j^* / dCov_i > 0$ if $V''_{is_is_j} < 0$, and iii) $ds_j^* / dCov_i < 0$ if $V''_{is_is_j} > 0$.

Proof: In Appendix 2.

Similar propositions have been derived in earlier papers, and the intuition is as follows. If $\rho_i > 0$ an increase in covariance shifts the best response function further down, compared to the risk neutral case. If $\rho_i < 0$ an increase in covariance implies lower risk, which shifts the best response function down towards the risk neutral best response function. The effect on the rival firm is that with strategic substitutes (complements) its equilibrium strategy increase (decrease).

Let *dCov* denote an equal change in the covariance of both firms, that does not affect expected profits. We informally refer to this as a mean preserving change in market risk. The following proposition prove that the effect of a change in market risk on equilibrium strategies is unambiguous only for strategic complements.

Proposition 3: If both firm i's and firm j's utility function are of the HARA class, with positive and non-increasing absolute risk aversion, and the distribution of profits and marginal profits is bivariate normal, then changing the market risk have the following effects:
i) $ds_i^* / dCov < 0$ if $V''_{is_is_j} > 0$, ii) $ds_i^* / dCov < 0$ if $V''_{is_is_j} < 0$ and $V''_{js_jCov}V''_{is_is_j} < V''_{is_iCov}V''_{js_js_j}$, and
iii) $ds_i^* / dCov > 0$ if $V''_{is_is_j} < 0$ and $V''_{js_jCov}V''_{is_is_j} > V''_{is_iCov}V''_{js_js_j}$.

Proof: In Appendix 3.

For strategic substitutes, $V''_{is_is_j} < 0$, and demand uncertainty, this result was independently derived by Tessiotore (1994). Increasing the covariance cause both firms' best response functions to shift down, but depending on the relative magnitude of these shifts, one of the firms may, in equilibrium, produce more than before. This is analogous to the case where both firms' marginal cost increase, which may, in a model with strategic substitutes, result in an equilibrium where one of the firms produce more than before. However, in Tessiotore's work there are no results that extends to strategic complements. Proposition 3 shows that there is an unambiguous negative effect on the equilibrium strategies. The intuition is that increases in covariance causes each firm's best response function to shift down to be better adjusted in the worst realisations. The best response to a lower price of the rival, is to reduce price further. This gives an unambiguous negative effect on each firm's equilibrium strategy.

One of the most interesting properties of risk aversion is that fixed costs matter for the strategy choice, if the utility function displays decreasing absolute risk aversion ($\gamma < 1$ in

(A2)). More fixed costs reduce firm's net wealth, which increases the absolute risk aversion. This was shown in the early works of Baron (1970), Sandmo (1971), and Leland (1972), but no work seems to have incorporated this insight into an oligopoly framework.

Proposition 4: If firm i's utility function is of the HARA class, with positive and decreasing absolute risk aversion, and the distribution of profits and marginal profits is bivariate normal, then changing the fixed costs of firm i have the following effects:

1) for $\rho_i > 0$ i) $ds_i^* / df_i < 0$, ii) $ds_j^* / df_i > 0$ if $V''_{is_is_j} < 0$, and iii) $ds_j^* / df_i < 0$ if $V''_{is_is_j} > 0$.

2) for $\rho_i < 0$ i) $ds_i^* / df_i > 0$, ii) $ds_j^* / df_i < 0$ if $V''_{is_is_j} < 0$, and iii) $ds_j^* / df_i > 0$ if $V''_{is_is_j} > 0$.

Proof: In Appendix 4.

For a positive covariance, an increase in firm i's fixed cost leads to a lower equilibrium strategy, but the effect on the other firm is, as usual, dependent upon the nature of competition. The intuition is simple. If the fixed costs are high, the bad realisations are even worse for a risk averse firm. To reduce the impact of these, the risk averse firm reduce its quantity or price to be better adjusted in the bad outcomes. If the covariance is negative, firm i will instead increase its price to limit the effect of the worst outcomes.

Proposition 4 have implications for the work on strategic investments. In Dixit (1980) it is assumed that an incumbent firm can invest in a cost reducing technology, which is a credible commitment to output expansion in a second stage. Since the initial investment is sunk in the second stage, the only effect it has for a risk neutral firm is that the cost reduction shifts the best response function up (even if there is uncertainty in the second stage). On the other hand, with a risk averse incumbent and uncertain demand ($\rho_i > 0$), the cost of investment reduce the wealth and thereby increase risk aversion, which may shift the best response function down enough to offset the cost reduction effect. For strategic complements the two effects work in the same direction. The cost reduction will reduce the best response strategy, and the investment cost further decrease it.

3. A two stage game

This section analyses the effect of previous profits on subsequent decisions. The intuition is that a firm's wealth level is dependent upon past profits. If these have been high, the firm is probably more willing to accept risk, and its behaviour will be closer to maximising expected profits. The strategic effect is illustrated in a two stage game, where first period profits influence the second period behaviour (under decreasing absolute risk aversion). It is shown that a risk neutral firm can rationally attempt to influence a risk averse competitors' profits, to soften future competition. As before, the uncertainty is normally distributed and it is assumed to be uncorrelated between periods 1 and 2. We also assume that the covariance between profits and marginal profits is positive in each period, and further that both firms are active in both periods, that is, exit is not possible. To clarify the effects it is assumed that only firm i is risk averse with expected utility function $EU_i\left[\pi_{i1}\left[s_{i1},s_{j1}\right]+\pi_{i2}\left[s_{i2},s_{j2}\right]\right]$, and that firm j is risk neutral and maximises expected profits $E\left[\pi_{j1}\left[s_{j1},s_{i1}\right]+\pi_{j2}\left[s_{j2},s_{i2}\right]\right]$.

As usual, the problem is solved by first studying firm i's first order condition in the second stage, when first period profits have been realised, $\bar{\pi}_{i1}$,

$$V'_{i2s_{i2}} = E\pi'_{i2s_{i2}} - R_i\left[\bar{\pi}_{i1}+\pi_{i2}\right]Cov(\pi_{i2},\pi'_{i2s_{i2}}) = 0. \tag{9}$$

Differentiation of (9) w.r.t s_{i2} and s_{j1} gives

$$BR'_{i2s_{j1}}\left[s_{j2}\right] = \frac{R'_{i\pi_{i1}}\bar{\pi}'_{i1s_{1j}}Cov(\pi_{i2},\pi'_{i2s_{i2}})}{V''_{i2s_{i2}s_{i2}}}\ . \tag{10}$$

Under the assumption of decreasing absolute risk aversion we have $R'_{i\pi_{i1}}<0$ and with positive covariance, the sign of $BR'_{i2s_{j1}}\left[s_{j2}\right]$ is that of $\bar{\pi}'_{i1s_{1j}}$. For example, if increasing firm j's strategy in the first period results in a reduced first period profit for firm i, it implies that firm i's second period strategy is lower. Turning to the first stage and using firm j's first order condition gives

$$V'_{j1s_{j1}} = E\left[\pi'_{j1s_{j1}} + \pi'_{j2s_{j2}}s'_{j2s_{j1}} + \pi'_{j2s_{i2}}s'_{i2s_{j1}}\right] = E\left[\pi'_{j1s_{j1}} + \pi'_{j2s_{i2}}s'_{i2s_{j1}}\right] = 0. \tag{11}$$

The second term is the strategic effect. From (10), the sign of $s'_{i2s_{1j}}$ will be that of $\overline{\pi}'_{i1s_{1j}}$, and assuming that $\text{sign}(\overline{\pi}'_{i1s_{1j}}) = \text{sign}(E\pi'_{j2s_{i2}})$, the strategic effect is positive. By an argument similar to that in the proof of Proposition 0, this implies that the expected marginal profit of firm j in the first stage has to be negative. Stated differently, the first period best response strategy of firm j, denoted $BR_{j1}^{RNS}[s_{i1}]$, will be higher than the best response without strategic effect, $BR_{j1}^{RN}[s_{i1}]$.

There is no strategic effect in firm i's first stage FOC, because firm j is risk neutral and will maximise expected profits in the second period, whatever its first period profit was. The first order condition for firm i is

$$EU'_{i\pi}\left[\pi'_{i1s_{i1}} + \pi'_{i2s_{i2}}s'_{i2s_{i1}} + \pi'_{i2s_{j2}}s'_{j2s_{i1}}\right] = EU'_{i\pi}\pi'_{i1s_{i1}} = 0. \tag{12}$$

By previous derivations this is equivalent to

$$V'_{i1s_{i1}} = E\pi'_{i1s_{i1}} - R_i\left[\pi_{i1} + \pi_{i2}\right]Cov(\pi_{i1}, \pi'_{i1s_{i1}}) = 0. \tag{13}$$

By the proof of Proposition 1 the best response function of firm i with the strategic effect, $BR_{i1}^{RAS}\left[s_{j1}\right]$ will be below its risk neutral best response function without the strategic effect, $BR_{i1}^{RN}\left[s_{j1}\right]$. Thus, compared to a risk neutral game, the risk averse firm's first period best response function shifts down, and the risk neutral firm's best response function shifts up.

Proposition 5: If firm j is risk neutral and firm i's utility function is of the HARA class, with positive and decreasing absolute risk aversion, and the distribution of profits and marginal profits is bivariate normal with positive correlation within each time period, but independent between periods, then $BR_{j1}^{RNS}\left[s_{i1}\right] > BR_{j1}^{RN}\left[s_{i1}\right]$ and $BR_{i1}^{RAS}\left[s_{j1}\right] < BR_{i1}^{RN}\left[s_{j1}\right]$.

With strategic substitutes, the strategic effect make the risk neutral firm to be more aggressive in the first period, to reduce the rival's profit and thereby increase its risk aversion, which makes the second period competition softer.[8] [9] By Proposition 1, the risk averse firm has lower best response strategies than under risk neutrality. For strategic substitutes the risk neutral (averse) firm's first period equilibrium quantity is higher (lower), compared to the case wbere both firms are risk neutral. For strategic complements, the equilibrium price of the risk averse firm is lower, compared to the risk neutral game. The effect on the first period price of the risk neutral firm is ambiguous. The risk neutral firm's expected profits will be lower in both periods for strategic complements. This is illustrated in Figure 2A and 2B, where the first period best response functions in the case both firms are risk neutral are lines, and those where firm i is risk averse are dashed.

FIGURE 2A AND 2B ABOUT HERE.

In this section it has been assumed that both firms will continue to compete in the second period, whatever their first period profits were. However, it is quite possible that the risk averse firm will exit in the second stage if the expected utility of continuation is lower than a reservation utility, even though the expected profit is positive. The game would then be one of predation, where the risk neutral firm may attempt to depress the risk averse firm's first period profits, to make its risk premium large enough to induce exit. This is an alternative explanation to why predatory behaviour can be rational. However, if predation fails in the strategic complements case, such that the risk averse firm remains in the second

[8]Note that the strategic effect would exist even if there was no uncertainty in the first period, but only in the second period. The first period best response function of the risk averse firm would then be the same as under risk neutrality. However, the risk neutral firm still has an incentive to influence the risk averse firm's first period profit to soften second period competition.
[9]The insight that past profits matter for future competition is related to the model by Glazer (1994) wbere indebted firms compete in two stages. Limited liability of equity bolders make them risk seeking, and more so the more outstanding debt there is. First period profits will determine the net debt in the second period. With quantity competition, it is shown that firms will reduce their quantities to give the competitor higher profits, which makes second period competition softer. Risk aversion has the opposite effect, the risk neutral firm expands output to reduce the profits of the rival and thereby increases its second period risk aversion to soften competition.

period, this could be detrimental to the risk neutral firm's profits in the second period, when the risk averse firm (with low net wealth) set even lower prices. A more rigorous treatment of this question awaits future research.

The empirical prediction from Proposition 5 is that risk aversion is low when past industry profits have been high, and the current "net wealth" of firms is high. Even though, as shown in Section 2, it is impossible to exactly predict how reduced risk aversion influence competition without information on the strategic variable and the type of uncertainty, we know that in three out of four combinations of uncertainty and strategic variable, this leads to lower expected price-cost margins in the next period. The implication is that under "good" times, firms are more willing to accept risks, which make competition more intense. This tend to make price-cost margins less pro cyclical over the business cycle. Of course, this can also be explained by other models e.g. price rigidities, and "price wars during booms". Empirical evidence on this seems mixed, but in Domowitz et al (1987) there is some support for counter cyclical price-cost margins in concentrated industries. However, there is not enough evidence to either accept or reject the hypothesis that past profit influences risk aversion and thereby future strategic interaction.

4. Endogenous number of firms

This section considers two well known oligopoly models where firms pay a fixed cost to enter the market and thereafter compete in prices or quantities. The number of firms in equilibrium is determined by a reservation utility rather than the usual non negative profit condition. The fixed cost can be thought of as the cost to participate in a lottery, where the payoff in the lottery is the uncertain profit from one period of competition. The utility of participating in the lottery is lower for a risk averse firm, which will therefore not participate in all lotteries which a risk neutral firm accepts. Our point is that the lotteries they face are dependent upon the risk aversion of the participants as well as the cost of participating. In a related paper, Appelbaum and Katz (1986) considers a model of perfect competition where risk averse firms set quantities under demand uncertainty. They show that the aggregate quantity on the market is lower, but that the number of firms can be either lower or higher than under risk neutrality. In contrast, our focus is on strategic models with both quantity and price setting risk averse firms. Moreover, we use numerical examples to illustrate the effects.

We first use a version of the circular city model of Salop (1979), where firms pay a fixed (sunk) entry cost, and thereafter compete in prices.[10] Consumers are uniformly distributed around the unit circle, and each consumer buys either one or zero units of the good. Their willingness to pay, ε, is uncertain with (bounded) support $\left[\underline{\varepsilon},\bar{\varepsilon}\right]$, and perfectly correlated across consumers [11] In the first stage firms decide on whether or not to enter and pay f. In the second stage, when firms $i=1,..,n$ have entered the market, they set prices p_i while consumers willingness to pay is unknown. The travel cost is t per unit and firms are assumed to be located at maximum distance, $1/n$, from each other.[12] The marginal cost of production is set to zero.

Define d as the location of the marginal consumer, who is indifferent between buying from firm 1 and paying p_1 plus the travel cost td, and from the closest neighbour with the price p, and incurring travel cost $t(1/n-d)$. Formally

$$p_1 + td = p + t(1/n-d). \tag{14}$$

Solving for d yields the demand facing firm 1 from one side

$$d = \frac{p - p_1 + t/n}{2t}. \tag{15}$$

[10]See Eaton and Lipsey (1989), and Tirole (1988) for surveys and extensions of the bare bones model used here.

[11]The reason for not using the normal distribution of the uncertainty is that we wish to calculate the expected utility and the certainty equivalent, and to do this we would need to truncate the normal distribution at some arbitrary level.

[12]It can also be shown that this is the unique equilibrium locations Kats (1995).

Next, define the critical realisation of the uncertainty, ε^c, for which the consumer located at d is indifferent between buying from firm 1 (or the competitor) and not buying at all,[13] by

$$\varepsilon^c = p_1 + td = p_1 + t\frac{p - p_1 + t/n}{2t} = \frac{1}{2}\left(p_1 + p + t/n\right). \tag{16}$$

For the realisations $\underline{\varepsilon} \leq \varepsilon < \varepsilon^c$, firm 1 is a monopolist for the consumers who are located near it. The location of the marginal consumer, x, is defined by

$$\varepsilon = p_1 + tx. \tag{17}$$

Solving for x gives the demand for firm 1 on one side

$$x = \frac{\varepsilon - p_1}{t}. \tag{18}$$

The risk firm 1 faces is that by setting a high price, there will be realisations where consumers' valuation is so low that the market will not be covered, and consumers located near d will not buy.

To simplify the analysis, it is assumed that consumers valuation is uniformly distributed on $[\underline{\varepsilon}, \bar{\varepsilon}]$ with density $\phi(\varepsilon) = 1/(\bar{\varepsilon} - \underline{\varepsilon})$ and cumulative probability $\Phi(\varepsilon) = (\varepsilon - \underline{\varepsilon})/(\bar{\varepsilon} - \underline{\varepsilon})$. Thus $\varepsilon^c \leq \varepsilon \leq \bar{\varepsilon}$ with probability $1 - \Phi(\varepsilon^c) = (\bar{\varepsilon} - \varepsilon^c)/(\bar{\varepsilon} - \underline{\varepsilon})$. The profit (net of fixed costs) in these high realisations is $\pi_1\left[p_1, p \middle| \varepsilon > \varepsilon^c\right] = 2p_1 d = p_1\left(p - p_1 + t/n\right)/t$. For any lower valuation $\underline{\varepsilon} \leq \varepsilon < \varepsilon^c$ the profit is $\pi_1\left[p_1, p \middle| \varepsilon \leq \varepsilon^c\right] = 2p_1 x = 2p_1\left(\varepsilon - p_1\right)/t$.

To illustrate, assume that all firms have a negative exponential utility function[14] with the same absolute risk aversion

$$U_i\left[\pi_i\right] = -e^{-\alpha\pi_i}. \tag{19}$$

[13]Conditional on that there exist a $\underline{\varepsilon} \leq \varepsilon^c \leq \bar{\varepsilon}$.

[14]This implies that the fixed entry cost will not influence the choice of strategy, which simplifies the interpretation. Introducing strictly decreasing absolute risk aversion would strengthen the results, since prices are decreasing in fixed cost (Proposition 4). In terms of (A2) $\beta = 1$, and $\gamma = -\infty$.

The expected utility for firm 1 is

$$EU_1 = -\frac{\bar{\varepsilon} - \varepsilon^c}{\bar{\varepsilon} - \underline{\varepsilon}} e^{-\alpha \frac{p_1(p - p_1 + t/n)}{t}} - \int_{\underline{\varepsilon}}^{\varepsilon^c} e^{-\alpha \frac{2p_1(\varepsilon - p_1)}{t}} \frac{1}{\bar{\varepsilon} - \underline{\varepsilon}} d\varepsilon . \tag{20}$$

The first term is the utility of profits, conditional on $\varepsilon^c \leq \varepsilon \leq \bar{\varepsilon}$, weighted with the probability that the realisation is within this range. The second term is the expected utility of profits when $\underline{\varepsilon} \leq \varepsilon < \varepsilon^c$.

Differentiate (20) w.r.t. p_1 and note that by (16) $d\varepsilon^c/dp_1 = 1/2$ which yields the first order conditions

$$\begin{aligned} EU'_{1p_1} &= \left(\frac{1}{2} + \left(\bar{\varepsilon} - \varepsilon^c\right)\alpha \frac{\left(p - 2p_1 + t/n\right)}{t}\right) e^{-\alpha \frac{p_1(p - p_1 + t/n)}{t}} - \\ &\frac{1}{2} e^{-\alpha \frac{2p_1(\varepsilon^c - p_1)}{t}} + 2\alpha \int_{\underline{\varepsilon}}^{\varepsilon^c} \frac{(\varepsilon - 2p_1)}{t} e^{-\alpha \frac{2p_1(\varepsilon - p_1)}{t}} d\varepsilon = 0. \end{aligned} \tag{21}$$

We search for a symmetric (subgame perfect) Nash equilibrium where the prices for all firms are equal, $p_i^* = p^*$. It is impossible to get closed form solutions to (21) (except for the risk neutral case). However, it is easy to solve them numerically, and the results for a set of parameters are shown in table 1.

Table 1. Nash equilibrium prices, expected profits, and certainty equivalents for $t = 0.5, \underline{\varepsilon} = 0.2, \bar{\varepsilon} = 0.5$

n	α	$p_i^* = p^*$	$E[\pi_i]$	Certainty equivalent
1	0	0.1750	0.1225	0.1225
2	0	0.1684	0.07442	0.07442
3	0	0.1480	0.04836	0.04836
1	1	0.1724	0.1225	0.1207
2	1	0.1676	0.07422	0.07407
3	1	0.1479	0.04833	0.04833
1	2	0.1699	0.1224	0.1189
2	2	0.1668	0.07403	0.07373
3	2	0.1478	0.04831	0.04830

As in Proposition 1, the equilibrium price *for a given number of firms* is a decreasing function of the absolute risk aversion.[15] For a risk neutral ($\alpha=0$) duopoly the equilibrium price is *0.1684* and the expected profit is *0.07442*. With $\alpha=1$, the equilibrium price is *0.1676* and the certainty equivalent is *0.07407*. This implies that fewer firms may enter if they are risk averse. For example, the free entry equilibrium will be a monopoly for $0.07442<f<0.1225$ if $\alpha=0$, and $0.07407<f<0.1207$ if $\alpha=1$. That is, for $0.07407<f<0.07442$ the equilibrium outcome is a duopoly if firms are risk neutral but a monopoly if they are risk averse with $\alpha=1$. However, for $0.07442<f<0.1207$ the equilibrium is a monopoly in both cases, and the price is lower under risk aversion. The general point we want to stress with this example is that even if price competition is tougher with a given number of risk averse firms, free entry with fixed costs may lead to an equilibrium with fewer firms and softer competition. In this model, the number of firms is never lower (concentration never higher) under risk neutrality than under risk aversion. This result is at odds with Katz and Appelbaum (1986) where in a quantity setting framework, concentration can be either higher or lower compared to the risk neutral outcome. The intuition is as follows. A risk averse firm needs a risk premium to participate in a game with uncertainty. However, the expected profit will be lower for a given number of risk averse firms, since they set lower prices. To get the necessary risk premium may require that fewer firms enter in equilibrium. Another point to be noted is that in this example the difference between the Nash equilibrium strategies under risk aversion and risk neutrality is small, which is the case for many specifications. It remains to study how robust this finding is to alternative specifications of the demand and utility functions.

The effects of risk aversion on equilibrium outcomes are less clear cut with competition in strategic substitutes. Even though risk averse firms need compensation to bear risk, the competition is less intense and it may therefore be possible that the higher expected profits (for a given number of firms) more than offset the required risk premium, to give a less

[15] For $n\geq 4$ there are enough firms to cover the market even if $\varepsilon=\underline{\varepsilon}$. The equilibrium price is then independent of uncertainty. For $2\leq n\leq 3$ it can be verified that $\underline{\varepsilon} < \varepsilon^c < \bar{\varepsilon}$ in equilibrium.

concentrated industry structure. To check if this conjecture is correct, we study a homogenous good Cournot oligopoly, with free entry and entry cost f in the first stage. Marginal cost of production is set to zero.

Let the inverse demand function (when n firms have entered) be linear with uncertain intercept, ε,

$$p_i = \varepsilon - bq_i - b\sum_{j\neq i}^{n} q_j . \tag{22}$$

As in the circular city model, ε is assumed to be uniformly distributed with support $\left[\underline{\varepsilon},\bar{\varepsilon}\right]$. The operating profit is

$$\pi_i = q_i\left(\varepsilon - bq_i - b\sum_{j\neq i}^{n} q_j\right). \tag{23}$$

Let all firms have expected utility functions of the HARA class, with the same parameters

$$EU_i = \int_{\underline{\varepsilon}}^{\bar{\varepsilon}} \frac{1-\gamma}{\gamma}\left(\frac{\alpha\left(\pi_i - f\right)}{1-\gamma} + \beta\right)^{\gamma} \frac{1}{\bar{\varepsilon}-\underline{\varepsilon}} d\varepsilon . \tag{24}$$

Differentiating (24) w.r.t. q_i yields the first order conditions

$$EU'_{iq_i} = \alpha \int_{\underline{\varepsilon}}^{\bar{\varepsilon}} \left(\varepsilon - 2bq_i - b\sum_{j\neq i}^{n} q_j\right)\left(\frac{\alpha(q_i(\varepsilon - bq_i - b\sum_{j\neq i}^{n} q_j) - f)}{1-\gamma} + \beta\right)^{\gamma-1} d\varepsilon = 0. \tag{25}$$

(25) is solved numerically and the symmetric Nash equilibrium quantities, profits, expected utility, and certainty equivalent for a set of parameters are shown in Table 2.

Table 2. Nash equilibrium prices, expected profits, and certainty equivalents for $\underline{\varepsilon} = 1, \bar{\varepsilon} = 1.5, b = 1$

n	f	α, β, γ	$q_i^* = q^*$	$E[\pi_i]$	Certainty equivalent
1	0	0,1,-∞	0.625	0.391	0.391
2	0	0,1,-∞	0.417	0.174	0.174
3	0	0,1,-∞	0.312	0.0977	0.0977
4	0	0,1,-∞	0.250	0.0625	0.0625
1	0	2,1,-∞	0.612	0.390	0.383
2	0	2,1,-∞	0.411	0.176	0.172
3	0	2,1,-∞	0.309	0.0996	0.0976
4	0	2,1,-∞	0.248	0.0640	0.0628
1	0.0625	2,0.125,-2	0.589	0.389	0.305
2	0.0625	2,0.125,-2	0.390	0.183	0.105
3	0.0625	2,0.125,-2	0.293	0.109	0.0345
4	0.0625	2,0.125,-2	0.235	0.0728	0.00138
1	0.0975	2,0.125,-2	0.587	0.389	0.269
2	0.0975	2,0.125,-2	0.387	0.184	0.0690
3	0.0975	2,0.125,-2	0.290	0.110	-0.000776
4	0.0975	2,0.125,-2	0.232	0.0745	-0.0336

As in Proposition 1, the equilibrium strategies are decreasing in absolute risk aversion, for any number of firms.[16] It can also be noted that the difference between strategies under risk aversion and risk neutrality is quite small, as in the circular city model.

In the constant absolute risk aversion case, ($\beta = 1, \gamma = -\infty$), comparing risk neutrality ($\alpha = 0$) with risk aversion ($\alpha = 2$) reveals that for $1 \leq n \leq 3$ the certainty equivalent is lower than the expected profit under risk neutrality. This implies, that e.g. a duopoly is sustainable for $0.0977 < f < 0.174$ if $\alpha = 0$ but only for $0.0976 < f < 0.172$ if $\alpha = 2$. Thus there is a range of fixed cost where only a monopoly is sustainable under risk aversion, but a duopoly is the equilibrium outcome under risk neutrality. However, for $n = 4$ the certainty equivalent is higher than the expected profit under risk neutrality. For $0.0625 < f < 0.0628$ four firms are sustainable under risk aversion but only three if they are risk neutral. By this we have demonstrated that the competition reducing effect of risk aversion may dominate the risk

[16]It can be verified that the equilibrium price is non negative in all cases discussed here, even in the worst realisation.

premium effect. A market with risk averse firms can thus be less concentrated than with risk neutral firms. This extends the finding of Appelbaum and Katz (1986) to the case with quantity setting oligopolies.

With decreasing absolute risk aversion ($\alpha=2, \beta=0.125, \gamma=-2$) the equilibrium strategies are decreasing in the fixed cost, for a given number of firms, as shown in Proposition 4. For $f=0.0625$ and $n=4$ the certainty equivalent is 0.00138, but the expected profit minus the fixed costs is zero under risk neutrality. This implies that a risk neutral firm is indifferent between participating and staying out, whereas the risk averse firm strictly prefers participating in a quadropoly with fixed cost $f=0.0625$. Again, the competition reducing effect dominates and may lead to a market with lower concentration under risk aversion. In contrast, if $f=0.0975$ and $n=3$ the expected net profit is strictly positive for a risk neutral firm, but the certainty equivalent is -0.000776 and the risk averse firm therefore strictly prefers to stay out.

5. Concluding remarks

There exist very little systematic empirical evidence on how firms risk preferences are reflected in their choice of product market strategies. It is therefore useful to see if there are some indirect evidence from early structure-conduct-performance (SCP) studies. One of the most well documented regularities in SCP studies is that profitability is positively correlated with some measure of capital requirements in the industry, (see e.g. Schmalensee (1989)). This is usually explained by that a large capital requirement is a proxy for a large minimum efficient scale, or the amount of sunk investments, which may be an entry barrier. It was shown in the paper that risk aversion provide a complementary explanation. More fixed cost increase the risk aversion, which tend to result in higher expected operating profits. Moreover, if firms are risk averse in their entry decision, industries where initial investments are large, and where there is significant uncertainty about future conditions, can sustain high profit levels without attracting new entry, a possibility that was illustrated in the models of free entry. This is also consistent with evidence, which suggest that entry rates are negatively correlated with capital requirements. Relatively few empirical studies have directly tested if some measure of risk is correlated with profitability at the firm and industry level. However, of those that include this variable, the majority find a positive relation. In Schmalensee's (1989) survey, five studies found a significant positive correlation, three were insignificant, and two found a negative correlation. If one wants to explain this finding without risk aversion, measurement errors or selection bias are the primary candidates. Selection bias may arise if very low profits drive firms into bankruptcy, which make it unlikely that they are included in the datasets, while those with very high profits remain in the industry and are included.

One point to comment on is the, admittedly unrealistic, assumption in the paper that each firms' utility function is known to the rivals. Even it is not known to the competitors exactly, they may still have an idea on the degree of risk aversion of the other. Some information which may be available to them, and serve as proxies of risk attitude, are the use of insurance, or the amount of corporate hedging. This type of variables could then also be used in the traditional SCP studies, since they will have no explanatory power under risk neutrality, but may show up significant if firms are risk avert. Other proxies can be the ownership and control structure of firms, e.g. if the owners and management are undiversified. When it comes to selection of the relevant risk measure, it will depend on whether it primarily effects the demand or cost conditions. It was demonstrated in the paper that the effect of cost uncertainty had an unambiguous positive effect on expected price-cost margins, while the results in the demand uncertainty case depends on the nature of competition. This is a testable prediction, which requires good measures of demand and cost uncertainty. In the latter case, standard deviations of the input prices, and wages would provide reasonable proxies. To find proxies for demand uncertainty, one carefully needs to examine each industry in order to find the relevant measures.

A final point to comment on is the relation between risk aversion and limited liability of equity holders in debt financed firms. That a risk averse firm wish to be well adjusted in bad realisations of uncertainty is directly opposite from what is expected for indebted firms with limited liability, which can ignore bad realisations (see Brander and Lewis (1986), and Asplund (1995)). When models of limited liability have been tested, only weak evidence is found for the risk increasing effect predicted by the limited liability of equity. In Phillips (1995) a sample of firms which made a leveraged buy out is studied. The assumption of limited liability suggest that these firms would expand output, but instead they seem to bave contracted output. In terms of risk aversion, this would be interpreted as if the new owners were more risk averse, since a large fraction of their wealth are now tied to the firm. It is quite possible that this effect dominates the limited liability effect. An interesting topic for future research is to study the case where risk averse equity holders control the firm, while still enjoying limited liability in the worst realisations.

References

Appelbaum, E., and E. Katz (1986), Measures of risk aversion and comparative statics of industry equilibrium, *American Economic Review*, 76:524-9.

Asplund, M. (1995), Oligopoly and limited liability, Stockholm School of Economics working paper No.45.

Baron, D.P. (1970), Price uncertainty, utility, and industry equilibrium in pure competition, *International Economic Review*, 11:463-480.

Baron, D.P. (1971), Demand uncertainty and imperfect competition, *International Economic Review*, 12:196-208.

Brander, J.A., and T.R. Lewis (1986), Oligopoly and financial structure: The limited liability effect, *American Economic Review*, 76:956-970.

Bulow, J., J.Geanakopolos, and P.Klemperer (1985) Multimarket oligopoly: strategic substitutes and complements, *Journal of Political Economy*, 93:488-511.

Dixit, A. (1980), The role of investment in entry deterrence, *Economic Journal*, 90:95-106.

Dixit, A. (1986), Comparative static's for oligopoly, *International Economic Review*, 27:107-122.

Domowitz, I., R.G. Hubbard, and B.C. Petersen (1987), Oligopoly supergames: Some empirical evidence on prices and margins, *Journal of Industrial Economics*, 35:379 -398.

Eaton, B.C., and R.G.Lipsey (1989), Product differentiation, in R.Schmalensee and R.D.Willig eds., Handbook of Industrial Economics Vol. 1, Elsevier Science Publishers B.V. North Holland

Eldor, R. and I.Zilcha (1990),Oligopoly, uncertain demand, and forward markets, *Journal of Economics and Business*, 42:17-26.

Froot, K.A., D.S. Scharfstein, and J.C. Stein (1993), Risk management: Coordinating corporate investment and financial policies, *Journal of Finance*, 48:1629-1658.

Glazer,J. (1994), The strategic effects of long-term debt in imperfect competition, *Journal of Economic Theory*, 62:428-443.

Harris, F.H.Deb. (1986), Market structure and price-cost performance under endogenous profit risk, *Journal of Industrial Economics*, 35:35-59.

Hay, D.A. and D. J. Morris (1991), Industrial economics and organisation. Theory and evidence, Oxford University Press, Oxford.

Holmstrom, B.R. and J.Tirole (1989), The theory of the firm, in R.Schmalensee and R.D.Willig eds., Handbook of Industrial Economics Vol. 1, Elsevier Science Publishers B.V. North Holland

Holthausen, D.M. (1979), Kinky demand, risk aversion and price leadership, *International Economic Review*, 20:341-348.

Huang, C-f. and R.H.Litzenberger (1988), Foundations of financial economics, Elsevier Science Publishers B.V.

Hvid, M. (1989), Risk-averse duopolists and voluntary information transmission, *Journal of Industrial Economics*, 38:49-64.

Kats, A. (1995), More on Hotelling's stability in competition, *International Journal of Industrial Organization*, 13:89-93.

Ingersoll, J.E. (1987), Theory of financial decision making, Rowman and Littlefield Publishers Inc.

Leland, H. (1972), Theory of the firm facing random demand, *American Economic Review*, 62:278-291.

Nance, D.R., C.W.Smith,Jr, and C.W. Smithson (1993), On the determinants of corporate hedging, *Journal of Finance*, 48:267-284.

Perrakis, S. and G.Warskett (1983),Capacity and entry under demand uncertainty, *Review of Economic Studies*, 50:495-511.

Phillips, G.M. (1995), Increased debt and industry product markets. An empirical analysis, *Journal of Financial Economics*, 37:189-238.

Rubinstein, M. (1976), The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics and Management Science*, 7:407-425.

Sandmo,A. (1971), On the theory of the competitive firm under price uncertainty, *American Economic Review*, 61:65-73.

Salop, S.C. (1979), Monopolistic competition with outside goods, *Bell Journal of Economics and Management Science*, 10:141-156.

Schmalensee, R. (1989), Inter-industry studies of structure and performance, in R.Schmalensee and R.D.Willig eds., Handbook of Industrial Organization Vol. 2, Elsevier Science Publishers B.V. North Holland .

Scherer, F.M. and D.Ross (1990), Industrial market structure and performance, third edition, Houghton Mifflin Company, Boston.

Tessiotore, A. (1994), Market segmentation and oligopoly under uncertainty, *Journal of Economics and Business*, 46:65-76.

Tirole, J. (1988), The theory of industrial organisation, MIT Press, Cambridge, Mass.

Appendix 0

Consider the two profit functions, based on linear demand and cost functions

1) $\pi_i[s_i,s_j]=s_i(a-b_1s_i-b_2s_j-c)$, with $\pi'_{is_i}[s_i,s_j]=a-2b_1s_i-b_2s_j-c$

2) $\pi_i[s_i,s_j]=(s_i-c)(a-b_1s_i+b_2s_j)$, with $\pi'_{is_i}[s_i,s_j]=a-2b_1s_i+b_2s_j+b_1c$

The covariances, $Cov\{\pi_i,\pi'_{is_i}\}$, are then:

If $a\sim N\{Ea,\sigma_a^2\}$ then: 1) $s_i\sigma_a^2>0$. 2) $(s_i-c)\sigma_a^2>0$.

If $b_1\sim N\{Eb_1,\sigma_{b_1}^2\}$ then: 1) $2s_i^3\sigma_{b_1}^2>0$. 2) $\left((s_i-c)^2s_i+(s_i-c)s_i^2\right)\sigma_{b_1}^2>0$.

If $b_2\sim N\{Eb_2,\sigma_{b_2}^2\}$ then: 1) $s_is_j^2\sigma_{b_2}^2>0$. 2) $(s_i-c)s_j^2\sigma_{b_2}^2>0$.

If $c\sim N\{Ec,\sigma_c^2\}$ then: 1) $s_i\sigma_c^2>0$. 2) $-b_1(a-b_1s_i+b_2s_j)\sigma_c^2<0$.

In the demand uncertainty cases, the covariance is positive for both profit functions. When the marginal cost is uncertain, the covariance is positive in the first profit function where firms set quantities, but negative in the second where prices are the strategic variables.

Appendix 1

Differentiating the first order conditions yields

$$\begin{aligned}V''_{is_is_i}ds_i+V''_{is_is_j}ds_j+V''_{is_i\theta_i}d\theta_i=0\\ V''_{js_js_i}ds_i+V''_{js_js_j}ds_j+V''_{js_j\theta_i}d\theta_i=0\end{aligned}$$

The direct effect of a change in firm i's risk aversion is zero for firm j, so by Cramers rule

$$\frac{ds_i}{d\theta_i}=-\frac{V''_{is_i\theta_i}V''_{js_js_j}}{DET},\text{ and }\frac{ds_j}{d\theta_i}=\frac{V''_{is_i\theta_i}V''_{js_js_i}}{DET}.$$

Under the stability condition $DET=V''_{is_is_i}V''_{js_js_j}-V''_{is_is_j}V''_{js_js_i}>0$ and by the second order condition $V''_{js_js_j}<0$. The derivative of the first order condition w.r.t. θ_i is

$$V''_{is_i\theta_i}=-R'_{i\theta_i}\left[\pi_i[s_i,s_j],f_i\right]Cov(\pi_i[s_i,s_j],\pi'_{is_i}[s_i,s_j]).$$

If $\rho_i>0$ ($\rho_i<0$) the covariance is positive (negative), and by assumption $R'_{i\theta_i}>0$, and thus $ds_i^*/d\theta_i<0$ ($ds_i^*/d\theta_i>0$). The change in equilibrium strategy of firm j has the opposite sign of $V''_{js_js_i}$. Q.E.D.

Appendix 2

Similar to the proof in Appendix 1. The sign of $ds_i^* / dCov_i$ is that of $V''_{is_iCov_i}$. The derivative of the first order condition w.r.t. Cov_i is

$$V''_{is_iCov_i} = -R_i - R'_{iCov_i} Cov_i .$$

The first term R_i is positive when the absolute risk aversion is positive, which we have assumed in (A2). The sign of the second term $R'_{iCov_i} Cov_i$ can also be determined. Note that the Arrow Pratt measure for the class of HARA utility functions considered here is strictly convex, except for the limiting case of constant absolute risk aversion. For constant absolute risk aversion then $R'_{iCov_i} = 0$. For $Cov_i > 0$, increasing the covariance means an increase in the risk, whereas it is a decrease in risk if $Cov_i < 0$. In the first case it implies that the Arrow-Pratt measure is increasing by Jensen's inequality, since it is a mean preserving increase in risk. The sign of the second term is then positive and therefore $V''_{is_iCov_i} < 0$. For negative covariance, the Arrow-Pratt measure is decreasing, also by Jensen's inequality, since it is a mean preserving reduction in risk. The sign of the second term is again positive (since it is the product of two negative numbers), and thus $V''_{is_iCov_i} < 0$. From this follows that $ds_i^* / dCov_i < 0$, and the effect on $ds_j^* / dCov_i$ is dependent on the sign of $V''_{js_js_i}$. Q.E.D.

Appendix 3

This is similar to the proof in Appendix 2, except that both firms best response functions now are shifted. First differentiate the first order conditions,

$$\begin{aligned} V''_{is_is_i} ds_i + V''_{is_is_j} ds_j + V''_{is_iCov} dCov &= 0 \\ V''_{js_js_i} ds_i + V''_{js_js_j} ds_j + V''_{js_jCov} dCov &= 0 \end{aligned}$$

By Cramers rule

$$\frac{ds_i}{dCov} = \frac{V''_{js_jCov}V''_{is_is_j} - V''_{is_iCov}V''_{js_js_j}}{DET}.$$

However, from the proof of Proposition 2 it is clear that the direct effect of a change in the covariance for each firm is $V''_{is_iCov} < 0$ and $V''_{js_jCov} < 0$. The proposition follows from the second order conditions $V''_{is_is_i} < 0$ and $V''_{js_js_j} < 0$, and the stability condition *DET*> *0*. Q.E.D.

Appendix 4

This also follows directly from the differentiation of the first order conditions, where the sign of ds_i^* / df_i will be that of $V''_{is_if_i}$. By (A1) fixed costs of firm *i* enters only in its risk aversion term, and if the absolute risk aversion is decreasing ($\gamma < 1$) then $R'_{if_i} > 0$. The derivative is

$$V''_{is_if_i} = -R'_{if_i}\left[\pi_i\left[s_i, s_j\right], f_i\right]Cov(\pi_i\left[s_i, s_j\right], \pi'_{is_i}\left[s_i, s_j\right]).$$

If the covarianec is positive then $ds_i^* / df_i < 0$. The effect on the other firm depends on the if there are strategic substitutes or complements. The argument is symmetric for negative covariance. Q.E.D.

FIGURE 1 A.

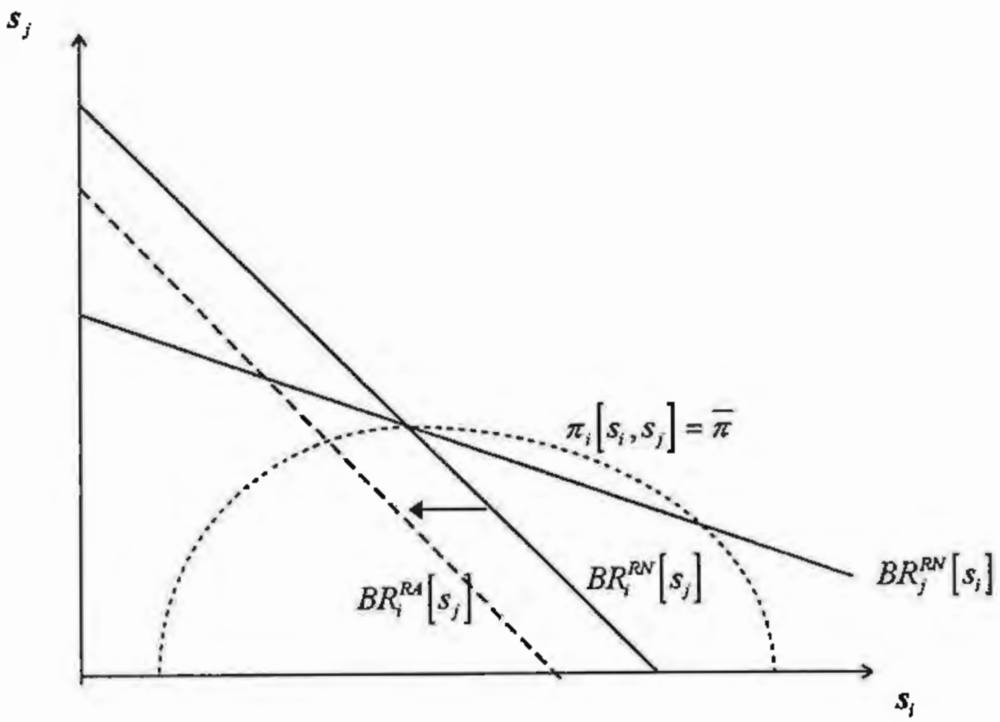


FIGURE 1 B.

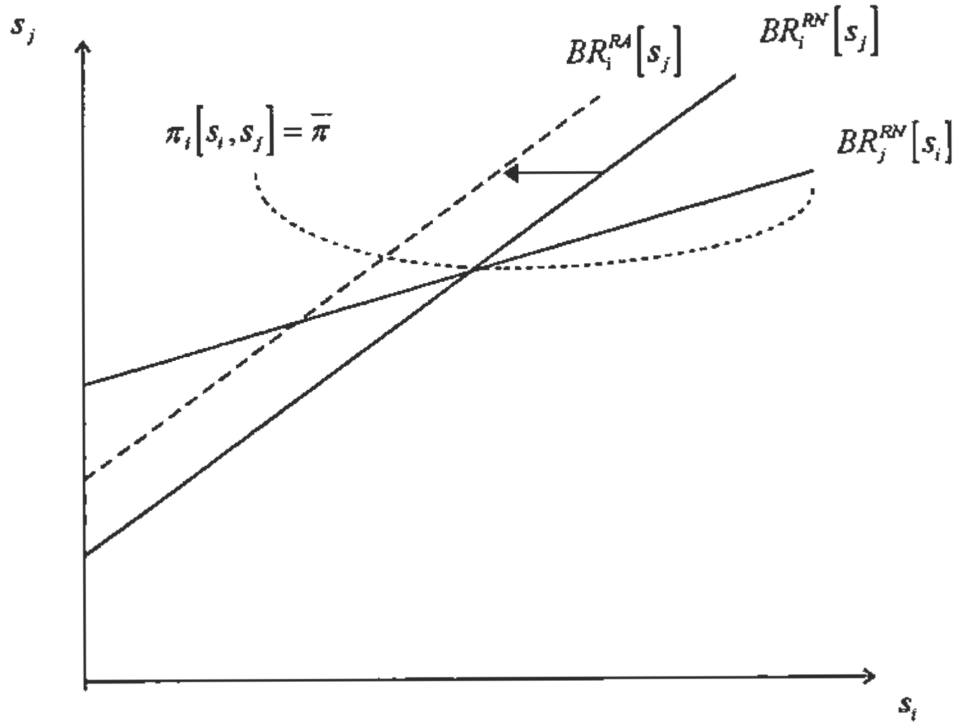


FIGURE 2 A.

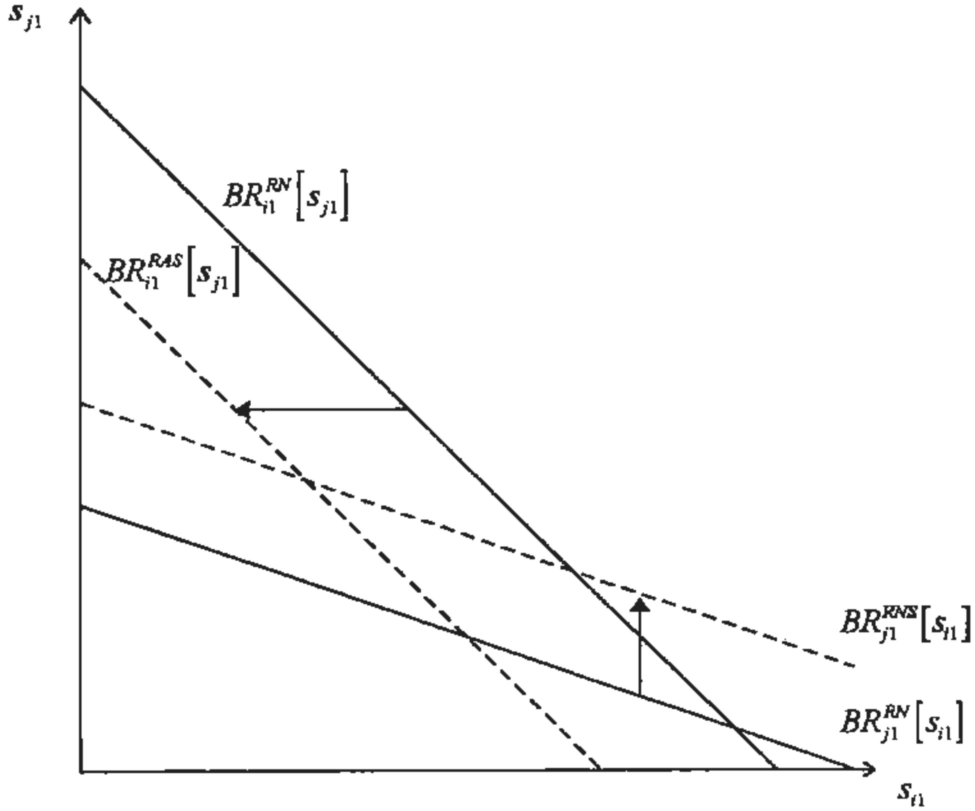
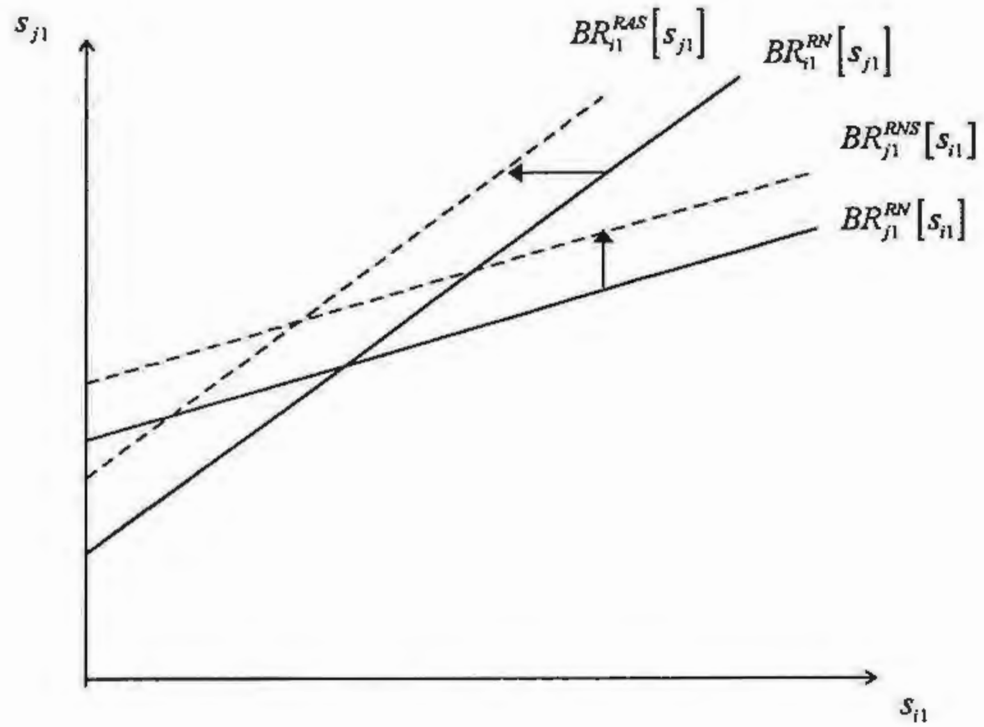

s_{j1}
$BR_{i1}^{RN}\left[s_{j1}\right]$
$BR_{i1}^{RAS}\left[s_{j1}\right]$
$BR_{j1}^{RNS}\left[s_{i1}\right]$
$BR_{j1}^{RN}\left[s_{i1}\right]$
s_{i1}

FIGURE 2 B.



Oligopoly and Limited Liability*

Marcus Asplund

Department of Economics

Stockholm School of Economics

Box 6501, S-113 83 Stockholm

e-mail: nema@hhs.se

Abstract

This paper studies the strategic interaction on oligopolistic markets where firms have debt obligations. If competitors set sufficiently high (low) quantities (prices) then there exists no unique strategy that maximise equity holders payoff, since whatever quantity (price) an indebted firm sets, operating profits will not cover debt. The result is an infinite number of weak, and not necessarily any strict, Nash equilibrium. However, many of these involve weakly dominated strategies. For low debt levels, there exist a unique strict N.E. , which is the only strategy to survive iterated elimination of weakly dominated strategies. For high debt levels, it is only possible to give upper and lower bounds of the surviving strategies. Generally, the bounds of the surviving strategies (prices or quantities) are increasing in debt levels. The analysis substantially generalises earlier work and provides some new insights into the relation between financial structure and product market behaviour.

Keywords: Oligopoly; limited liability; weakly dominated strategies; financial structure.

JEL specification: D43; D81; G32; G33; L13; L21.

---

*I would like to thank Tore Ellingsen, and participants in the Stockholm School of Economics workshops in Corporate finance, and Institutional economics and economic theory, for many helpful comments and suggestions. Financial support from the Swedish Competition Authority is gratefully acknowledged.

1. Introduction

Most firms are financed by a combination of debt and equity. Equity holders are in control and can choose strategies as long as they meet debt obligations. If they fail to repay due debt, the creditors can take control and liquidate the firm, and equity holders claims become worthless. Equity holders are residual claimants when profits are greater than debt obligations, while they get zero payoff when profits are insufficient to cover debt. This well known limited liability effect increases the incentives of equity holders to adopt risky strategies (see e.g. Jensen and Meckling (1976)).[1]

The limited liability effect is studied in an oligopoly context in Brander and Lewis (1986). BL compare the choice of strategy in an oligopoly game (Cournot) played under uncertainty, between firms which maximise firm value, i.e. the value of equity *and* debt, and firms which maximise only equity value. The Nash equilibrium under limited liability typically involve higher quantities than the game with unlimited liability.[2] The general point is that debt shift the best response function, since limited liability permits the firm to ignore the bad realisations where low quantities would have been optimal.[3]

Our results generalise BL's and offer additional insights into why and when financial structure matters for product market strategies. The key insight is that debt restricts the set of strategies for which equity holders get a strictly positive payoff and therefore the range over which best response functions are uniquely defined. Debt serves as a commitment to high quantities (or prices), since the opposite inevitably lead to bankruptcy, whatever the competitors do. This

[1] Equity is in a sense a call option on the firm, and it is a standard result in option pricing theory that the value of a call option is increasing in the risk of the underlying asset.

[2] Glazer (1994) extends the insights to a model with two periods of competition and shows that firms will set lower quantities the first period which increase the competitors profit to make them less aggressive in the second period. McAndrews and Nakamura (1992) study debt as a commitment to aggressive strategies to deter entry.

[3] Brander and Lewis (1988) show that fixed bankruptcy cost also shift the best response function, even when firms maximise firm value. In another setting Maksimovic (1990) show that firms best response function shift out, if they can obtain a loan commitment prior to competition. A loan commitment is a strategic investment where a fixed fee is paid in return for a commitment by a creditor to lend money at rates below market rates.

effect is very general and does not rest upon the existence of uncertainty and the possibility to gamble on the good realisations of uncertainty.

With debt financing and no uncertainty, there exists an infinite number of weak Nash equilibria where the indebted firm gets zero payoff, and not necessarily any strict Nash equilibrium. To narrow down the set of equilibria, we use iterated elimination of weakly dominated strategies, which greatly restricts the set.[4] For low debt levels there exists a unique strict Nash equilibrium (the Nash equilibrium under unlimited liability), and it is the only to survive elimination. For higher debt levels, there is no strict Nash equilibrium, but it is proved that the surviving strategies (quantities or prices) are increasing in debt. For very high debt levels elimination of weakly dominated strategies can not be applied. The model is generalised to allow for uncertainty. Again there may exist many weak Nash equilibria, but for very low levels of debt the only surviving strategies is the strict Nash equilibrium with unlimited liability (i.e. very low debt levels are irrelevant for the equilibrium). For an intermediate range of debt levels only the strict Nash equilibrium of BL survives elimination. At higher levels it is possible to give lower and upper bounds on the surviving strategies, but for very high debt levels elimination of weakly dominated strategies can not be used.

The structure of the paper is as follows. Section 2 describes the oligopoly model (consistent with both strategic complements and substitutes). It also defines the concept of iterated elimination of weakly dominated strategies (IEWDS), and provides a simple example. Section 3 states the results in the special case of no uncertainty and only one debt financed firm. The most important result is that when the Nash equilibrium profit is less than debt obligations, the indebted firm goes bankrupt in all strategies that survive IEWDS. The bounds of the surviving strategies are characterised. The following section contains a numerical example of

[4]Elimination of dominated strategies have previously been applied in oligopoly contexts. It has been shown by Moulin (1984) that in the 2-firm Cournot game (without limited liability), the Cournot-Nash equilibrium is the only to survive iterated elimination of strictly dominated strategies. Recently, Börgers (1992) has applied IEWDS to a Bertrand model with capacity constraints, to study the surviving (discrete) pricing strategies. His interest is under what conditions the surviving prices are close to Walrasian prices. For a general discussion of this equilibrium concept see Fudenberg and Tirole (1991 p. 460-4).

a Cournot model. Section 5 contains some generalisations. First, the case where both firms are debt financed. Second, it is shown that IEWDS can be applied even when there is uncertainty. A third extension is to a two stage model, where the debt is determined endogenously. Section 6 summarises and concludes the paper.

2. Model

Let s_i denote the strategies of Firm i, which have non negative support, $s_i \in S_i = [0,\infty]$. Firm i's equity holders receive a strictly positive payoff if its operating profits, $\pi_i\left[s_i, s_j\right]$, are strictly greater than its outstanding debt, D_i, otherwise they get zero payoff. In the duopoly considered here, equity holders in each firm simultaneously choose a strategy to maximise their respective payoff $U_i\left[\pi_i\left[s_i, s_j\right], D_i\right] = \max\left[0, \pi_i\left[s_i, s_j\right] - D_i\right]$, where $i, j = 1,2, i \neq j$. It is assumed that creditors can not influence the strategy of the firm.[5] The best response correspondences are $R_i\left[s_j\right] = \arg\max\left[\max\left[0, \pi_i\left[s_i, s_j\right] - D_i\right]\right]$. It is important to note that the best response is not unique for all s_j. For strategic substitutes, high values of s_j will cause Firm i's profit to be less than D_i, whatever s_i it chooses. For strategic complements, if Firm j selects a low s_j, there is no unique best response of Firm i, since it is bankrupt whatever strategy it uses. The lowest (highest) s_j where the best response is unique for strategic substitutes (strategic complements) is defined by max $\pi_i\left[s_i, s_j^{DD}\right] = D_i$ for $s_i \in S_i$. Define $s_i^D = R_i\left[s_j^{DD}\right]$, the payoff maximising strategy which yields profits equal to the outstanding debt. In addition, define s_i^M as the monopoly quantity (best response to an infinitely high price of j) with strategic substitutes (strategic complements),

$$s_i^M \equiv \begin{cases} R_i[0] & if \quad R_i'\left[s_j\right] < 0 \\ R_i[\infty] & if \quad R_i'\left[s_j\right] > 0, \end{cases}$$

[5]This can be motivated by asymmetric information between equity holders and creditors, or where there are many creditors which only hold small claims and renegotiation is costly. Thus we do not allow equity holders to renegotiate the debt prior to the choice of strategy.

which yield profit $\pi_i\left[s_i^M\right]$. The formulation of strategies is consistent with either a Cournot-type game where s_i are quantities, or a Bertrand-type game where they are prices. Informally, we term s_i quantities when best response functions are downward sloping (strategic substitutes) $R'_{is_j}\left[s_j\right]<0$, and prices when best response functions are upward sloping (strategic complements) $R'_{is_j}\left[s_j\right]>0$. The sign of the slope of the best response functions are assumed to be the same for both firms. Our results are robust to both cases, conditional on the assumptions stated below.

To ensure a well behaved oligopoly game it is assumed that there exist an unique, strict Nash equilibrium in the game where players maximise their respective firm values, which is equivalent to maximising profits, in the paper informally denoted "the Nash equilibrium ". Let s_1^*, s_2^* denote this Nash equilibrium, characterised by $s_1^* = R_2\left[s_1^*\right], s_2^* = R_1\left[s_2^*\right]$. The support of the best response functions is bounded, $0 \le R_i\left[s_j\right] < \infty$, and best response functions are continuous and monotone for $\forall s_j \in S_j$.

Definition: Iterated elimination of weakly dominated pure strategies (IEWDS) is done in the following steps: The set of surviving strategies for Firm i after n rounds of elimination is denoted S_i^n, and let $S_i^0 = S_i$. Define recursively the strategies for i which are not weakly dominated. A strategy in S_i^{n-1} is weakly dominated if there exist some strategy also in S_i^{n-1}, which have at least as high payoff for all j's strategies in S_j^{n-1}, and strictly higher payoff for at least one of them.

$$S_i^n = \left\{ s_i \in S_i^{n-1} \middle| \nexists s_i' \in S_i^{n-1} \quad \text{such that} \quad \begin{matrix} U_i\left[s_i', s_j, D_i\right] \ge U_i\left[s_i, s_j, D_i\right] & \forall s_j \in S_j^{n-1} \\ U_i\left[s_i', s_j, D_i\right] > U_i\left[s_i, s_j, D_i\right] & \exists s_j \in S_j^{n-1} \end{matrix} \right\}.$$

The surviving pure strategies are defined as

$$S_i^\infty = \bigcap_{n=0} S_i^n .$$

In words, in a round of iteration all weakly dominated strategies are eliminated. In the next round, we eliminate from the new, smaller set of remaining strategies of one player, all which are weakly dominated given the new, smaller set of strategies for other player. This iteration continues until it converges, where the pure strategies for i that survives IEWDS is S_i^∞. Note that only pure strategies weakly dominated by another pure strategy are considered. It is well known that even if a pure strategy is not dominated by another pure strategy, it may still be dominated by a mixed strategy. However, for the purpose of this paper we can get the interesting results even with this weak condition. Below it is shown that no surviving pure strategy is weakly dominated by any mixed strategy. Moreover, which strategies that survive IEWDS may be dependent upon the assumed order of iteration, (see e.g. Fudenberg and Tirole (1991 p 460-4)).[6] In the problem we analyse the surviving strategies could be different if we assumed some other order of iteration, this is discussed after Theorem 2.

To illustrate the ideas we consider the following simple example: Firm 1 and Firm 2 play a Cournot game under no uncertainty. Firm 1 is run by equity holders and has outstanding debt, D_1, whereas Firm 2 has no debt. The assumption is made that the Cournot-Nash equilibrium profit of Firm 1, in the game where firms maximise their respective firm value (profits), $\pi_1\left[s_1^*, s_2^*\right]$, is less than the debt obligations. However, there exists some quantities of Firm 1 and Firm 2 that would yield profits strictly greater than the debt. In Figure 1 the dashed segment is the profit maximising best response of Firm 1, where its profit is less than the debt.

FIGURE 1 ABOUT HERE

If Firm 1 produces s_1^* then the strictly best response of firm 2 is s_2^*. Similarly, if Firm 2 produces s_2^* then Firm 1 can not do strictly better than by setting s_1^*, since its payoff is zero for all strategies. Thus s_1^*, s_2^* is a weak (pure strategy) Nash equilibrium, but this equilibrium is no longer unique. Given that Firm 1 set $s_1 = 0$, Firm 2 can of course do no better than by

[6] E.g. if first all weakly dominated strategies of Player 1 are eliminated, and thereafter the strategies which are weakly dominated for Player 2, given that some of Player 1's strategies have already been eliminated.

setting s_2^M, its monopoly quantity. Given that Firm 2 produce s_2^M , Firm 1 can not do strictly better than it does by producing $s_1 = 0$. This is another weak Nash equilibrium, of which there exist an infinite number. However, no strict Nash equilibrium exist.[7] Nash equilibrium is not a good equilibrium concept in this particular game, in the sense that it does not exclude many "unreasonable" strategies. The problem is that weakly dominated strategies are not eliminated. The weak Nash equilibrium $s_1 = 0$ and s_2^M, involves a weakly dominated strategy for Firm 1. For all of Firm 2's strategies, Firm 1's payoff is at least as high if it sets s_1^D. Moreover, if Firm 2 would produce a quantity strictly lower than s_2^{DD}, the payoff for Firm 1 is strictly positive, whereas for $s_1 = 0$ it is zero. Thus $s_1 = 0$, and every strategy below s_1^D from the concavity of the profit function, is weakly dominated by s_1^D. Then if Firm 1 never sets quantities lower than s_1^D, all quantities for Firm 2 greater than $R_2\left[s_1^D\right]$ are weakly dominated. Moreover, all strategies for Firm 1 with higher quantity than s_1^M (the monopoly quantity), are of course weakly dominated. When Firm 1 never sets quantities greater than s_1^M, all quantities lower than $R_2\left[s_1^M\right]$ are weakly dominated for Firm 2. In this example it is not possible to eliminate any additional strategies, since the payoff for Firm 1 is zero when Firm 2 never sets quantities lower than $R_2\left[s_1^M\right]$. Firm 1's surviving quantities are between s_1^D and s_1^M, and Firm 2's between $R_2\left[s_1^M\right]$ and $R_2\left[s_1^D\right]$. The shaded area in Figure 1 show the surviving strategies. (In general the elimination process will continue until it converge, i.e. when there are no additional strategies that are weakly dominated).

3.Results

In this section it is assumed that only Firm 1 is debt financed. The payoffs are $U_1\left[s_1, s_2, D_1\right] = \max\left[0, \pi_1\left[s_1, s_2\right] - D_1\right]$ and $U_2\left[s_2, s_1, D_2\right] = U_2\left[s_2, s_1, 0\right] = \pi_2\left[s_2, s_1\right]$. It is assumed that Firm 1's profit in the Nash equilibrium may be insufficient to cover the debt

[7]Note that even if Firm 1's debt is less than the Nash equilibrium profit, in the game where firms maximise firm value, there would still exist an infinite number of weak Nash equilibria, but also a strict Nash equilibrium, namely the Cournot-Nash equilibrium.

payments $\pi_1\left[s_1^*, s_2^*\right] < D_1$, but that for at least one combination of strategies the profit is greater than debt obligations, such that the debt is bounded by $\pi_1\left[s_1^M\right] > D_1$.

We first prove that Firm 2 will always choose a best response to one of the strategies in S_1^∞.

Lemma 1: If $S_1^\infty = \left[\underline{s}_1^\infty, \overline{s}_1^\infty\right]$ and $R_{is_j}'\left[s_j\right] < 0$ then $S_2^\infty = \left[R_2\left[\overline{s}_1^\infty\right], R_2\left[\underline{s}_1^\infty\right]\right]$.
If $S_1^\infty = \left[\underline{s}_1^\infty, \overline{s}_1^\infty\right]$ and $R_{is_j}'\left[s_j\right] > 0$ then $S_2^\infty = \left[R_2\left[\underline{s}_1^\infty\right], R_2\left[\overline{s}_1^\infty\right]\right]$.

Proof: An informal proof for the strategic substitutes case $R_{is_j}'\left[s_j\right] < 0$ is as follows. Firm 2's payoff is equal to its profit. Since Firm 1 never sets quantities lower than $\underline{s}_1^\infty$, all quantities greater than $R_2\left[\underline{s}_1^\infty\right]$ give Firm 2 lower profit than $R_2\left[\underline{s}_1^\infty\right]$, whatever quantity Firm 1 produce. If Firm 1 would set $\underline{s}_1^\infty$, Firm 2's profit can be no higher than if it uses $R_2\left[\underline{s}_1^\infty\right]$. Therefore $R_2\left[\underline{s}_1^\infty\right]$ can not be weakly dominated as an upper bound of the surviving strategies of Firm 2. The proof for the lower bound is similar. For strategic complements, the argument is symmetric. Q.E.D.

The interesting property of Lemma 1 is that the iterative process must yield surviving strategies which are spanned by the best response function of the firm without debt. Using Lemma 1 we can immediately prove that the payoff for Firm 1 will be zero in all strategies that survive the IEWDS, if the debt is greater than the Nash equilibrium profit.

Theorem 1: If $\pi_1\left[s_1^*, s_2^*\right] < D_1$ and $D_2 = 0$ then $U_1\left[s_1, s_2, D_1\right] = 0$ for $\forall s_1 \in S_1^\infty, \forall s_2 \in S_2^\infty$.

Proof: The theorem is proved by contradiction, here in the strategic substitutes case $R_{is_j}'\left[s_j\right] < 0$. If there is any combination of strategies that yields strictly positive payoff for Firm 1, it is definitely positive at the lower bound $\underline{s}_2^\infty$ of Firm 2's surviving quantities (since Firm 1's profit is decreasing in s_2). All Firm 1's quantities which are greater than $R_1\left[\underline{s}_2^\infty\right]$ give strictly lower profits (and therefore never strictly higher payoff) for all of Firm 2's surviving quantities. Moreover, $R_1\left[\underline{s}_2^\infty\right]$ is not weakly dominated by some lower quantity, since it gives maximum payoff if Firm 2 uses its lowest quantity $\underline{s}_2^{\infty}$. Thus the surviving upper bound is $R_1\left[\underline{s}_2^{\infty}\right]$. However, from Lemma 1 the surviving lowest quantity for Firm 2 is $\underline{s}_2^{\infty} = R_2\left[\overline{s}_1^{\infty}\right]$ and therefore $\underline{s}_2^{\infty} = R_2\left[\overline{s}_1^{\infty}\right] = R_2\left[R_1\left[R_2\left[\overline{s}_1^{\infty}\right]\right]\right]$ but the only quantities where this hold is at the Nash equilibrium s_1^*, s_2^*. Since by assumption the Nash equilibrium profits are insufficient to cover debt payments we have established the contradiction. The argument for strategic complements is similar, except that we study the surviving upper bound for Firm 2. If the payoff is strictly positive anywhere, it must definitely be so at $\overline{s}_2^{\infty} = R_2\left[\overline{s}_1^{\infty}\right]$. Q.E.D.

The intuition is simple. As long as the indebted firm's profit is greater than its debt for some remaining strategies, it has an incentive to eliminate the strategies which are best responses to strategies that the other firm never uses. This is understood by the other firm which in turn never uses any strategies that are best responses to strategies that the indebted firm never use. This stepwise movement towards the Nash equilibrium does not stop until the indebted firm is bankrupt, for all remaining strategies of the firm without debt. Even though the indebted firm could have been solvent for some combination of strategies, the incentive to eliminate weakly dominated strategies makes it bankrupt in all surviving strategies.[8]

For high debt levels, the indebted firm's profit is always less than its debt in the surviving strategies. But then, it can be argued, it is totally irrelevant if it produces at all or uses the surviving strategies. There are at least two answers to this. First, equity holders may continue because of some, non modelled, private benefit of control. Second, at least equity holders do not get strictly lower payoff from continuing and use the surviving strategies, and using some other strategy definitely involves weakly dominated strategies, which is hardly a rational thing to do.

[8]The payoff for Firm 1 is also zero in any mixed strategy. If $\left\{\Sigma_1^{\infty}, \Sigma_2^{\infty}\right\}$ is the set of mixed strategies that survives IEWDS, then payoff for Firm 1 is $U_1\left[\sigma_1, \sigma_1, D_1\right] = 0$ for $\sigma_1 \in \Sigma_1^{\infty}, \sigma_2 \in \Sigma_2^{\infty}$. Moreover, we can not eliminate any additional pure strategies, as weakly dominated by some mixed strategy. Since Firm 1 get zero payoff in all surviving strategies, it is indifferent to all mixed strategies, such that no mixed strategy can weakly dominate any other (in particular no pure strategy). Any of Firm 2's surviving strategies are a strictly best responses to one pure strategy of Firm 1, of which we could not eliminate any by a mixed strategy.

Having stated this property of the surviving strategies, we continue the characterisation of the upper and lower bounds of these strategies. In Theorem 2 we characterise the bounds of the strategies which are not eliminated in the first round, and the bounds of the indebted firms strategies that survive all rounds of iteration. The general idea is that each firm wants to be on its best response function, given the remaining strategies of the other firm. No firm wants to use strategies which are best responses to strategies which the other firm never uses. This will cause a "stepwise" movement towards the Nash equilibrium which will end either 1) at the Nash equilibrium, or 2) when the indebted firm is bankrupt in all remaining strategies.

Theorem 2:

1) If $\pi_1\left[s_1^*, s_2^*\right] \geq D_1$ and $D_2 = 0$ then $S_1^1 = \left[s_1^D, s_1^M\right]$ and $S_1^\infty = \left[s_1^*\right]$.

2) If $\pi_1\left[s_1^*, s_2^*\right] < D_1$ and $D_2 = 0$ then $S_1^1 = \left[s_1^D, s_1^M\right]$ and $S_1^\infty = \left[s_1^D, R_1\left[\underline{s}_2^n\right]\right]$, where $\underline{s}_2^n < s_2^{DD} \leq R_2\left[R_1\left[\underline{s}_2^n\right]\right]$.

Proof: First consider the case of strategic substitutes. It is obvious that Firm i will never set quantities higher than its monopoly quantity $s_i^M = R_i\left[0\right]$, which would give lower profit whatever quantity Firm j would set. s_i^M is not weakly dominated by some lower quantity, since it gives maximum payoff if Firm j produce zero (which is the lower bound of S_j^0). Therefore the upper bound is s_i^M after the first round. The lower bound of Firm 2's strategies is the best response to an infinite quantity by Firm 1, (which is the upper bound of S_1^0) $\underline{s}_2^1 = R_2\left[\infty\right] = 0$. The profit of Firm 1 is always less than its debt if Firm 2 sets a higher quantity than s_2^{DD}. However, if Firm 2 would set a quantity lower than s_2^{DD}, the payoff is strictly greater at s_1^D than for any lower quantity, by the concavity of the profit function. Moreover, no higher quantity can weakly dominate s_1^D since there is always some quantity just below s_2^{DD} where the profit is higher (and payoff strictly positive) at s_1^D.

Next there are two cases. 1) $R_2\left[s_1^D\right] \leq s_2^{DD}$. In the second round of elimination, s_1^D will be weakly dominated by $R_1\left[R_2\left[s_1^D\right]\right]$, which is unique and strictly higher. This is because s_1^D is a best response to higher quantities than $R_2\left[s_1^D\right]$, which Firm 2 never uses. Continued iteration then gives the new upper bound of Firm 2's strategies $R_2\left[R_1\left[R_2\left[s_1^D\right]\right]\right]$ and so forth. The elimination continues, and the lower bound converge to the strict Nash equilibrium s_1^*. This is similar to the application of elimination of strictly dominated strategies in a Cournot model, which converge to the unique Nash equilibrium, Moulin (1984) and Fudenberg and Tirole (1991). 2) $R_2\left[s_1^D\right] > s_2^{DD}$. In the second round of elimination s_1^D will not be weakly dominated by any higher quantity. By the argument above, s_1^D gives strictly higher payoff for some s_2 just below s_2^{DD}.(Conditional on there being strategies below s_2^{DD}. If it does not, then the payoff is zero in all remaining strategies and no strategy can weakly dominate s_1^D.) The situation will be the same in all future rounds of elimination. This implies that s_1^D is the surviving lower bound.

Assume that after $n \geq 0$ rounds of elimination $\underline{s}_2^n$ is the lower bound of Firm 2's remaining strategies and $\pi_1\left[R_1\left[\underline{s}_2^n\right], \underline{s}_2^n\right] > D_1$. All strategies greater than $R_1\left[\underline{s}_2^n\right]$ will be weakly dominated, and it can not be weakly dominated by some lower quantity, since it gives maximum payoff for $\underline{s}_2^n$. In the following round the lower bound of Firm 2's quantities will be $R_2\left[R_1\left[\underline{s}_2^n\right]\right]$ since it will not set quantities which are not best responses to quantities which are never used by Firm 1. As before there are two cases. 1) $R_2\left[R_1\left[\underline{s}_2^n\right]\right] < s_2^{DD}$. $R_1\left[\underline{s}_2^n\right]$ will be weakly dominated by the lower quantity $R_1\left[R_2\left[R_1\left[\underline{s}_2^n\right]\right]\right]$ and the elimination process continues. 2) $R_2\left[R_1\left[\underline{s}_2^n\right]\right] \geq s_2^{DD}$. There are no quantity that weakly dominates $R_1\left[\underline{s}_2^n\right]$, since the payoff is zero for all remaining strategies. Therefore the surviving upper bound of Firm 1's strategies is $\bar{s}_1^* = R_1\left[\underline{s}_2^n\right]$.

The argument for strategic complements is symmetric. In the first round of elimination all strategies greater than $R_i\left[\infty\right]$ (which is bounded by assumption) are weakly dominated. For Firm 2, the lower bound is $R_2\left[0\right]$. For Firm 1 all strategies lower than s_1^D (where the best response function is not unique) are weakly dominated. The rest of the proof is identical and omitted. Q.E.D.

Theorem 2 shows that for low levels of debt ($\pi_1\left[s_1^*, s_2^*\right] \geq D_1$) the only surviving strategy is the strict Nash equilibrium. Low debt is irrelevant for the equilibrium. For higher debt levels there exists a set of surviving strategies, for which it is possible to give upper and lower bounds.[9]

Theorem 2 is illustrated in Figure 2a (strategic substitutes) and Figure 2b (strategic complements). The dashed segment is the profit maximising best response, where the profit is less than debt obligations. The shaded areas are the surviving strategies.

FIGURES 2A AND 2B ABOUT HERE

Lemma 1 gives the surviving strategies for Firm 2.

Corollary 1:

1) If $\pi_1\left[s_1^*, s_2^*\right] \geq D_1$ and $D_2 = 0$ then $S_2^\infty = \left[s_2^*\right]$.

2) If $\pi_1\left[s_1^*, s_2^*\right] < D_1$ and $D_2 = 0$ and $R_{is_j}'\left[s_j\right] < 0$ ($R_{is_j}'\left[s_j\right] > 0$) then $S_2^\infty = \left[R_2\left[R_1\left[\underline{s}_2^n\right]\right], R_2\left[s_1^D\right]\right]$ ($S_2^\infty = \left[R_2\left[s_1^D\right], R_2\left[R_1\left[\overline{s}_2^n\right]\right]\right]$)

Since $s_1^* < s_1^D = \underline{s}_1^\infty$ the Nash strategy is not among the surviving strategies for Firm 1 (and by Lemma 1, the Nash strategy for Firm 2 is not among the surviving strategies), for high debt levels.

Corollary 2:

1) If $\pi_1\left[s_1^*, s_2^*\right] \geq D_1$ and $D_2 = 0$ then $s_1^* = \underline{s}_1^\infty$

2) If $\pi_1\left[s_1^*, s_2^*\right] < D_1$ and $D_2 = 0$ then $s_1^* < s_1^D = \underline{s}_1^\infty$

[9] In the paper, simultaneous elimination of dominated strategies are assumed. As noted above, this can influence the set of surviving strategies, and this is so for Theorem 2. For strategic complements, assuming sequential elimination and beginning with Firm 2, it is not necessarily possible to eliminate any strategies of Firm 1. When the highest prices of Firm 2 are already eliminated, none of the remaining may be sufficiently high to give Firm 1 a strictly positive payoff. Thus, all of Firm 1's strategies survive elimination. If the iteration starts with Firm 1 then Theorem 2 still holds. An alternative to assuming any arbitrarily order, is to study the strategies which survive all orders of elimination (which s_1^D does), but that is beyond the scope of this paper.

An interesting corollary is the relation between debt and the lower bound on the strategies Firm 1 can rationally use (according to weak dominance criteria). As noted above, the profits are increasing to the right on Firm 1's best response function, and more debt will cause s_1^D to increase. The upper bound will take discrete values, because over some range of debt the iteration converge in the same round.

Corollary 3: If $\pi_1\left[s_1^*, s_2^*\right] < D_1$ and $D_2 = 0$ then:

1) s_1^D is strictly increasing in D_1.

2) $\overline{s_1}^{\infty}$ is weakly increasing in D_1.

Corollary 3 can be compared to with BL's analysis of a stochastic Cournot duopoly. Their main result is that debt will cause firms to behave more aggressively and set higher quantities. In their paper, the results for the game with strategic complements are not explicitly stated. However, it can be showed that debt cause firms to set higher prices.[10] The intuition is that debt shifts the best response function to the right, when equity holders can ignore the low realisations of uncertainty, where low quantities (or prices) would have been optimal.[11] Theorem 2 and Corollary 3 gives a similar result, even when there is no uncertainty. Debt limits the strategies the equity holders can get a positive payoff from. Low quantities (or prices) will not yield sufficiently high profit to cover debt payments, whatever strategy the competitor use.

The smallest possible set of surviving strategies, for high debt levels, is arbitrarily close to a singleton.[12]

---

[10] E.g. differentiated products with linear demand functions, and uniform distribution of the intercept of the demand function.

[11] Brander and Lewis are somewhat misleading in this game.
"While this paper focuses nearly exclusively on the Cournot doupoly market structure, the first central insight we offer applies quite generally. This central insight is that higher debt levels tend, in the standard case ($R_{iz}^i > 0$), to increase a firm's desired *output*." p. 961, italics added. $R_{iz}^i > 0$ refers to a positive covariance between profits and marginal profits, wbich is likely to hold even with strategic complements, under demand uncertainty. This point has also been noted by Showalter (1995).

[12] Note that the largest possible surviving strategies (in the case $\pi_1\left[s_1^*, s_2^*\right] < D_1$ and $D_2 = 0$), is when no additional strategies are eliminated after the first round and $S_1^n = \left[s_1^D, s_1^M\right]$ (for $n = 1,..,\infty$).

Corollary 4: If $\pi_1\left[s_1^*, s_2^*\right] < D_1$ and $D_2 = 0$ then the smallest surviving strategy space is $S_1^{\infty} = \left[s_1^D, s_1^D + \varepsilon\right]$, where $\varepsilon \to 0^+$.

Define $\hat{s}_1$ from $\pi_1\left[\hat{s}_1, R_2\left[\hat{s}_1\right]\right] = \pi_1\left[s_1^*, s_2^*\right]$. This is the intersection between the isoprofit function for the Nash equilibrium profit for Firm 1, and the best response function of Firm 2. Even though Firm 1 can never avoid bankruptcy when $\pi_1\left[s_1^*, s_2^*\right] < D_1$ the creditors loss is affected. Define the loss of the creditors as $L_1^{\infty} = D_1 - \pi_1\left[s_1, s_2\right]$ for $s_1 \in S_1^{\infty}, s_2 \in S_2^{\infty}$. This is to be compared with their loss in the Nash equilibrium $L_1^* = D_1 - \pi_1\left[s_1^*, s_2^*\right]$. The loss of the creditors is lower (higher) in the surviving strategies than in the Nash equilibrium, if the upper (lower) bound of the surviving strategies is lower than $\hat{s}_1$.

Corollary 5: If $\pi_1\left[s_1^*, s_2^*\right] < D_1$ and $D_2 = 0$:

1) If $\hat{s}_1 > s_1^{\infty}$ then $L_1^{\infty} < L_1^*$.

2) If $\hat{s}_1 < s_1^{\infty}$ then $L_1^{\infty} > L_1^*$.

For some range of surviving strategies the difference between the loss in the Nash equilibrium and the surviving strategies is indeterminate. Corollary 5 shows that it need not be a conflict of interest between creditors and equity holders. Creditors understand that they will not get all their money back, but since equity holders have a strong incentive to set high quantities (or prices), this gives higher profits which reduce the loss. Below the incentive to renegotiate debt is discussed.

An assumption was that there existed some combination of strategies where the indebted firm could be solvent. If this is not the case, weak dominance criteria can not be used, since the payoff is always zero and it is not possible to eliminate any strategies for the indebted firm. More specifically, if the debt is greater than the monopoly profit $\pi_1\left[s_1^M\right] < D_1$ no upper and lower bounds can be given for the indebted firm.

4. Numerical example

To illustrate the surviving strategies in a numerical example we study a standard Cournot game.

Assume a linear demand function

$$P = \left(1 - \left(s_i + s_j\right)\right),$$

which yields the profit function for firm i

$$\pi_i = s_i \left(1 - \left(s_i + s_j\right)\right).$$

The Cournot-Nash equilibrium is $s_i^* = 1/3$ with profit $\pi_i^* = 1/9$. The monopoly quantities are $s_i^M = 1/2$ with profit $\pi_i^M = 1/4$.

In the example we look at cases where the debt is between Cournot-Nash profit and the monopoly profit. Less debt result in the Nash equilibrium. Figure 3 show the surviving upper and lower bounds for Firm 1's strategies, as functions of the debt level. The range of strategies between the bounds are the surviving strategies. The lower bound is strictly increasing in debt, while the upper bound takes discrete values depending on the number of iterations required for convergence, which is Corollary 3. When the outstanding debt is equal to the Nash equilibrium profit, the upper and lower bound converge to the Nash equilibrium quantity, which for higher debt levels is not within the surviving strategies, from Corollary 2.

FIGURE 3 ABOUT HERE

5.Extensions

5.1 Both firms debt financed

Suppose now that both firms are debt financed. The additional results follow almost immediately from the fact that all strategies lower than s_i^D and higher than s_i^M are weakly dominated in the first round of elimination. There will be different cases depending on whether the debt of Firm 2 is greater or less than its Nash equilibrium profit. Begin by studying the case of strategic substitutes.

If $s_1^D > s_1^*$ and $s_2^D < s_2^*$ (i.e. Firm 1's (Firm 2's) debt is greater (less) than its Nash equilibrium profit), the surviving strategies can have the property that both firms go bankrupt. It is easy to verify that Firm 1 will be bankrupt in all surviving strategies, and if the lower bound of its surviving strategies is high enough it will also reduce the other firms profit enough to drive it into bankruptcy. When $s_1^D > s_1^*$ and $s_2^D \geq s_2^*$, both firms will be bankrupt in the surviving strategies. After the first round of elimination the remaining strategies will be greater than (or equal to) s_i^D, which give both firms zero payoffs in the remaining (the surviving) strategies.

For strategic complements the results are different. If $s_1^D > s_1^*$ and $s_2^D < s_2^*$ Firm 2 (Firm 1) will be solvent (bankrupt) in all surviving strategies. Firm 1 never sets prices below s_1^D so the profits of Firm 2 will always be at least as great as the Nash equilibrium profits. As long as Firm 1 is solvent for some remaining strategies, it has an incentive to move to be on its best response function. In Theorem 1 it was shown that this stepwise movement continue until Firm 1 gets zero payoff in all remaining strategies. The more interesting case is when $s_1^D > s_1^*$ and $s_2^D \geq s_2^*$ and both firms set higher prices than at the Nash equilibrium, which may save one of the firms from bankruptcy. This occur when e.g. Firm 2's highest remaining price after the first round of elimination is not high enough to give Firm 1 a strictly positive payoff, while some of the highest remaining prices of Firm 1 are sufficiently high to keep Firm 2 from bankruptcy.

5.2 Endogenous debt

In Theorem 2 it was proved that debt serves as a credible commitment to high quantities (or prices). This commitment can be exploited in a two stage game, where debt levels are selected in the first, and competition takes place in the second. The general idea is that since creditors must get at least zero net payoff, the surviving strategies must have the property that the value of equity and debt, is greater than the value of equity when there is no debt.

To illustrate with an example, take the game and parameters in Section 4. Assume that Firm 1 can select a debt level at t=0 by issuing public bonds (Firm 2 is fully equity financed), at t=1 both firms set quantities. At t=2 the firms are liquidated and due debt are paid, and equity holders receive the residual (if any).

A special feature of linear demand and quantity competition is that Stackelberg leader quantity is equal to the monopoly quantity. By issuing large enough debt, Firm 1 can essentially commit to a quantity just below the monopoly quantity and get the Stackelberg leader outcome. This is done by issuing public bonds $D_1^{opt} = \pi_1\left[s_1^M, R_2\left[s_1^M\right]\right] - \varepsilon$ and committing to pay an interest satisfying $\left(1+r_1^{opt}\right) = \pi_1\left[s_1^M, 0\right] / D_1^{opt} - z$, where $z, \varepsilon \to 0^+$.[13] The proceeds D_1^{opt} are then distributed to equity holders (for the argument ignoring legal rules of dividends) which can not be recouped by creditors at t=2. The only strategies of Firm 1 which can give it a positive payoff at t=2, are very close to s_1^M. The lower bound is arbitrarily close to s_1^M ($s_1^D \to s_1^M$). Firm 2 never set quantities higher than $R_2\left[s_1^D\right] \to R_2\left[s_1^M\right]$. In the surviving strategies the profit of Firm 1 is arbitrary close to $\pi_1\left[s_1^M, R_2\left[s_1^M\right]\right]$, which is sufficient to give creditors at least zero net payoff, since equity holders borrowed only $\pi_1\left[s_1^M, R_2\left[s_1^M\right]\right] - \varepsilon$. Equity holders get zero payoff at t=2, but the dividends at t=0 was greater than the Cournot profit. Even though the example studied the optimal debt of one firm, similar arguments can be used to study simultaneous selection of debt levels. [14]

[13] $\varepsilon = 0$ yield zero payoff for all strategies, and thus no strategy can be eliminated.

[14] There will be no symmetric pure strategy equilibrium. Consider the case where Firm 2 has low debt. Firm 1's equity holders are then better off if they, as before, take on high debt levels to commit to s_1^M. Similarly, if Firm 1 has a high debt level it is better for Firm 2 to select a low debt level. If both had high debt levels, it would (as demonstrated in the preceding section) result in both setting high quantities, and Firm 2's equity holders are worse off than without debt.

For this two stage model to work it is necessary that debt is a credible commitment. If Firm 1 declares that it has borrowed D_1^{opt} and has to pay back $D_1^{opt}(1+r_1^{opt})$ at t=2 , it must be credible to Firm 2. The reason is that if Firm 2 produce $R_2[s_1^M]$ the profit maximising strategy (which maximise the value of equity and debt) for Firm 1 is not s_1^M but the (profit maximising) best response $R_1[R_2[s_1^M]]$. This implies that there are gains to Firm 1's equity holders and creditors from renegotiating the debt repayment, if Firm 2 believes that it is $D_1^{opt}(1+r_1^{opt})$. However, Firm 2 understands this incentive to renegotiate and does not set $R_2[s_1^M]$. To make renegotiation unprofitable, Firm 1 want to have many small creditors for which it is costly renegotiate (or where free rider problems in debt restructuring are severe c.f. junk bonds), instead of one large creditor (e.g. a bank). An alternative is to have strict debt covenants which prohibits renegotiation. (For more on debt renegotiations see Gertner and Scharfstein (1991)). We ignored any direct bankruptcy costs, (e.g. legal fees) that have to be borne by the creditors at t=2. If these are large there may not be possible to find a positive D_1^{opt} where the profits minus bankruptcy costs are greater than the Nash equilibrium profit without debt. However, if bankruptcy costs are small there will be some positive D_1^{opt} where the value of debt plus equity is greater than without debt.

5.3 Uncertainty

In the previous derivations we have assumed no uncertainty, which is now relaxed. It is well known that limited liability gives an incentive to increase risk taking. Brander and Lewis (1986) have shown that debt make equity controlled firms behave aggressively by setting higher quantities, compared to maximising firm value. To be optimally prepared for the good realisations firms set high quantities (or prices). In bad realisations these strategies will give low profits leading to bankruptcy, but this is ignored by equity holders. To ensure the existence of a unique Nash equilibrium, debt can not be too large. In this section we demonstrate that IEWDS can also be applied under uncertainty to obtain the Nash equilibrium of BL. In addition IEWDS can be used at higher debt levels to give upper and lower bounds on the surviving strategies. It is pointed out that very low debt levels are irrelevant for the Nash equilibrium.

Define a weakly dominated strategy for *i* to be one where the *expected* payoff is weakly lower for all *j*'s strategies, and strictly lower for some. The definition of iterated weak dominance follow symmetrically.

Let $f[z]$ with finite support $[\underline{z}, \bar{z}]$ denote a density function (for simplicity it is assumed to be uniform) of the parameter *z*. Again assume that only Firm 1 is debt financed and that there exists at least one combination of s_1, s_2, z such that $\pi_1[s_1, s_2, z] > D_1$. For illustration consider the case of strategic substitutes. Let $\delta\pi_i / \delta z > 0$ and $\delta^2 \pi_i / \delta s_i \delta z > 0$, i.e. both profits and marginal profits are increasing in the realised *z*. Begin by implicitly defining the lowest realisation where Firm 1 is solvent, $\hat{z}_1$, by $\pi_1[s_1, s_2, \hat{z}_1] = D_1$.
Assuming that $\underline{z} < \hat{z} < \bar{z}$, BL shows that the best response function of Firm 1 is

$$R_1[s_2] = \arg\max \int_{\hat{z}_1}^{\bar{z}} \left(\pi_1[s_1, s_2, z] - D_1\right) f(z) dz.$$

Firm 1 ignore the realisations below $\hat{z}_1$, while Firm 2 maximise expected profits (over all realisations). Its best response function is given by

$$R_2[s_1] = \arg\max \int_{\underline{z}}^{\bar{z}} \pi_2[s_2, s_1, z] f(z) dz.$$

The first order conditions are

$$\int_{\hat{z}_1}^{\bar{z}} \frac{\delta\pi_1[s_1, s_2, z]}{\delta s_1} dz = 0 \text{ and } \int_{\underline{z}}^{\bar{z}} \frac{\delta\pi_2[s_2, s_1, z]}{\delta s_2} dz = 0.$$

By differentiating first order conditions, BL shows that the sign of ds_1^* / dD_1 is equal to the sign of $\delta\pi_1[s_1, s_2, \hat{z}_1] / \delta s_1$. Noting that in the first order conditions the integral over $\hat{z}_1$ and strictly better states is zero, it follows by assumption that the marginal profit evaluated at $\hat{z}_1$ must be negative. Thus s_1^* is increasing in D_1. However, there are two points to note about this formulation.

First, it is assumed that $\underline{z} < \hat{z}$, stating that Firm 1 is insolvent in the worst state. In previous work it has not been pointed out that this place a restriction on the size of the debt. The debt has to be sufficiently large to make the firm insolvent in the worst state, in the Nash equilibrium with unlimited liability $\pi_1\left[s_1^*, s_2^*, \underline{z}\right] < D_1$. If this is not the case then the firm does not want to ignore the worst realisation. For low debt levels (and little uncertainty) and high profits in the Nash equilibrium with unlimited liability, this assumption is likely to be invalid. This implies that for low debt levels the Nash equilibrium is independent of debt, which is similar to the first part of Theorem 2.

Second, for the integral to be well defined (conditional debt being large enough to make $\underline{z} < \hat{z}$) it is required that $\hat{z} < \bar{z}$. By increasing s_2 the profit of Firm 1 is reduced, which increases $\hat{z}_1$, i.e. the lowest z where Firm 1 is solvent. It means that for high values of s_2 the best response function of Firm 1 is not well defined, since its payoff is zero whatever Firm 2 produces and in all realisations of uncertainty. To find the lower bound of Firm 2's strategies where Firm 1's best response function is uniquely defined, we note that Firm 1 has to be insolvent in all realisations, specifically $\bar{z}$. There exists a lower bound quantity of Firm 2, s_2^{DD}, where Firm 1's profit is not sufficient to cover its debt (as in the case without uncertainty), even in the best realisation defined by $\max \pi_1\left[s_1, s_2^{DD}, \bar{z}\right] = D_1$ for $s_1 \in S_1$. The lower bound of Firm 1's strategies, where the best response function is unique is $s_1^D = R_1\left[s_2^{DD}\right]$. For all Firm 2's strategies above s_2^{DD} Firm 1 is bankrupt in all realisations. The best response function of Firm 1, where it is well defined, is the same as in BL. However, it is not unique for all of Firm 2's strategies.

Now we proceed as before and in the first round eliminate all strategies for Firm i which are greater than $s_i^M = R_i\left[0\right]$. Moreover, s_i^M is not weakly dominated by any lower quantity, since it gives maximum payoff if Firm j would produce zero. The lower bound of Firm 2's strategies is as before $R_2[\infty]$. For Firm 1, strategies lower than s_1^D are weakly dominated. The lower bound can be no lower since the payoff is zero whatever Firm 2 produces and whatever realisation of uncertainty. After this we proceed as before by successively eliminating weakly dominated strategies. All theorems and corollaries in Section 3 have their counterpart in case of uncertainty. Moreover, if the debt is higher than the monopoly profit at the best realisation $\pi_1\left[s_1^M, 0, \bar{z}\right] \leq D_1$ no strategies are weakly dominated since it bankrupt in all realisations and strategies.

6.Concluding remarks

In most work on corporate governance it is assumed that equity holders have considerable discretion in the choice of strategy, and that creditors remain rather passive as long as due debt is paid. It is well known that limited liability of equity holders give them incentives to choose risky strategies, which may also effect the equilibrium outcomes in oligopoly models. In this paper, oligopoly games were studied from a different angle to provide new insigbts to why financial structure matters for product market interaction. Instead of directly asking which strategies will be used, we began by eliminating all strategies that will not be used by equity holders in control. Equity holders will eliminate a strategy which (to them) is not a strictly payoff maximising best response to any of the strategies the opponent can use. We then studied the upper and lower bounds of the strategies which survived iterated elimination of weakly dominated strategies. The upper and lower bounds of an indebted firm were shown to be (weakly or strictly) increasing in debt levels. The results of this paper generalises previous work in two ways. First, it was shown that even without uncertainty debt may affect the outcome of oligopoly games. Second, under uncertainty we extended the range of debt levels where predictions can be made on the outcome of the game.

One interesting feature of the surviving set of strategies is that there need not be a conflict of interest between equity holders and creditors. This holds even at high debt levels where the profit is insufficient to cover debt obligations in all surviving strategies. The intuition is that debt committed an indebted firm to strategies with high quantities (or prices) which may result in high operating profits and lower losses to the creditors. However, if there are conflicting interests we open up the interesting question of the possibilities to restructure existing debt. In this work we merely assumed that creditors were unable to influence equity holders, motivated by the argument that there were many creditors which each held only a small claim, and the existence of some direct costs of debt restructuring. Future work on the relation between financial structure and product market interaction have to deal carefully with the renegotiation possibilities as in Beaudry and Poitevin (1994). Another topic for future research is to study alternative assumptions about the observability of competing firms financial arrangements, which may or may not be perfectly observable.

References

Beaudry, P. and M.Poitevin, 1994, The commitment value of contracts under dynamic renegotiation, *RAND Journal of Economics*, 25:501-517

Brander, J.A. and T.R.Lewis, 1986, Oligopoly and financial structure: the limited liability effect, *American Economic Review*,76:956-970.

Brander, J.A.and T.R.Lewis, 1988, Bankruptcy costs and the theory of oligopoly, *Canadian Journal of Economics*, 21:221-243.

Börgers, T., 1992, Iterated elimination of dominated strategies in a Bertrand-Edgeworth model, *Review of Economic Studies*, 59:163-176.

Fudenberg, D. and J.Tirole, 1991, Game theory, The MIT Press, Cambridge, Mass. and London, England.

Gertner, R. and D.Scharfstein, 1991, A theory of workouts and the effects of reorganization law, *Journal of Finance*, 46:1189-1222.

Glazer, J., 1994, The strategic effects of long-term debt in imperfect competition, *Journal of Economic Theory*, 62:428-443.

Jensen, M.C. and W.Meckling, 1976, The theory of the firm: managerial behaviour, agency cost; and capital structure, *Journal of Financial Economics*, 3:305-360.

McAndrews, J.J. and L.I. Nakamura, 1992, Entry deterring debt, *Journal of Money, Credit, and Banking*, 24:98-110.

Maksimovic, V., 1988, Capital structure in repeated oligopolies, *RAND Journal of Economics*, 19:389-407.

Maksimovic, V., 1990, Product market imperfections and loan commitments, *Journal of Finance*, 45:1641-1653.

Moulin, H., 1984, Dominance solvability and Cournot stability, *Mathematical Social Sciences*, 7:83-102.

Showalter, D.M., 1995, Oligopoly and financial structure: Comment, *American Economic Review*, 85:647-653.

FIGURE 1.

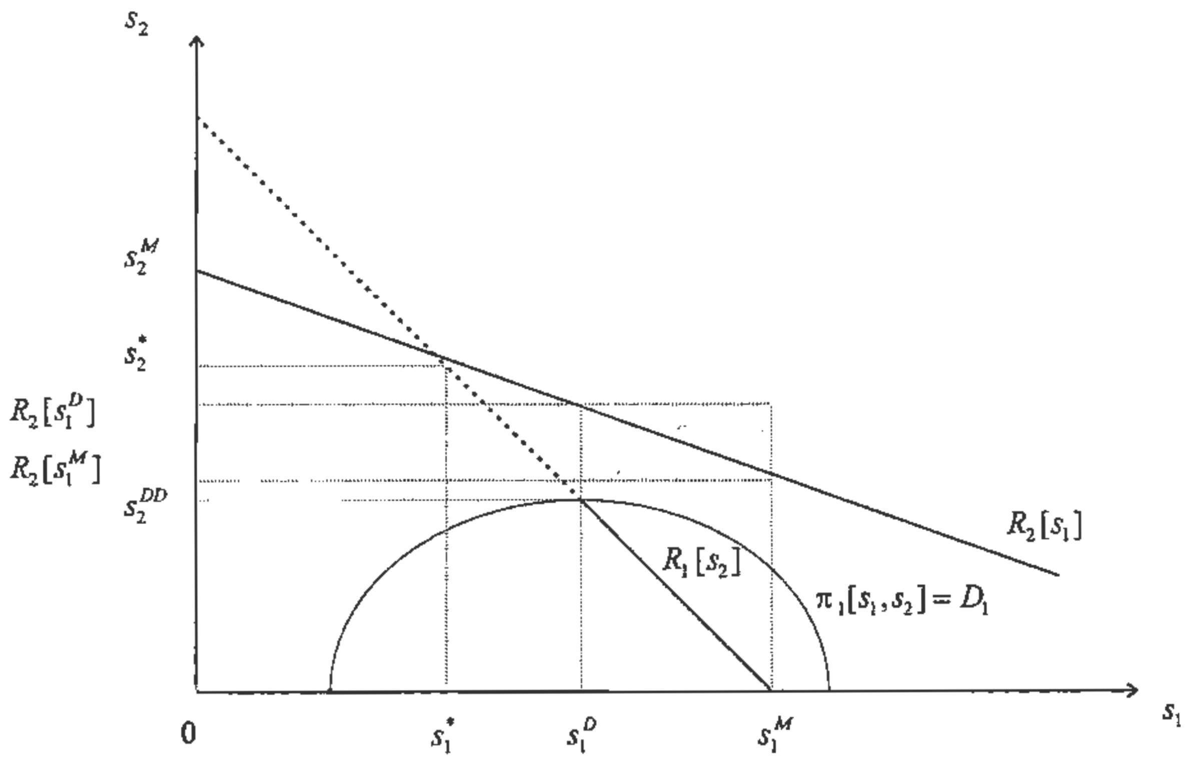


FIGURE 2 A.

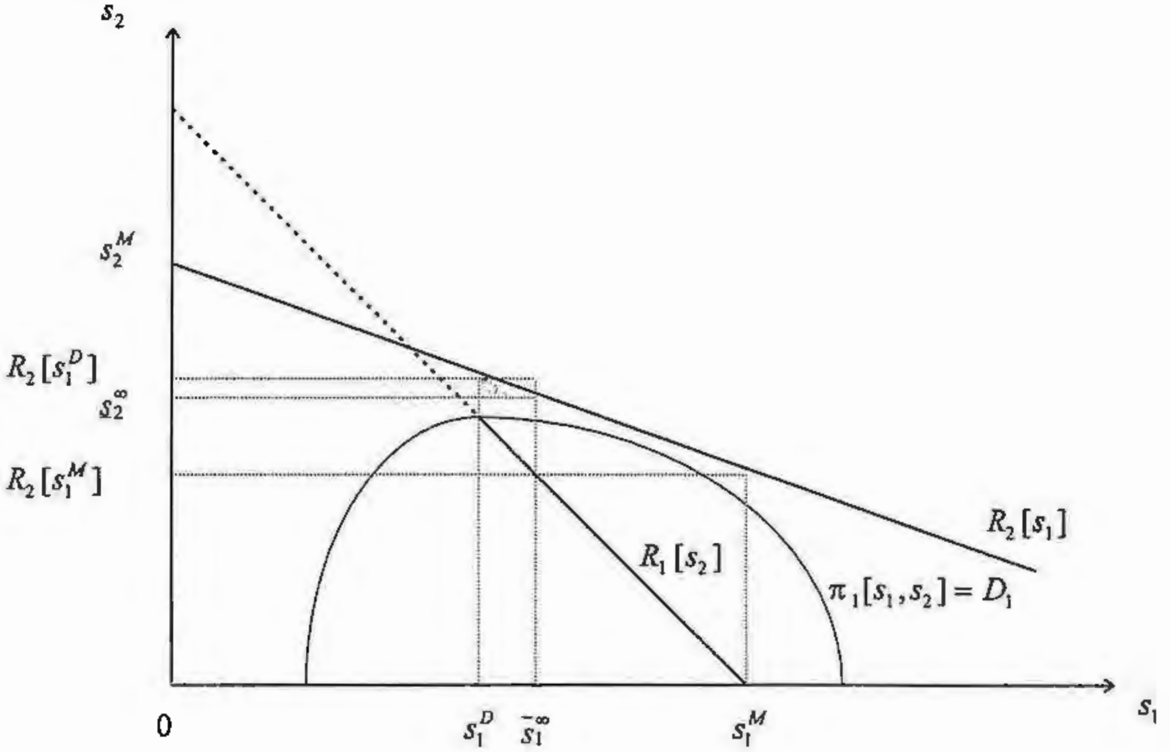

s_2
s_2^M
$R_2[s_1^D]$
s_2^∞
$R_2[s_1^M]$
$R_2[s_1]$
$R_1[s_2]$
$\pi_1[s_1, s_2] = D_1$
0
s_1^D
$\bar{s}_1^\infty$
s_1^M
s_1

FIGURE 2 B.

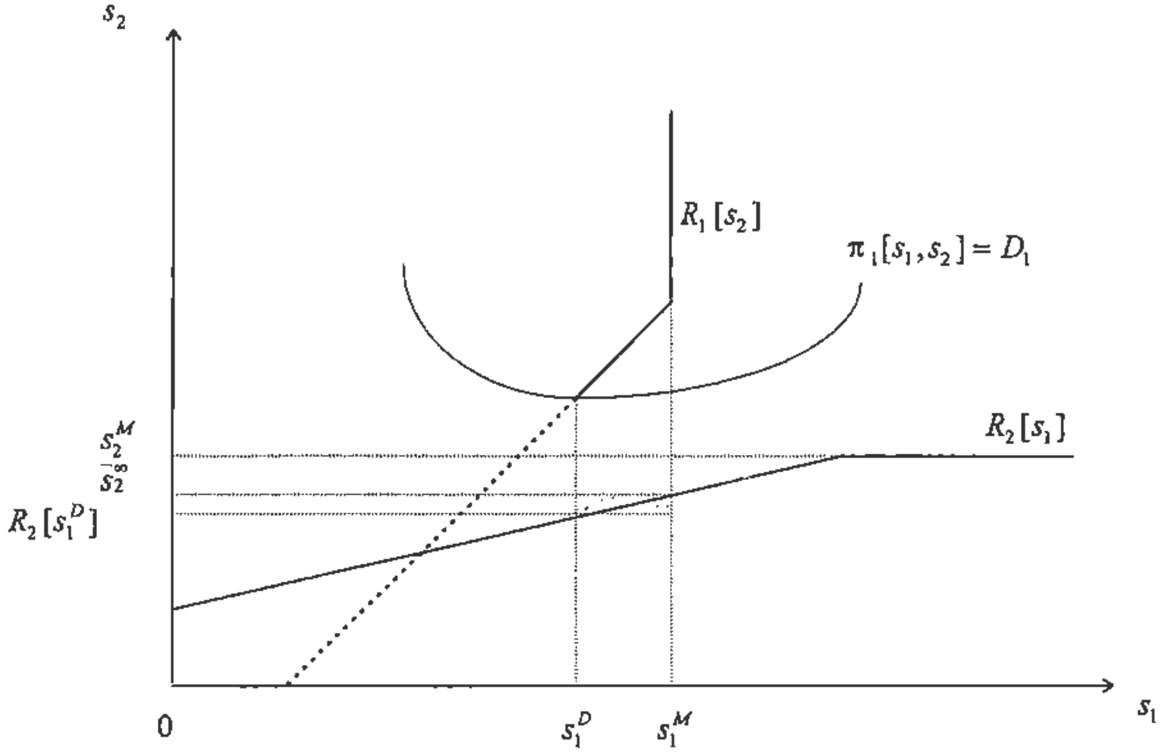
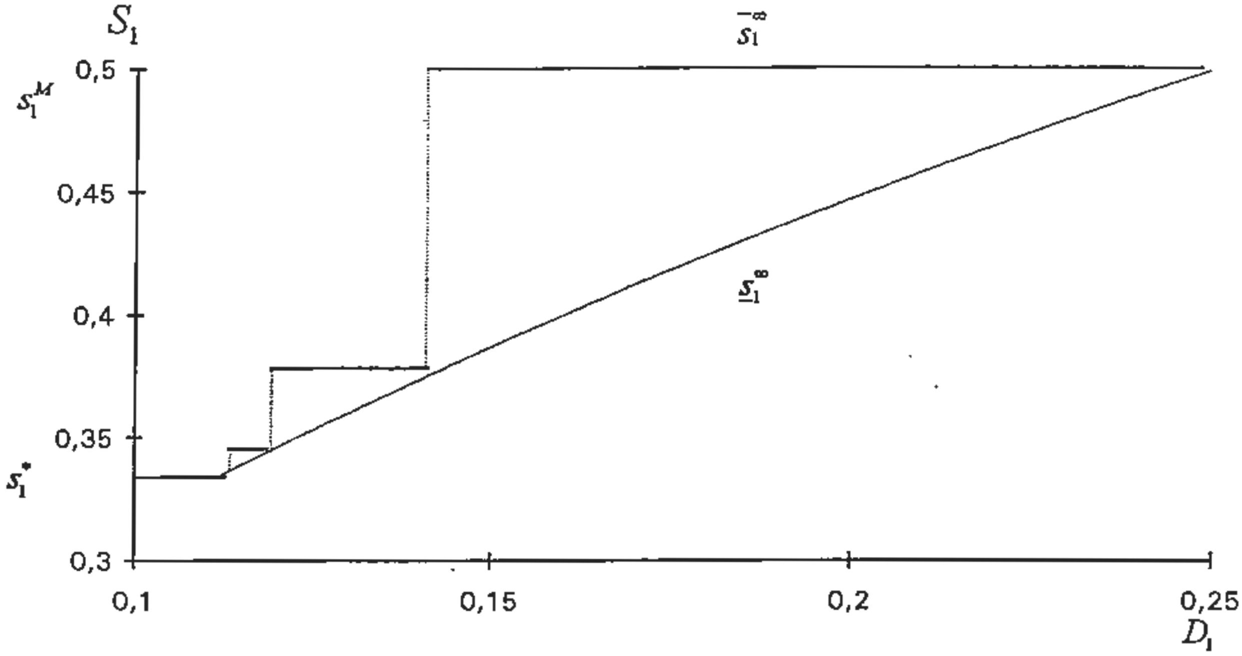


Figure 3.



S_1
$\bar{s}_1^{\infty}$
s_1^M
0,5
0,45
0,4
$\underline{s}_1^{\infty}$
0,35
s_1^*
0,3
0,1
0,15
0,2
0,25
D_1

What Fraction of a Capital Investment is Sunk Cost?*

Marcus Asplund

Department of Economics

Stockholm School of Economics

Box 6501, S-113 83 Stockholm

e-mail:nema@hhs.se

Abstract

We study the valuation of used capital assets, specifically metalworking machinery used in Swedish manufacturing industries. Two samples, collected 1960 and 1990, with information on salvage values of individual machine tools, show that 47% and 78% of the discarded assets were scrapped rather than sold on second-hand markets. Using information on both sold and scrapped assets, we estimate the expected salvage value, both conditional and unconditional on that the asset can be sold. The results indicate that in both samples these investments are largely sunk costs. For an average "new" machine, firms can only expect to get back 20-60% of the initial price. It is argued that this is largely explained by high fixed costs in buying and selling used capital, which cause second-hand markets to work poorly.

Keywords: Sunk cost; second-hand market; salvage value; censored regression; machine tools.

JEL numbers: C24; D24.

---

* I wish to thank Tore Ellingsen, Lars Bergman, Ernst Berndt, Lena Edlund, Gunnar Eliasson, Zvi Griliches, Christina Hartler, Thomas Lindh, Erik Mellander, Jan Södersten, and the participants at the Industrial Institute of Economic Research (IUI) seminar "Capital; Its Value, Its Rate of Return and Its Productivity" Stockholm, March, 1991 for many helpful suggestions. Of course, any remaining errors are my responsibility. Financial support from The Royal Swedish Academy of Science and IUI are gratefully acknowledged.

1. Introduction

In this paper we consider the valuation of used capital assets. We do this by a study of firm's decision to sell or scrap, and the salvage values of, metalworking machinery used in Swedish manufacturing industries. One would expect that these machine tools are assets with relatively little specificity for which there exist well functioning second-hand markets. However, our results indicate that these assets are essentially sunk cost investments. This conclusion is based on the findings that most discarded machines (even at low age) are scrapped rather than sold, and salvage values of the sold pieces are low.

With our data sets we can give an estimate of the fraction of a capital investment which is sunk, which has until now remained an important, but quantitatively largely unknown, factor in many economic models. Sunk costs is an often stressed determinant of entry/exit decisions and strategic investment (Baumol et al. (1981), Dixit (1980), and Eaton and Lipsey (1980)). Empirical work which demonstrate the importance of sunk costs for competition include Kessides (1990), Sutton (1991), and Worthington (1995). In economic relations, asset specificity and sunk costs may also lead to ex post opportunism and can therefore influence contracting between agents (Williamson (1975), Klein, Crawford, and Alchian (1978)). Even in a non strategic framework, the sunk cost component is an important factor in investment decisions, not only because low salvage values reduce the net present value of a given investment. In recent work on the timing of investments under uncertainty, surveyed in Dixit (1992), and Pindyck (1991), it is shown that sunk costs introduce an option value of waiting for new information.

Previous studies of capital asset depreciation have not been interested in the sunk cost component, but instead in the gradual decline in value as a reflection of capital's decline in productive efficiency. Under an assumption of well functioning second-hand markets, it is possible to use prices paid on these to deduce the sum of the decline in efficiency and obsolescence of capital as it ages, which is useful in estimating the efficiency and value of aggregated capital stocks.[1] However, the price paid for used capital will also be a function of the properties of the second-hand markets, and if these are inefficient there will be a wedge between the value of capital inside and outside the firm. List and transaction prices obtained from second-hand market dealers (used in e.g. Beidleman (1973) and Hulten and Wykoff (1981)) are useful primarily as an upper bound of salvage values of discarded assets, since they include the dealer's profit margin. Another potentially important problem with second-hand market data is that there may be a selection bias. This would arise if the assets sold in second-hand markets are not representative of the average asset, which may be impossible to sell. To avoid these problems, this study uses a different source of data compared to previous studies. With information on both sold and scrapped machine tools it is possible to study the expected salvage value of an average asset, and estimate the probability that it is sold rather than scrapped. From this we obtain estimates on the sunk cost component for this type of capital investments.

The paper is organised as follows. Section 2 contains a description of the two data sets and some observations based upon the descriptive statistics of them. Additional survey evidence is also presented. Section 3 gives the econometric model, which is a two-stage Heckman procedure. Section 4 gives the econometric results. Section 5 concludes the paper with a discussion of second-hand markets for this type of capital assets.

2. Data

The data has been collected directly from firms in Swedish manufacturing industries and contain information on discarded (sold or scrapped) metalworking machinery. The first sample was collected in 1960 from nine large firms, and include machines which were discarded 1954-59. The second sample was primarily collected in 1990-91 from four large

[1]For references on theoretical and econometric studies and more exact relationships between depreciation and decay in efficiency according to certain patterns see Hulten and Wykoff (1981), and Jorgenson (1989).

firms, and the assets were discarded 1983-91.[2] The machines were not discarded in connection with a major disinvestment as a bankruptcy, but rather as a matter of day to day business operations, e.g. due to capacity adjustments or changes in production requirements. The machines are mainly lathes, drilling machines, milling cutters and presses, and are aggregated due to the small sample sizes. Ideally, one would like to have detailed information on each machine, and estimate salvages values for each type separately. However, aggregation has also been used in previous studies (e.g. Hulten and Wykoff (1981)) and has the advantage of giving a broad picture of the salvage values of this class of assets.[3] Data consists of information on which year the machine was bought and when it was discarded, and the nominal initial price and salvage value (if any). The salvage value of scrapped machines have been set to zero, which is an approximation. However, the scrap value is insignificant, and is not even recorded separately in firms own accounting, so this bias is likely to be small. Since observed prices are nominal rather than real, it requires a correction for the inflationary impact on prices. In this study we deflate the nominal prices with index for ready-made goods in mining and manufacturing industries (from Statistical Abstracts of Sweden).[4] For obvious reasons this is not an ideal measure of general inflation, since it is not known exactly how it corrects for technological progress in various goods.[5] Some descriptive statistics on the number of the assets sold and scrapped, average ages and prices of the discarded assets in the samples can be found in Table 1.

---

[2]Most of the 1990 sample was collected by the author but 53 observations were collected in 1988 of which 38 were used in Hartler (1988). The 1960 sample was collected in 1960 by Jan Wallander, The Industrial Institute of Economic and Social Research, IUI, and presented in Wallander (1962).

[3]Some statistics about life lengths for various metalworking machines reveal that they are rather similar. Wallander (1962) estimate the average economic life lengths for lathes, drilling-machines, milling cutters and presses to 22, 29, 26 and 28 years respectively. The conclusion in Prais (1986) is similar.

[4]This index is not availiable for the years prior to 1924, where we used the index of metal working machinery calculated by Wallander (1962).

[5]Other studies have dealt with this problem in a number of ways. Beidleman (1973), also uses some price indices to deflate, but it is not clear how these are calculated. Hulten and Wykoff (1981) use a trend variable to take account of the inflation, which is possible because of their large sample, containing time series data as well as cross section data.

Table 1. Descriptive statistics of sold and scrapped machines in samples. Prices in SEK, with base year 1959 (1USD=5.20 SEK 1959). Average age in years. (Standard deviation in parenthesis).

Sample	1960 Sold	1960 Scrapped	1990 Sold	1990 Scrapped
No. of observations	310	276	102	364
Mean age	24.6	26.1	13.1	18.5
	(12.6)	(12.5)	(8.5)	(9.5)
Mean initial price	35000	14000	240000	40000
	(59000)	(23000)	(450000)	(88000)
Mean salvage value	4000	0	31000	0
	(7600)		(57000)	

There are several facts to be noted in Table 1. First, and most important, the high percentages, 47% (=276/(276+310) and 78% (=364/(102+364)), of the assets which are scrapped rather than sold. Second, the salvage values for sold machines are low, approximately 12% of the initial price. Third, the average ages of the discarded machines are lower than in comparable studies, in particular in the later sample. Beidleman (1973), and Oliner (1990) indicate that the service lives are approximately thirty years for this class of assets. Fourth, the average age of scrapped and sold machines are similar within the samples but different between samples. It is important to note that the machines are discarded at a much younger age in the 1990 sample.

The question is if the finding of high fractions scrapped machines is representative of the manufacturing industries, or if our sample is biased towards firms with highly specific capital. To answer this question a complementary survey was conducted among manufacturing firms in Sweden, with questions concerning the mean age of sold, scrapped and remaining machinery for the years 1990 and 1991. A short summary of the replies of the 38 firms which answered all these questions (out of 102 firms in industries similar to those in the samples) are presented in Table 2.

Table 2. Descriptive statistics from survey of 38 Swedish manufacturing firms. Average ages in years. (Standard deviation in parenthesis).

Year	1990	1990	1990	1991	1991	1991
No. of employed 1991	<500	<1000	>1000	<500	<1000	>1000
No. of observations	11	14	13	11	14	13
No. of sold machines	89	194	355	48	379	518
	(9.9)	(13)	(21)	(11)	(28)	(44)
Mean age	11.2	15.7	9.5	11.0	15.1	10.9
	(6.8)	(5.4)	(4.4)	(3.5)	(6.0)	(5.3)
No. of scrapped machines	98	467	4200	75	773	2900
	(11)	(44)	(503)	(16)	(73)	(290)
Mean age	14.2	18.3	12.1	10.9	15.5	12.0
	(6.7)	(4.2)	(4.4)	(4.6)	(6.6)	(4.0)
Mean age of remaining machines	n.a.	n.a.	n.a.	11.2	11.6	9.3
				(4.2)	(2.5)	(2.8)

For the firms with less than 500 employees, the average percentage scrapped machines 1990-91 is 55%, and for the other groups the percentages are 68%, and 89%. The percentage in the 1990 sample was 78%, which is not extreme compared to the average in the survey. Scrapped machines have a somewhat higher average age than the sold, which is again consistent with Table 1. It is also interesting that the average ages are even lower than in the later sample which is, as noted above, lower than comparable studies. The 1990 sample (sold and scrapped) has an average age of 17 years, so it may actually overstate the average age of discarded machines. As expected, the average age of the discarded assets are higher than for the remaining, (except for the smallest firms in 1991). However, the age of discarded machines for the smallest and the largest firms is only 1-3 years more than for the average remaining machines, while for the medium sized firms the difference is 4-7 years. The conclusion of the comparison of the descriptive statistics in Table 1 and Table 2, is that the 1990 sample is reasonably representative of Swedish manufacturing industries over this time period. The main differences are that for smaller firms the percentage scrapped is lower, and that large firms scrap machines at a younger age.

3. Econometric model

A machine is sold when the valuation of its services to another firm exceeds that of the current owner, and it will be scrapped when the value to the current owner is always higher than that of other firms. The assumption (although implicit in most papers) in estimating the salvage values of used capital assets, is that there exist a demand function which is a function of the asset's age and other observable characteristics, and that the supply of assets varies with the production requirements of the current owner. This assumption make it possible to identify the demand function, which corresponds to the salvage value function we want to estimate.

Estimating salvage values for the discarded machines raises two specific problems. The first, which has been the focus of several earlier papers, is to find a flexible functional form to discriminate between the different depreciation patterns suggested in the literature. Box-Cox transformations[6] have been used in several studies following Hulten and Wykoff (1981). However, it has been demonstrated that the original maximum likelihood estimator has several less desirable properties, such as scale invariance of the t-statistics. The non-linear least squares estimator[7] of the Box-Cox class suggested by Berndt, Showalter, and Woolridge (1993) was used in an earlier version and could not reject the simple specifications used here.[8] The estimator used instead have coefficients which are easier to interpret than those of the Box-Cox transformations, and can capture much of the (possible) nonlinearities in the depreciation patterns.

The second, and more important, problem is to find a consistent method to handle the sample selection problem if sold assets are not representative of the average asset. The assets which

---

[6]Estimating λ, β from the equation $\left((y_t - 1)/\lambda\right)^{\lambda} = \beta x_t + \varepsilon_t$ under the assumption that the transformations induce normality of the error term.

[7]Estimating λ, β from the equation $y_t = \left(1 + \lambda\beta x_t\right)^{1/\lambda} + \varepsilon_t$.

[8]Results are availiable from the author on request.

are actually sold and not scrapped is presumably not a random sample of the population, for example they may have been more expensive as new. There have been some attempts to correct for this bias in previous studies. Hulten and Wykoff (1981) suggest that each transaction price is multiplied with the probability that the asset has survived until it is observed on the second-hand market, to give the expected value for an average asset. However, these probabilities can not be estimated from within the sample, instead they are calculated from survival distributions of similar assets, estimated by Winfrey (1935). This problem can be studied more rigorously here due to the nature of our data, and by applying the two-stage procedure of Heckman (1979) (see Amemiya (1985 p. 360-408), and Greene (1993 p. 691-714)).

Let the selection mechanism be $z_t^* = \gamma w_t + u_t$, where w_t denote the variables determining if the machine is sold or scrapped, and γ a parameter vector to be estimated. Set $z_t = 1$ if we observe that the machine is sold and $z_t = 0$ if it is scrapped

$$z_t = \begin{cases} 1 & if \quad z_t^* > 0 \\ 0 & if \quad z_t^* \le 0 \end{cases}.$$

That is, if γw_t is large it is more likely that the asset is sold. For the sold machines the salvage value, y_t, is

$$y_t = \beta x_t + e_t .$$

In Heckman's two-stage procedure the assumption is that the error terms (u_t, e_t) are bivariate normal $(0,0,1,\sigma_e, \rho)$.[9]

The expected salvage value of the sold machines is then

$$E\left[y_t \middle| x_t, z_t = 1\right] = \beta x_t + E\left[e_t \middle| z_t = 1\right].$$

The conditional expectation of the error term is

$$E\left[e_t \middle| z_t = 1\right] = E\left[e_t \middle| u_t > -\gamma w_t\right] = \sigma IMR(\psi_t),$$

where

$$IMR(\psi_t) \equiv \frac{f(\psi_t)}{F(\psi_t)}, \quad \psi_t \equiv \gamma w_t, \qquad \sigma \equiv \rho\sigma_e.$$

[9]In the setting we consider, the Heckman procedure is only a slight generalisation of the standard Tobit model, where we allow the independent variables in w and x to be different.

$f(\psi_t)$ is the density function for a standard normal variable and $F(\psi_t)$ is its cumulative density function. The variable $IMR(\psi_t)$ is the inverse of Mills ratio. The first stage is a standard Probit regression with the dependent variable *SELLDUM*, which takes the value 1 (0) if the asset is sold (scrapped). From this we obtain consistent estimates of $IMR(\psi_t)$. The second step is to use these as an independent variable in the formulation of the expected value conditional on the salvage value being strictly positive

$$E[y_t|x_t, z_t = 1] = \beta x_t + \sigma IMR(\psi_t). \tag{1}$$

The expected salvage value, unconditional it being strictly positive, is

$$E[y_t\, x_t, z_t = 1] = F(\psi_t)(\beta x_t + \sigma IMR(\psi_t)). \tag{2}$$

This is the expected value conditional on being above the limit, weighted with the probability of being above the limit (McDonald and Moffit (1980)).

The starting point in the specification of the equation to be estimated is the assumption of a relationship between age of the asset, denoted *AGE*, and the normalised salvage value, $Q \equiv SELL/BUY$, where *SELL* is the deflated salvage value and *BUY* is the deflated price of the asset as new.

To capture the possibly non-linear relation we use the quadratic specification

$$Q_t = \beta_0 + \beta_1 AGE_t + \beta_2 AGESQR_t + \sigma IMR_t + e_t. \tag{3}$$

As a comparison a log-linear specification is also estimated (where the coefficients are different from those in (3))

$$LOGQ_t = \beta_0 + \beta_1 AGE_t + \beta_2 AGESQR_t + \sigma IMR_t + e_t. \tag{4}$$

For frictionless second-hand markets and if capital is flexible, it is expected that if a firm decides to discard a new machine, it will be sold and the salvage value will be close to the price initially paid. In terms of (3) and (4) the estimated Q at AGE=0, will be close to 1. If the coefficients of *AGESQR* and *IMR* are zero in (4) it implies an exponential depreciation pattern with a constant depreciation rate β_1.

A third specification is obtained by multiplying both sides of (3) with *BUY* and adding a constant

$$SELL_t = \beta_0 + \beta_1 BUY_t + \beta_2 AGEBUY_t + \beta_3 AGESQRBUY + \sigma IMR_t + e_t. \qquad (5)$$

The reason for using (5) is that information is lost if there are differences in depreciation between more and less expensive machines. Since the data have been collected from different firms, there may be differences between the probability that a machine is sold, which would be the case if some firms have less specific machines or are paying more attention to selling them. Moreover, there may be differences in the salvage values for sold machines, also due to asset specificity. To control for these differences we use dummy variables for each firm, denoted *FIRMDUM*, in both the Probit and the least squares estimation. In the specification with *SELL* as the dependent variable this implies that *FIRMDUM* are multiplied with *BUY*. To reduce potential problems with multicollinearity, additional transformations of the independent variables *AGE* and *BUY* are used in the Probit regression.

4. Results

In the initial Probit regression the dependent variable is *SELLDUM* and the independent variables are transformations and cross products of the initial price and asset age. We expect that younger assets, which probably are less obsolete and have more remaining years of productive possibilities, are more likely to be sold than scrapped. The same is expected for assets which were more expensive as new. The results are reported in Table 3.

Table 3. Results from Probit regression

Sample Variable	1960 *SELLDUM*	1990 *SELLDUM*
CONSTANT	0.271	-1.35***
	(0.384)	(0.386)
BUY	0.387 E-4***	0.589 E-5***
	(0.682 E-5)	(0.151 E-5)
BUYSQR	-0.173 E-10	-0.101 E-11***
	(0.121 E-10)	(0.334 E-12)
AGE	-0.0249	-0.0276
	(0.0196)	(0.0257)
AGESQR	0.487 E-3	0.477 E-3
	(0.347 E-3)	(0.582 E -3)
AGEBUY	-0.622 E-6***	-0.105 E-6 *
	(0.223 E-6)	(0.610 E-7)
FIRM1DUM		1.28***
		(0.356)
FIRM2DUM		0.572*
		(0.306)
FIRM3DUM		0.301
		(0.288)
FIRM4DUM	-0.940**	
	(0.366)	
FIRM5DUM	-0.482	
	(0.319)	
FIRM6DUM	-0.838***	
	(0.302)	
FIRM7DUM	-0.213	
	(0.314)	
FIRM8DUM	0.704**	
	(0.347)	
FIRM9DUM	-0.537*	
	(0.312)	
FIRM10DUM	-0.906**	
	(0.360)	
% correct predictions	70%	84%

Standard errors in parenthesis.
Significance levels: 10% *=abs(t)>1.65, 5% **=abs(t)>1.96, 1% ***=abs(t)>2.58 .

The predictive abilities of the estimated equations are good, 70% and 84% respectively.[10] The coefficient of *BUY* is highly significant with the expected positive sign in both samples, which is interpreted as a higher probability to sell (rather than scrap) machines which were expensive

[10]However, it should be noted that predicting only sold (scrapped) in the 1960 (1990) sample would yield 55% (80%) correct predictions.

as new. In the 1990 sample *BUYSQR* is significantly negative, indicating that the most expensive assets are less likely to be sold than *BUY* would suggest. It can be noted that in neither of the samples are the coefficients of *AGE* and *AGESQR* significant. Thus, there is no independent effect of age on the sell/ scrap decision, which one would have expected if technological progress was rapid. It is important that the cross term *AGEBUY* is negative and significant in both samples. The interpretation is that for a given initial price, the probability to sell a discarded machine decreases with its age. Some of the dummy variables are significant, which indicate differences in selling/ scrapping behaviour across firms. For example, in the 1960 sample, FIRM8 is more likely to sell a discarded machine than FIRM4, given its age and initial price.

To give a picture of the estimated coefficients in Table 3, Figures 1A and 1B show the probability that a discarded asset is sold rather than scrapped, as a function of its age and initial price. If second-hand markets are frictionless and capital is flexible between uses, the probability to sell a new machine (*AGE*=0), would be close to 1. To show the probabilities of an average firm, the simulations are based on the mean of the estimated coefficients of the firm dummies in each sample. The range of initial prices reflect the mean in each sample (25000 and 84000), but for the range of asset ages one has to remember that the mean age in the 1990 sample is 17 years, and only a few are more than 30 years old. The probability that an asset is sold decreases sharply with its age, in both samples. Likewise, machines with low initial prices are unlikely to be sold. The probabilities are generally lower in the 1990 sample, that is, an average asset was less likely to be sold in 1990, than in 1960.

Figure 1A. Probability that a discarded machine is sold in the 1960 sample.

Figure 1B. Probability that a discarded machine is sold in the 1990 sample.

In the second stage equations (3)-(5) are estimated by least squares, and the results are summarised in Table 4. Appendix 1 shows the result of a specification without firm dummies. However, based on a F-test that specification was rejected against this alternative.

Table 4. Results from least squares regression of equations (3)-(5)

Sample Variable	1960 *Q*	1990 *Q*	1960 *LOGQ*	1990 *LOGQ*	1960 *SELL*	1990 *SELL*
CONSTANT	0.353*** (0.0497)	0.358*** (0.0855)	-0.981*** (0.211)	-1.18*** (0.346)	237 (1010)	-9420 (9330)
AGE	-0.0115*** (0.340 E-2)	-0.0279*** (0.550 E-2)	-0.0640*** (0.0147)	-0.149*** (0.0247)		
AGESQR	0.111 E-3*** (0.531 E-4)	0.469 E-3*** (0.127 E-3)	0.460 E-3* (0.261 E-3)	0.248 E-2*** (0.580 E-3)		
BUY					0.278*** (0.0571)	0.427*** (0.0612)
AGEBUY					-0.835 E-2** (0.378 E-2)	-0.0359*** (0.671 E-2)
AGESQRBUY					0.120 E-3** (0.598 E-4)	0.866 E-3*** (0.210 E-3)
FIRM1DUM		0.0148 (0.0722)		0.0695 (0.285)		-0.552 E-2 (0.0108)
FIRM2DUM		-0.150** (0.0655)		-0.727*** (0.237)		-0.0882*** (0.0185)
FIRM3DUM		-0.131* (0.0684)		-0.781*** (0.234)		-0.142* (0.0611)
FIRM4DUM	0.154* (0.0799)		0.108 (0.268)		0.0315 (0.0582)	
FIRM5DUM	-0.0950*** (0.0357)		-1.03*** (0.225)		-0.0684 (0.0528)	
FIRM6DUM	-0.0812** (0.0330)		-0.640*** (0.221)		0.855 E-2 (0.0255)	
FIRM7DUM	-0.109*** (0.0206)		-1.04*** (0.132)		-0.0992*** (0.0204)	
FIRM8DUM	-0.0114 (0.0281)		-.0666 (0.141)		0.0158 (0.0504)	
FIRM9DUM	-0.0671 (0.0410)		-0.943*** (0.221)		-0.0660 (0.0491)	
FIRM10DUM	0.0873 (0.0787)		-0.210 (0.303)		0.0417 (0.0580)	
IMR	0.0772** (0.0336)	0.137*** (0.02811)	0.694*** (0.199)	0.742*** (0.165)	53.7 (919)	10600* (6170)
Adjusted R^2	0.37	0.40	0.44	0.40	0.61	0.83

White (1980) heteroskedastic-consistent standard errors in parenthesis.
Significance levels: 10% *=abs(t)>1.65, 5% **=abs(t)>1.96, 1% ***=abs(t)>2.58 .

The explanatory power of the regressions are reasonably high for these samples where different types of machines have been pooled. The high R^2 in the last specification is due primarily to high success in estimating the salvage values of machines with high initial price. It is particularly noteworthy that the two samples reveal striking similarities in the parameter estimates, since the time span is about thirty years.

We focus first on the regressions with *Q* and *LOGQ* as dependent variables (specifications (3) and (4)). As noted above, if capital is flexible and second-hand markets well functioning, we expect the estimated *Q* at *AGE*=0 to be close to 1. This is not consistent with the results reported in Table 4. For example, in the 1960 sample with e.g. *Q* as dependent variable, the estimated *CONSTANT* is only 0.353. The highest firm dummy, *FIRM4DUM*, is 0.154, and the effect of *IMR* (calculated from the estimated parameters in Table 3) for an asset from this firm with initial price 25000, is 0.0772*0.624=0.0482. Adding these effects yields a conditional expected salvage value (1) of 55% of the initial value. The estimate is similar for the specification with *LOGQ* as dependent variable, but would be lower for the other firms. In the 1990 sample and specification (3), the highest conditional expected value (for assets from FIRM1 with initial price 84000) is 58%. Hartler (1988), which use a subset of the 1990 sample with only sold machines, estimate the value at *AGE*=1 (due to the variable transformations it is impossible to calculate the value at *AGE*=0) to be 60% of the initial price. For both specifications and samples *AGE* and *AGESQR* are significant. In (3) the negative coefficient of *AGE* and positive *AGESQR* give a convex salvage value function. This implies that the hypothesis that these assets have a linear depreciation pattern can be rejected.[11] Similarly, in (4) the constant depreciation rate hypothesis is rejected, since *AGESQR* are significant. The sample selection variable *IMR* is highly significant in both samples. This means that selection sample bias really is an issue here, i.e. the sold assets are not fully representative of the total population.

[11]A convex depreciation pattern is consistent with many other studies. Hulten and Wykoff (1981 p. 106) summarize the results of previous studies and note "even though methods do vary across studies, the general conclusion which emerges is that the age-price patterns of various assets have a convex shape".

The regression with *SELL* as the dependent variable yields even lower conditional salvage values than (3) and (4), in the 1960 sample, whereas the opposite is true in the 1990 sample. As an example, in the 1960 sample at *AGE*=0 the coefficients of *BUY*, and *FIRM1DUM* implies that the salvage value is about 27.8+3.15% of the initial price, and the *CONSTANT* and *IMR* add a few more percent. The effect of *AGEBUY* and *AGESQRBUY* gives a slow decline in salvage values after the initial drop in value. *IMR* is significant only (at the 10% level) in the 1990 sample. This implies that the problem of selection bias is reduced in this specification and the prices paid can be representative of the conditional salvage values (1). However, the information on scrapped machines is still valuable in determining the unconditional salvage value (2), since that requires information on the probability that the machine is sold rather than scrapped.

Figures 2A and 2B shows the conditional expected salvage values, based on the log-linear specification (4) and the mean of the estimated coefficients for the firm dummies. The surfaces show the sharp initial drop in salvage values, which also explain why (for these samples) it is less interesting to use a flexible functional form (as in Hulten and Wykoff (1981)) to discriminate between, for instance, log-linear and linear specifications. For this to be a relevant question to focus on, it is necessary that the initial drop in value is rather small, since the assumption is that prices reflect declines in efficiency. This can hardly be the case here, where it would imply that machines lose more than 50% of their efficiency during the first year. For an average machine in the 1960 sample, with initial price 25000, the conditional expected salvage value (1) is 30% at *AGE*=0. The unconditional value (2) is 24%, which is obtained by multiplying the value in Figure 1A with that in Figure 2A For the 1990 sample, the drop in value is even more dramatic. The conditional expected salvage value is 45% and the unconditional is 17%, for the average asset with initial price 84000. In the 1990 sample, this implies that 55-83% of the capital cost is sunk! These are very high estimates, especially for the assets in this study, which would be expected to be less specific. However, caution is needed in giving this interpretation since there are no observations with *AGE*=0 and relatively few are scrapped at low age, in particular in the earlier sample. Still, the initial drop in value is so significant in both samples, and estimated salvage values so low in all specifications, that there is no alternative explanation than that the sunk cost component is substantial. The main difference between the specifications is that (3) and (5) give a faster decline in salvage values than (4), where the estimated salvage values are bound to be nonnegative even at high age.

Figure 2A. Expected salvage value, conditional on the machine being sold, in the 1960 sample.

Figure 2B. Expected salvage value, conditional on the machine being sold, in the 1990 sample.

From Table 3 and Table 4 it is clear that there exists significant differences across firms in both the probability that a discarded machine is sold, and the prices paid for sold machines. In the 1990 sample, *FIRM1DUM* has the highest estimated coefficient in both the Probit and the least squares regression. This indicates that FIRM1's capital is less sunk, in the sense that it is both more likely that a discarded machine is sold and, when it is, a relatively high price is paid for it. In the 1960 sample, *FIRM8DUM* has the highest coefficient in the Probit regression, but its coefficient in the least squares regression is low. *FIRM4DUM*, on the other hand, has a low coefficient in the Probit regression, but the highest in the least squares regression. This is surprising, since one would expect firms which scrap most of their machines to get low salvage values for the machines which are sold. In this particular case, part of the explanation may be that the estimate of *FIRM4DUM* in the least squares regression is based on only 13 observations. This give poor precision, indicated by the fact that the variable is significant only (at the 10% level) with Q as dependent variable. An alternative explanation is that FIRM4's machines differ substantially in specificity, where some are flexible and can be sold at high prices whereas many other are firm specific and scrapped. However, our data is not detailed enough to investigate this interesting question further, since it requires information on the characteristics of individual machines.

5. Concluding remarks

The most significant finding of this study is that capital investments in metalworking machinery (machine tools) appear to be largely sunk costs. This is supported by the fact that in the examined data sets, from 1960 and 1990, as much as 47% and 78% of the machine tools are scrapped rather than sold. The finding is also supported by a survey of the selling and scrapping decisions of 38 Swedish manufacturing firms. The econometric estimation showed that the expected salvage value of an average "new" machine, conditional on it being sold, is only about 20-60% of its initial price. This implies that 40-80% of the investment in machine tools is sunk cost. The fraction is even higher for the unconditional expected salvage values, where we take into account that many machines are scrapped rather than sold. However, there exists significant differences across firms, both in their sell/ scrap decisions and the expected salvage values for the sold machines. The estimated salvage values are lower than one would expect for this class of assets, and undoubtedly much lower than for assets like aeroplanes, ships, and buildings. It is therefore worthwhile to comment on some potential sources of bias.

Both in the samples and in the survey the size of the firms is quite large. The smallest firms in the survey (with <500 employees) had a lower percentage of scrapped machines than in the 1990 sample, which indicate that we underestimate the use of second-hand markets. Still, these firms scrapped 55% of their machines, which is a substantial percentage, and for the larger firms the percentage were even higher. It is possible small firms have lower percentages scrapped machines, but it is less clear that the prices of the machines that are sold would be much higher than estimated here. A second potential bias is if the years covered in the samples are in recession, where demand for second-hand machines is likely to be low (see e.g. Shleifer and Vishny (1992)). However, our samples cover years both in booms and recessions, so this bias is likely to be small. A third source is that we set the scrap value to zero, while in reality it is a positive number. Neither this is a convincing explanation to the low salvage values, since even the prices for the sold machines are very low, and scrap values are only a fraction of these. A fourth possible source of bias is if our measure of inflation deflate the price fall too much. This is possible, but when other price indices were used the results were similar. Moreover, the finding of large fractions of scrapped machines remains unexplained, which indicates that the question of deflation is not crucial. Below we briefly comment on some explanations to why salvage values are low for the assets in the samples.

Informational asymmetries between informed sellers and uninformed buyers on the quality of an asset may lead to an equilibrium with prices reflecting low quality and only low quality assets (lemons) entering the market (Akerlof (1970)). However, the assumption of uninformed buyers hardly describes the second-hand markets for machinery where, as noted by Beidleman (1973), most of them are professionals and well capable of judging the quality of a machine. Interviews with firms, machine dealers and auctioneers support this conjecture.

Capital embodied technological change may force the older vintages of capital out of firms, or into a second production line as spare capacity. Few companies would be willing to buy those obsolete machines. However, as Oliner (1990) remarks, the magnitude of the technological change in metalworking machinery is probably not very large. This is supported by the survey of Prais (1986) which showed old vintages side by side with brand new ones. Our survey also showed that the discarded machines were only slightly older than the remaining. The low salvage values give little economic incentive for firms to definitely discard capital, and economically dead machines might then be retained but little used, which would bias the results in the opposite direction and give higher average ages of the discarded machines. However, interviews with firms indicate that they discard machines which can no longer hold their position in the first production line.

For process/ product specific machines there will be few (if any) buyers, so rather than incurring costs of advertising machines for sale, or having someone occupied with trying to sell machines, they are scrapped. The surprising fact is the extensive scrapping of the machines in our samples, which are not expected to be specific. Moreover, the survey showed that our samples are not unique in this respect. An explanation could be that even if these machines seem to be general purpose equipment, there are some specificities which cause adjustment costs of the buyer. Fixed costs in adjustment and transport will favour new or expensive ones, since it is not worth the cost and effort to buy and install a cheap old machine which will only be used for a few years. This is supported by the higher probabilities to sell, and higher expected salvage values for machines which were more expensive as new. This is probably the most important explanation to why second-hand markets for this type of capital work poorly. However, a clear distinction has to be made between a marginal disinvestment, i.e. selling a single machine, and a large disinvestment, i.e. selling a whole production line or plant. Since salvage values are low, these costs are too high to motivate the effort of getting marginal disinvestments sold. For a large disinvestment on the other hand, the costs of finding buyers, or holding an auction, is of relatively minor importance and most assets will be sold. The interpretation of the salvage values estimated in this paper is then that they provide a lower bound on what firms can expect to get in a large disinvestment. Additional research is needed to compare the salvage values estimated in this paper with those of assets from liquidated firms, both when assets are sold piecemeal on an auction, and as a going concern. It would also be interesting to study the characteristics of the buyers. This could answer questions on whether capital from liquidated firms is transferred mainly between industries, or within the industry to remaining firms, and if these patterns differ in booms and recessions.

REFERENCES

Akerlof, G. (1970), The Market for Lemons, *Quarterly Journal of Economics*, 84:488-500.

Amemiya, T. (1985), Advanced Econometrics, Harvard University Press, Cambridge Mass.

Baumol, W.J. and R.D. Willig (1981), Fixed Costs, Sunk Costs, Entry Barriers, and Sustainability of Monopoly, *Quarterly Journal of Economics*, 95:405-431.

Beidleman, C.R. (1973), Valuation of Used Capital Assets, Studies in Accounting Research #7, American Accounting Association, Sarasota, Florida.

Berndt, E.R., M.H. Showalter, J.M. Wooldridge (1993), An Empirical Investigation of the Box-Cox Model and a Nonlinear Least Squares Alternative, *Econometric Reviews*, 12:65-102.

Dixit, A. (1980), The Role of Investment in Entry-Deterrence, *The Economic Journal*, 90:95-106.

Dixit, A. (1992), Investment and Hysteresis, *Journal of Economic Perspectives*, 6:107-132.

Eaton, B.C. and R.G. Lipsey (1980), Exit Barriers are Entry Barriers: The Durability of Capital as a Barrier to Entry, *The Bell Journal of Economics*, 11:721-729.

Greene, W.H. (1993), Econometric analysis, 2nd ed. Macmillan Publishing Company.

Hartler, C. (1988), Rate of Economic Depreciation of Machines - a Study with the Box-Cox Transformation, IUI Working Paper 190b.

Heckman, J. (1979), Sample Selection Bias as a Specification Error, *Econometrica*, 47:153-161.

Hulten, C.R. and F.C.Wykoff (1981), The Measurement of Economic Depreciation, in Charles R. Hulten ed., Depreciation, Inflation and Taxation of Income from Capital The Urban Institute Press.

Jorgenson, D.W. (1989), Capital as a Factor of Production, in Dale W. Jorgenson and Ralph Landau eds., Technology and Capital Formation, The MIT Press.

Kessides, I.N. (1990) Market Concentration, Contestability, and Sunk Costs, *Review of Economics and Statistics*, 72:614-622.

Klein, B. and R.Crawford, and A.Alchian (1978), Vertical Integration, Appropriable Rents, and the Competitive Contracting, *Journal of Law and Economics*, 21:297-326.

McDonald, J. and Moffit, R. (1980), The uses of Tobit Analysis, *Review of Economics and Statistics*, 62:318-321.

Oliner, S.D. (1990), New Evidence on the Depreciation and Retirement of Machine Tools, Mimeo, Board of Governors of the Federal Reserve System Division of Research and Statistics, Washington.

Pindyck, R.S. (1991), Irreversibility, Uncertainty, and Investment, *Journal of Economic Literature*, 29:1110-1148.

Prais, S.J. (1986), Some International Comparisons of the Age of the Machine-stock, *Journal of Industrial Economics*, 34:261-78.

Shleifer, A. and R.W. Vishny (1992), Liquidation Values and Debt Capacity: A Market Equilibrium Approach, *Journal of Finance*, 47:1343-66.

Sutton, J. (1991), Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration, The MIT Press., Cambridge Mass.

Wallander, J. (1962), Verkstadsindustrins Maskinkapital. (The Machinery of the Manufacturing Industries.) Almqvist och Wicksell. Stockholm.

White, H. (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48:817-838.

Williamson, O. (1975), Markets and Hierarchies, Free press.

Winfrey, R. (1935), Statistical Analyses of Industrial Property Retirements, Iowa Engineering Experiment Station, Bulletin 125.

Worthington, P.R., (1995), Investment Cash Flow, and Sunk Costs, *Journal of Industrial Economics*, 43: 49-61.

Appendix

Table A. Results from least squares regression without firm dummies

Sample Variable	1960 *Q*	1990 *Q*	1960 *LOGQ*	1990 *LOGQ*	1960 *SELL*	1990 *SELL*
CONSTANT	0.355***	0.370***	-1.25***	-1.09***	-418	-9550
	(0.0476)	(0.0448)	(0.191)	(0.193)	(695)	(11100)
AGE	-0.0158***	-0.0281***	-0.0815***	-0.152***		
	(0.353 E-2)	(0.532 E-2)	(0.0155)	(0.0237)		
AGESQR	0.168 E-3***	0.461 E-3***	0.600 E-3**	0.245 E-2***		
	(0.529 E-4)	(0.119 E-3)	(0.275 E-3)	(0.660 E-3)		
BUY					0.359***	0.357***
					(0.621)	(0.0699)
AGEBUY					-0.0186***	-0.0269***
					(0.518 E-2)	(0.736 E-2)
AGESQRBUY					0.277 E-3***	0.556 E-3***
					(0.102 E-3)	(0.201 E-3)
IMR	0.0848***	0.0582**	0.591***	0.313**	1350**	7630
	(0.0231)	(0.0251)	(0.133)	(0.140)	(661)	(7050)
Adjusted R^2	0.24	0.31	0.31	0.33	0.45	0.79

White (1980) heteroskedastic-consistent standard errors in parenthesis.
10% *=abs(t)>1.65, 5% **=abs(t)>1.96, 1% ***=abs(t)>2.58 .

Figure 1 A.

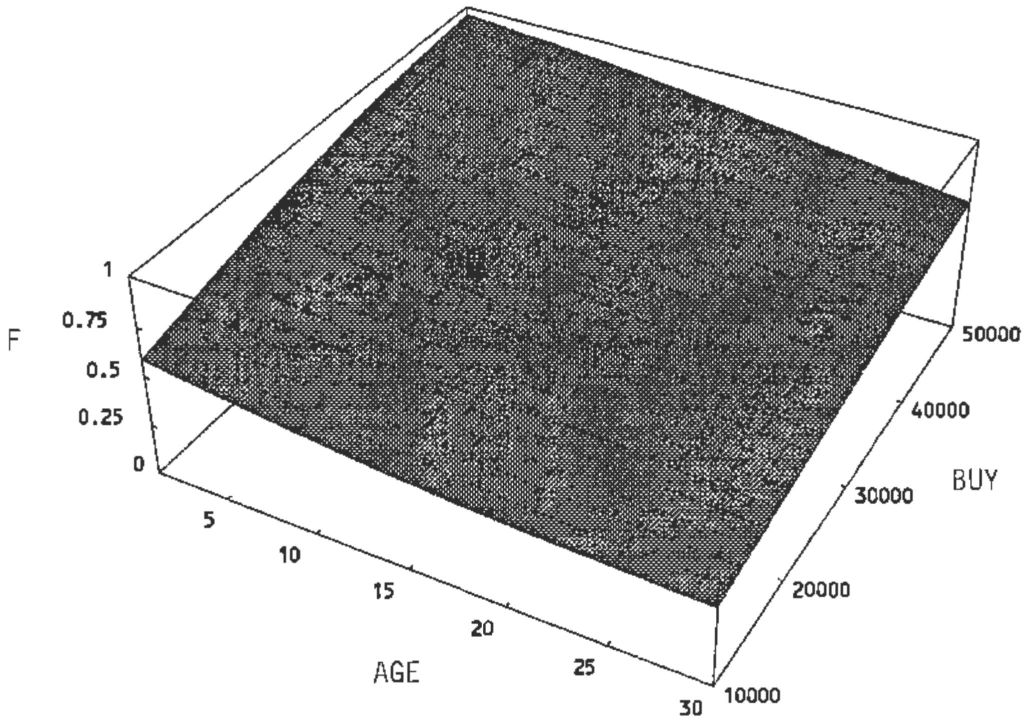


Figure 1 B.

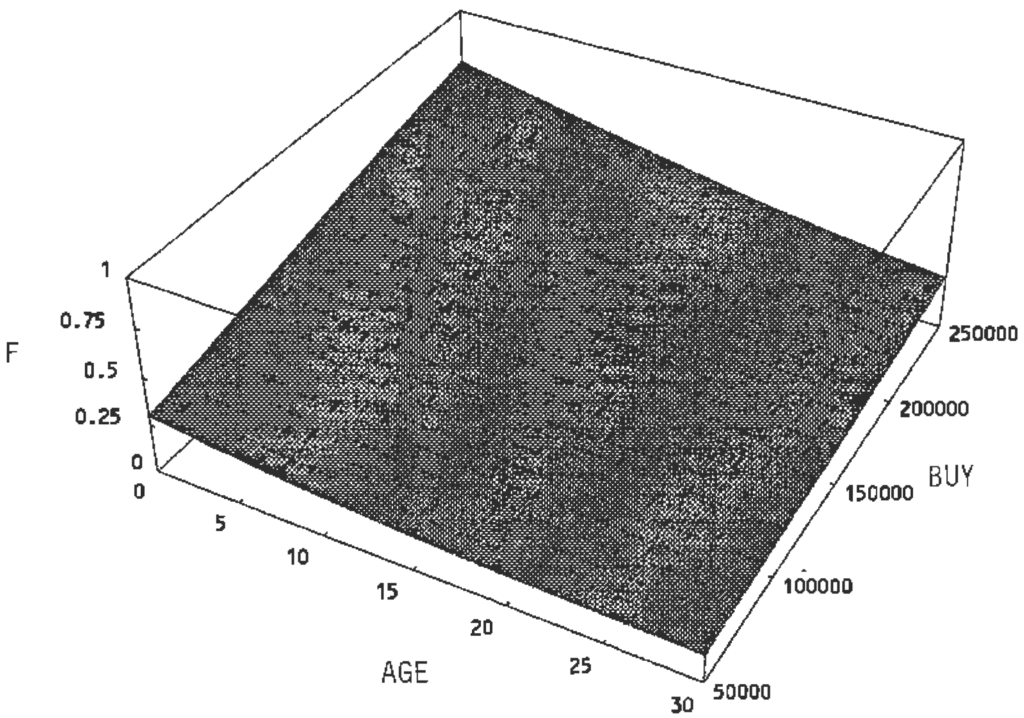


Figure 2 A.

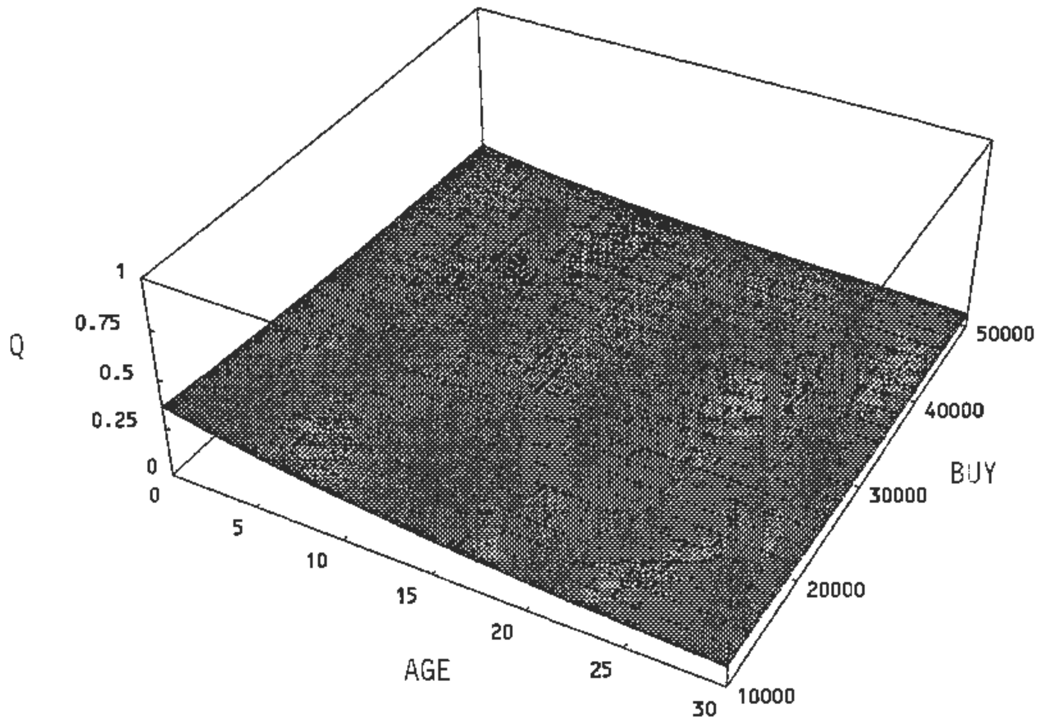
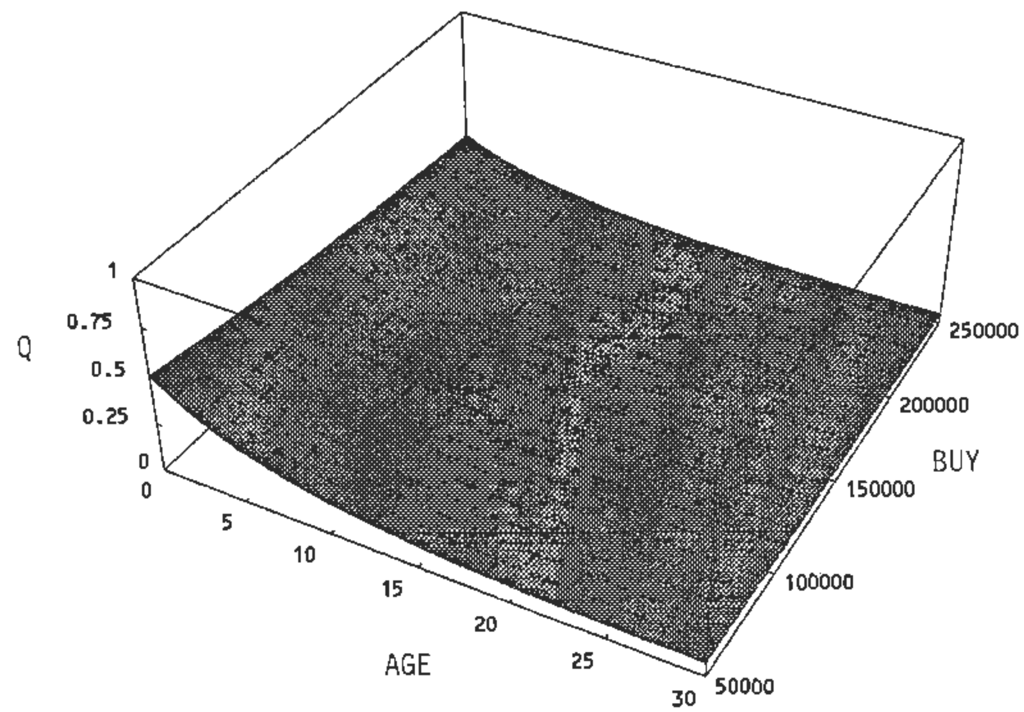


Figure 2 B.



Competition in Interrelated Markets: An Empirical Study*

Marcus Asplund and Rickard Sandin
Department of Economics
Stockholm School of Economics
Box 6501, S-113 83 Stockholm, Sweden
e-mail: nema@hhs.se; ners@hhs.se

Abstract

This paper studies competition in small, concentrated, and interrelated markets. Our data set consists of price information from 543 driving schools in 250 local markets in Sweden, which gives a large sample to test hypotheses on how market structure influences competition. The results show that if prices in nearby markets are low, and the distances to them are short, it reduces prices, as suggested in models of spatial competition. Moreover, we find that prices in closely located markets are interdependent. It is also shown that prices are increasing in firm concentration within a market, as most theories of oligopoly predict.

Keywords: Spatial competition; interrelated markets; oligopoly; driving schools.
JEL specification: C24; C31; D43; L13; L84; R32

* We wish to thank Tore Ellingsen, Per-Olov Johansson, Ulf Gerdtham, Kasper Roszbach and seminar participants at the Stockholm School of Economics and the Industrial Institute for Economic Research (IUI) for valuable comments. We also express our thanks to Hans-Christian Ageman at the Swedish Association of Driving Schools. The first author acknowledges financial support from the Swedish Competition Authority, and the second author from the Institute for Fiscal Policy Research.

1. Introduction

Although oligopoly theory has been successful in generating hypotheses on factors that may affect the intensity of competition, few of them are easy to test empirically. A set of models that have received little empirical attention, despite the fact that they yield testable implications, and for which data are relatively easy to find, are spatial models of competition. In these models the intensity of competition among firms is partly determined by the geographical dispersion of firms, which is easy to observe. Many of the industries where these models are applicable also have relatively homogeneous goods and prices that are readily observable. Our study uses data from a large number of concentrated markets within the same industry to test how prices in nearby markets, and the distance to them affect competition in a given market. At the same time we test how firm concentration within a market affect the intensity of competition.

The theory underlying this approach goes back to the early work of Kaldor (1935) who recognised that firms within a market compete not only with each other, but also, to a lesser extent, with firms in other nearby markets, which has been refered to as "overlapping oligopolies". However, most of the recent theoretical work on spatial competition focus on the interaction between firms at different locations. Factors that have been emphasised are the entry and locational decision of firms, and the distribution and travel cost of consumers. For surveys of this large literature, see Eaton and Lipsey (1989) and Martin (1993).

There are relatively few empirical studies on spatial models of competition. Johnson and Parkman (1983) studies the cement market and finds no significant support for the hypothesis that profitability falls with the geographical concentration of firms. In Cotterill's (1986) study of the retail food industry in Vermont, the distance between warehouses have negligible effects on the price level. Claycombe and Mahan (1993) study beef pricing, but finds no significant effects of commuting distances on prices. Fik (1988), however, finds that the distance to the nearest competitor has a significant effect on price when it is the only explanatory variable in a linear regression. A common feature of these studies is that only distances to, and not prices in nearby markets are considered to influence competition. An exception is Haining (1984), who finds some support for price clustering among neighbouring outlets in urban gasoline retailing. Slade (1986) on gasoline prices, and Horowitz (1986) on prices of meat, use time series data on price differences between markets to detect those that are part of the same geographical market. Even though we feel that this line of research is very interesting, it does not address the same questions as this study. For example, no measure of firm concentration within a market is included, and there is no measure of geographical distance between markets, as suggested by theories of spatial competition.

Much more attention have been assigned to factors that may affect the intensity of competition within a market. There exists a literature which focuses on competition in regional markets within the same industry. Examples of such studies are Geithman, Marvel, and Weiss (1981), and Koller and Weiss (1989) for the case of cement, and Marvel (1989) on gasoline. These studies mostly deal with the relation between prices and market concentration, and typically find a positive relation, Schmalensee (1989 p. 988).

Our choice of industry and market sample permits tests of competition both between and within markets. The results of our study of driving schools in Sweden show that the distances to other closely located markets and, in particular, the prices there play a significant role in explaining the price level in a given market. Moreover, we conclude that prices in nearby markets are endogeneously determined. Our results also show that prices increase with firm concentration within the market.

2. The Data

To undertake a study as outlined in the introduction, we searched for an industry where the product is fairly homogeneous and sold in many markets. We argue that an industry which meets these requirements is driving schools, with the product being a single driving lesson. One might argue that the quality of a lesson varies across firms (vertical product differentiation), but price differences are small within markets in our sample, which one would not expect if there were substantial quality differences. Furthermore, it appears that quality differences between teachers within the same firm are as common as any quality differences between firms, which again might explain the small price variation within markets. The product is horisontally differentiated to consumers at different locations, but probably quite homogeneous to consumers at similar location, as indicated by the low advertising intensity in the industry.

2.1 Markets

In this study a market is defined as a municipal. There are 288 municipals in Sweden of which 250 are included in our sample. We excluded Stockholm, Gothenburg, and Malmoe with surrounding suburb municipals because these three cities are much larger than the rest, and which may each consist of many local markets.[1] Furthermore, it was indicated that these markets had different cost and demand characteristics, as discussed in the Appendix. The markets in our sample differ in size as shown in Table 1. In 1993, the smallest market has a total population, *POPTOT* of only 2.865 while the largest market has a population of 178.011. The distribution is skewed towards smaller markets.

Although total population may describe the size of a market, the most important consumer group for the study of driving schools is the population of age 16-24, which is measured in hundreds, and denoted *POP1624*. We included the number, in hundreds, of students in the gymnasium (≈ high-school), *STUDENTS* as a separate measure of market size. This variable was chosen because there are municipals where there are no gymnasium, and where students must travel to other municipals, where it is likely that they take their driving lessons.

Municipal data on average personal wage income in 1993, *PINC* is used to control for differences in wealth levels that may affect demand. In Table 1 it is measured in SEK hundreds of thousand.[2]

[1] In the municipal of Stockholm, for example, the total population in 1993 was almost 700.000 and there are more than 50 driving schools, which should be compared to the mean population and number of firms in our sample which are 24.930 and 2.2 respectively.

[2] We are referring to the Swedish currency throughout the paper and we omit SEK from now on. In June 1, 1995, one US dollar equals SEK 7.35.

Table 1.– Sample Characteristics

Variable	mean	stand. dev.	min	max	median	number of observations
POPTOT	249	264	28.6	1780	143	250
POP1624	28.9	33.6	2.73	238	15.4	250
STUDENTS	8.29	11.8	0	65.8	3.39	250
PINC	1.23	0.0840	1.03	1.53	1.23	250
DIST1	27.0	17.1	6.00	119	21.6	250
DIST2	37.6	21.9	9.60	144	32.4	250
FIRMS	2.18	2.40	0	13	2	250
MCARS	5.30	6.69	0	33	3	250
FCARS	2.44	1.36	1	9	2	543
MPRICE	5.90	0.529	4.53	8.00	5.92	196
STDEV MPRICE	0.163	0.162	0	0.593	0.125	126
MPRICENEAR1	5.94	0.565	4.53	8.00	5.97	237
MPRICENEAR2	5.93	0.517	4.53	7.14	5.97	235
OFFICEVAL	3.03	0.595	2.23	5.81	2.87	24
WAGE	13.1	0.270	12.6	14.1	13.1	24

As mentioned, we study markets which are not necessarily isolated. In order to account for interdependency of markets we measured the distance (as straight lines) from the central town in each market to the nearest central town of the two closest markets, conditional on there being at least one firm in each of these markets.[3] This gives us two distance variables, *DIST1* and *DIST2,* measured in kilometres. As is evident from Table 1, the range hetween the minimum and maximum distance is large, identifying the fact that some markets are isolated while others are not.

2.2 Firms

The study contains all driver schools in our market sample as of June 1995. All together the data includes 543 firms, where the number of firms in a market is denoted *FIRMS.* Each firm was asked questions about price per lesson and how many minutes this lesson lasted. We

[3] In each municipal there is usually only one central town where the large majority of the population live in and where, with few exceptions, the firms are located. In most cases these towns also lie in the centre of the municipal.

also asked questions about the number of cars (including brand, model and vintage) and the number of teachers in the school working full-time as well as part-time. We deliberately restricted the number of questions to a few, since we believed that more questions would negatively affect the responding frequency. All except one firm answered all our questions, which is also explained by the fact that we did not ask questions on profitability or costs, but on features that are easily observed by any competitor.

Table 2 shows the distribution of markets with different number of firms. As one can see, there is a span of market structures ranging from zero to thirteen firms. In our sample, 207 out of 250 markets contain three firms or less, which by any standard are concentrated markets.

Table 2.– Market Size and Price, Related to the Number of Firms in the Market

FIRMS	*No. of markets*	*Mean POP1624*	*Mean STUDENTS*	*Mean MPRICE*
0	54	9.74	0.612	-
1	70	13.4	2.10	5.95
2	50	21.6	6.17	5.82
3	33	32.9	11.4	5.87
4	13	43.4	15.8	5.88
5	9	61.6	20.9	5.78
6	4	80.1	27.2	6.02
7	4	97.8	31.4	6.12
8	4	134	43.2	5.72
9	2	104	38.9	6.64
10	3	123	43.3	5.92
11	2	148	50.1	6.01
12	0	-	-	-
13	2	162	49.7	5.87

We use the number of cars in a market, *MCARS,* as a measure of total capacity and the number of cars in a firm, *FCARS* as a measure of firm size. From these variables a Herfindahl concentration index is defined as

$$HERFINDAHL_i = \sum_{k=1}^{FIRMS_i} \left(\frac{FCARS_{i,k}}{MCARS_i} \right)^2 .$$

Figure 1 shows a plot of the square root of the Herfindahl index as a function of market size measured by *POP1624*. The figure reveals a strong negative relation between market size and concentration, which could be even stronger if we had accounted for differences in personal income, distances to the closest markets, and other factors that affect demand. In the framework of Sutton (1991), the lower bound of concentration seems to be strictly decreasing in market size, and markets lie relatively close to this bound. This pattern is more likely to obtain when the product is homogeneous, supporting our argument that driving lessons do not differ much between firms and markets.

2.3 Prices

The price and duration of a single driving lesson varies among the firms in our sample. Some firms use second degree price discrimination, where they offer a package of several lessons (usually 5 or/ and 10 lessons) bundled with theory lessons, which gives a slightly lower price per lesson. We choose the price of a single lesson because it is very difficult to compare these packages, and more so because single lessons are most important quantitatively.

We use several different price variables in the paper. The mean market price per minute is denoted *MPRICE* and varies from 4.53 to 8.00, as shown in Table 1. However, there is almost no variation within a market, which is seen by the small standard deviation of *MPRICE* within markets with more than one firm, *STDEVMPRICE*. This can also be illustrated by the high correlation (0.85) between duopolists' prices within a market. There are relatively small differences in mean *MPRICE* for markets with different number of firms, as illustrated in Table 2. This is not surprising, because we have not adjusted for other factors, such as market spill-over effects and cost differences etc. Since prices in closely located markets are expected to influence competition in the market, we include the mean market price per minute in the closest and second closest market where there exists at least one firm, denoted by *MPRICENEAR1*, and *MPRICENEAR2* respectively. Table 3 shows the correlation between our price variables.

Table 3.– Correlation between Price Variables

Variable	*MPRICE*	*MPRICENEAR1*
MPRICENEAR1	0.557	1
MPRICENEAR2	0.423	0.320

As expected, *MPRICE* is more correlated with the price in the closest market, than with that in the second closest market. Moreover, both of these correlations are higher than between *MPRICENEAR1* and *MPRICENEAR2*, which may be on different sides of the market, and thus further away from each other.

2.4 Costs

The principal costs for driving schools are wages for the teachers, the cost of cars, gasoline, traffic insurance, and office and parking space. Wage costs are dominating. There is a central agreement on the wages of the teachers who are unionised, but according to their trade union, some regional differences exist primarily between Stockholm, Gothenburg, and Malmoe, and the rest of the country although there are no available data on this. Since these cities are excluded from our sample, regional differences are probably less important. Moreover, wage levels are traditionally quite compressed in Sweden. Still, to track potential wage differences in our sample we use county level data on monthly salary of skilled employees in the trade and commerce sector, measured in thousands, *WAGE*, as a proxy.

In order to account for cost differences in the use of office and parking space, we used the average assessed value of per square metre non-residential floor space in 1994, measured in thousands, *OFFICEVAL*. Again, only county level data are available, which means we cannot measure possible differences among municipals within a county. The remaining cost variables were excluded in our regressions because regional differences are very small.[4]

3. Econometric model

In the preceding section we informally discussed specificities considered to be important for competition in this particular industry. To reiterate, we expect the product - a single driving lesson - to be relatively homogeneous across firms within a market. However, the product is horizontally differentiated to consumers at different locations, who have the possibility to travel to nearby markets if prices there are significantly lower, and distances are short. Cost structures of firms and across markets are similar, and some fixed costs exist. There are many theoretical models that can be used to describe these features, but much of the insights can be found in a simple spatial model with three firms, serving consumers who are uniformly distributed on a line (see Eaton and Lipsey (1989) and Martin (1993) for discussions and references). Let Firm 1 and Firm 2 be Firm 0's closest neighbours, located L_1 and L_2 away, and let C_0 be its constant marginal cost, and t be the unit travel costs of consumers. It is a standard exercise to solve the first order condition for Firm 0 to obtain the best response function (under certain assumptions to guarantee existence)

$$P_0 = \frac{P_1 + P_2 + C_0(2 - \frac{dP_1}{dP_0} - \frac{dP_2}{dP_0}) + t(L_1 + L_2)}{4 - \frac{dP_1}{dP_0} - \frac{dP_2}{dP_0}} \tag{1}$$

Firm 0's price is increasing in the other firms' prices, the distance to them, and its own marginal cost. However, the magnitude is dependent upon the assumptions about the derivatives of dP_i / dP_0, which are the conjectural variation parameters (e.g. $dP_i / dP_0 = 0$ is

[4] See Appendix for detailed motivations.

Cournot conjectures and $0 < dP_i / dP_0 < 1$ are less aggressive conjectures).[5] The price P_0 can be reinterpreted as the collusive price for firms in market 0, given the prices in nearby markets and equal marginal costs, C_0. However, from static and dynamic oligopoly theory we know that this need not be the price in markets with more than one firm. For example, it may not be sustained in an infinitely repeated game, where more firms may lead to lower prices (see e.g. Tirole (1992)). We model this by introducing the number of competitors and concentration in the market, as independent variables.

As a general matter of specification, it can be argued that it is more appropriate to estimate a reduced form for the equilibrium price (the Nash equilibrium), rather than an equation based on (1). However, in a model with many markets located on a line, the equilibrium price in a market is a function of the all characteristics of all markets. This is clearly impossible to estimate.

The first point to note in the econometric specification, is that there are no firms in 54 out of 250 markets. It is unlikely that this is a random sample, as these markets may be small, lie close to other markets, or have high costs. To correct for this sample selection we use a Tobit II model (see Amemiya (1985)), which is estimated by the two-stage procedure of Heckman (1979). In the Tobit II model, the latent variable *FIRMS** determines where there exists at least one firm through the selection mechanism

$$FIRMS_i^* = \gamma W_i + \mu_i. \tag{2}$$

$$FIRMSPOS_i = 1 \qquad \text{if} \qquad FIRMS_i^* > 0,$$

$$FIRMSPOS_i = 0 \qquad \text{if} \qquad FIRMS_i^* \leq 0,$$

where γ is a vector of parameters to be estimated and W is a vector of exogenous variables. In the data, *FIRMSPOS=1* if there is at least one firm in the market, and *FIRMSPOS=0* if there are no firms. The mean price per minute, *MPRICE*, is determined by the regression model

$$MPRICE_i = \beta X_i + \varepsilon_i \qquad \text{if} \qquad FIRMSPOS_i = 1, \tag{3}$$

[5] It is well known that conjectural variation models can be criticised on several grounds, but they still provide a useful method to analyse different degrees of oligopolistic competition in a unified framework, see e.g. Dixit (1986).

where β is a parameter vector and X are explanatory variables. The distribution of the error terms, μ and ε, is assumed to be bivariate normal, $\mu,\varepsilon \sim BVN(0,0,1,\sigma_{\varepsilon}^{2},\rho)$. The expected price, conditional on there being firms in the market, is[6]

$$E[MPRICE_i \,|X_i, FIRMSPOS_i = 1] = \beta X_i + \sigma IMR_i, \tag{4}$$

where IMR is the inverse Mills ratio equal to

$$IMR_i = \frac{\phi(\psi_i)}{\Phi(\psi_i)}, \qquad \psi_i = \gamma W_i, \qquad \sigma = \rho\sigma_{\varepsilon}, \tag{5}$$

and where $\phi(.)$ $(\Phi(.))$ is the (cumulative) normal density function.

In the first stage we use a Probit model to estimate the probability that there exists at least one firm in the market. In the second stage we estimate prices, using only the markets which have at least one firm, with the estimated IMR as an independent variable. For markets with at least one firm we use the specifications[7]

$$\begin{aligned}
&MPRICE_i = \beta_0 + \\
&\beta_1 MONOPOLY_i + \beta_2 DUOPOLY_i + \beta_3 TRIOPOLY_i + \beta_4 QUADROPOLY_i + \\
&\beta_5 MPRICENEAR1_i + \beta_6 MPRICENEAR2_i + \\
&\beta_7 DIST1_i + \beta_8 DIST2_i + \beta_9 DIST1SQR_i + \beta_{10} DIST2SQR_i + \\
&\beta_{11} PINC_i + \beta_{12} WAGE_i + \beta_{13} OFFICEVAL_i + \sigma IMR_i + \varepsilon_i,
\end{aligned} \tag{6a}$$

and

$$\begin{aligned}
&MPRICE_i = \beta_0 + \\
&\beta_1 HERFINDAHL_i + \\
&\beta_2 MPRICENEAR1_i + \beta_3 MPRICENEAR2_i + \\
&\beta_4 DIST1_i + \beta_5 DIST2_i + \beta_6 DIST1SQR_i + \beta_7 DIST2SQR_i + \\
&\beta_8 PINC_i + \beta_9 WAGE_i + \beta_{10} OFFICEVAL_i + \sigma IMR_i + \varepsilon_i.
\end{aligned} \tag{6b}$$

It is likely that the *MPRICENEAR* variables are endogenous. For example, if the price in market *1* is increased, it will lead to a higher price in market *0*, which in turn leads to an even

[6]For derivations see e.g. Greene (1993) or Amemiya (1985).
[7]We also experimented with a measure of capacity utilisation. This variable was constructed as the number of cars in a market (which is a relatively fixed factor) minus the number of teachers (which can be hired on a part time basis to cover high demand periods), normalised with the former. However, this variable was too crude to yield any effect in our regressions.

higher price in market *1*, and an additional increase in price in market *0*, and so forth. A least squares (LS) estimator does not account for this interdependence, why we use a two-stage least squares (2SLS) estimator. With this 2SLS estimator we avoid the impossible task of specifying the full system of equilibrium prices, and need only to find instrumental variables for the prices in the nearest markets. If e.g. markets *j* and *k* are the nearest, we use all exogenous variables which correspond to (6a) in these markets as instruments for the price variables.

4. Results

Table 4 presents the results from the regressions with *FIRMSPOS* and *MPRICE* as dependent variables. As expected, the Probit regression shows that firms are less likely to operate in small markets (low *POP1624* or few *STUDENTS*), where *PINC* is low, and where *WAGE* is high. However, the distance to the nearest market had no significant effect, and the same holds for *OFFICEVAL*.[8] The predictive ability of the model is satisfying (83.2% compared to the naive 78.4%). The absence of firms in some smaller markets indicates the existence of some fixed cost, which need to be covered in equilibrium.

The LS and 2SLS regressions are based on 181 markets, where neither of the two closest markets are suburbs to Stockholm, Gothenburg, or Malmoe. The coefficients in the LS and 2SLS estimators are similar except for *MPRICENEAR1* and *MPRICENEAR2*. To test for endogeneity of these variables we conducted a Hausman-Wu test, which showed that endogeneity could not be rejected at the 10% level (P-value 0.085 and 0.070 in (6a) and (6b) respectively). We therefore focus on the 2SLS estimates in discussing the parameters.

The results support the hypothesis that competition between firms in different markets is important. *MPRICENEAR1* and *MPRICENEAR2* are positive and highly significant in all specifications, and the point estimates are high. Thus, if prices in the nearest markets are high,

[8] We also experimented with the number of firms in the nearest markets, and the distance to the closest markets where there are no firms, but these variables had no explanatory value.

this will tend to increase the price in a given market. Moreover, the sum of the two coefficients is higher in the 2SLS than in the LS (0.96 and 0.72 respectively). The LS underestimates the full effect of price competition, since it does not account for the fact that prices are interdependent. For example, if prices rise by $1 in both the nearest markets, the direct effect is an increase in price by $0.72. However, this induces a further increase in price by firms in the nearest markets, and so forth, to give a total price increase of $0.96. It must be noted that there is an alternative explanation to why the coefficients of *MPRICENEAR1* and *MPRICENEAR2* are positive and significant in the LS. If there are significant cost differences between markets, which are not captured by the included cost variables, then it is possible that the *MPRICENEAR* variables work as proxies for these left out variables. If unmeasured costs are high in the three closely located markets *1, 2*, and *3* and low in markets *4*, *5*, and *6* also lying close to each other, then prices will naturally be higher in the first three markets. The coefficients of the *MPRICENEAR* variables will then be positive since they capture the effect of costs differences on prices. However, we argue that this is not the main effect. First, in a regression where we included regional (county) dummies, and market type dummies (e.g. urban, farm area, less densely populated area), the coefficients of *MPRICENEAR1* and *MPRICENEAR2* were similar, and the restriction that all dummies are zero could not be rejected (pseudo F(27,143)=0.907 with P-value 0.600, Startz (1983)). We find it plausible that these regional, and market type characteristics are much more influential in determining costs than very local variations that affect only a small region. Discussing the matter with the main interest organisation of the firms, they clearly indicated that such very local variations are likely to be uncommon, and small in magnitude where they exist. Moreover, this would be an exogenous effect, but our Hausman-Wu test rejected that *MPRICENEAR1* and *MPRICENEAR2* are exogenous. The strong effect we find of prices in nearby markets raises questions on previous studies, where these variables are omitted e.g. Cotterill (1986), Koller and Weiss (1989), and Marvel (1978). If markets in these studies are not isolated, there may be a serious omitted variable bias, making the interpretation of the results difficult. Studies where markets were explicitly selected to be isolated, as in Bresnahan and Reiss (1991), are not sensitive to this criticism.

In the reported specifications we excluded *DIST2* and *DIST2SQR* because they were highly collinear with *DIST1* and *DIST1SQR*. The distance measures are neither individually, nor jointly significant. However, this test may not be the most interesting. A better way would be to test if the distance variables can be restricted both as regressors and as instrumental variables. We are not aware of any such formal test, but if distance variables are redundant, they will not have any explanatory power in a reduced form regression with *MPRICE* as dependent variable and all exogenous variables as explanatory variables. Even though this is not a formal test, the restriction that all distance variables are zero is strongly rejected ($F(6,152)=3.75$ with P-value 0.002). Based on this we conclude that the distance variables are important to include as explanatory variables in the regressions.[9] A robust finding in all our regressions is that the point estimate of *DIST1* is positive and *DIST1SQR* is negative, implying a concave effect of geographical distance on price, and not a linear effect as suggested by (1). Hence, the further away the nearest market is - the higher the price, but the effect decreases more than proportionally with distance. As a numerical example of the magnitude of the distance effect in (6a), a market where the closest market is 40 km away has about 0.07 higher price compared to a market where it is 10 km away (the effect of distance in *IMR* is negligible). This is comparable to the quadropoly coefficient of 0.06. Other studies that have included a measure of geographical distance find a positive, although not always significant effect on prices or profits, e.g. Cotterill (1986), Collins and Preston (1969), Fik (1988), and Johnson and Preston (1983).

In (6a), *MONOPOLY* is positive and highly significant in both the LS and the 2SLS regressions. The coefficients of the dummy variables are decreasing in the number of firms, as expected from theory. We tested restrictions on the market structure parameters, as shown in Table 5. It can be rejected that all market structure coefficients are equal, and that monopolies, duopolies, and triopolies have the same effect on price. It can not, however, be rejected that monopolies and duopolies have the same coefficient.

[9] Moreover, when the distance variables were excluded both as regressors and instruments, the 2SLS estimates of the other variables appeared to be quite different from those reported in Table 4.

Many previous studies have used some measure of concentration and found it to be positive and significant in explaining price, (see Schmalensee (1989 p. 987-8) for references). This also holds in (6b) where we use *HERFINDAHL* as a measure of concentration.[10] Much of its significance is due to the effect of monopolies, but it is significant even when monopolies are excluded.[11] The conclusion from (6a) and (6b) is that even though prices tend to be lower in markets with more firms, and in less concentrated markets, they are not decreasing steeply. Of course, this is consistent with many theories of oligopolistic behaviour, and we do not make any strong statement on which of these that is the most plausible.

The cost variables *WAGE* and *OFFICEVAL* are insignificant, and the former even has a negative sign. Again, it is an indication that costs differentials are not very important in determining prices. On the other hand, *PINC* is positive. As noted above, *PINC* can be interpreted either as a cost parameter or a demand parameter, and it is not possible to distinguish between them a priori. We favour the latter interpretation, since it is positive and significant in the Probit regression, where it should have been negative if it had measured costs.

Although insignificant, *IMR* is negative in both (6a) and (6b). This means that prices tend to be lower in markets where it is unlikely to observe a firm, but where there exists at least one firm. Failure to use information about markets where there are no firms, may bias the estimated parameters of the other variables. For comparison we report a regression with *IMR* excluded both as a regressor, and as instruments. The significance levels for the coefficients of *MPRICENEAR1*, *MPRICENEAR2* and *MONOPOLY* are lower, whereas *PINC* is significant. The point estimates also differ for most coefficients, suggesting the importance of including *IMR* in the specification.

To get another look at the price-concentration relation we estimated (6b) for different subsamples, based on the number of firms in the market. The results in Table 6 reveal that there is no effect of *HERFINDAHL* if we look only at markets with a given number of firms. This is due to the small variation in the variable (for a given number of firms), since firms are

[10]The same holds if we instead use the square root of the Herfindahl index. The coefficient is then 0.543*** (0.206), and the coefficients of the other variables are essentially the same as in (6b).
[11]Based on the 114 observations with two or more firms the coefficient is 0.563* (0.333).

of roughly the same size. *MPRICENEAR1* remains significant, and the point estimate is high. Even though *MPRICENEAR2* is positive in all subsamples, it is significant only for triopolies. *DIST1* is positive and *DIST1SQR* is negative in all subsamples, as in Table 4, but point estimates are largely insignificant. One further point to note is that adjusted R^2 are higher for markets with three or more firms. The very poor fit for duopolies partly explains why *DUOPOLY* did not show up significant in (6a). There is no obvious explanation to why it is much easier to estimate the price level in markets with three firms or more, than in duopoly markets.

5. Concluding remarks

Our sample of small firms, who provide a relatively homogeneous product in local, but not isolated markets is close to the textbook example of spatial competition. The econometric model used in this paper did not explicitly test any specific theory, but rather the main implications of spatial oligopoly models. The reason for this is simply that in the wide array of models, many of the results are dependent on specific, but unobservable, details e.g. whether travel costs are linear or quadratic, or if the pattern of entry is sequential or simultaneous. However, the general conclusion is that our results give support for the main hypotheses in this set of theories. We showed that high prices in nearby markets give a higher price level, and that prices in closely located markets are interdependent. Moreover, there is some support for the hypothesis that geographical distance between markets affect price levels.

Our future work in this area will be to use time series, as well as cross sectional data, to test more specific theories about spatial oligopoly. As a part of this extension we will use the accounting rates of return for the companies in the sample, to study the relation between prices and profitability. Although we feel that cost differences between firms and local markets in this study are small, there is scope for improvement. A challenge for future research is to find firm level cost data for some markets that fits into the spatial oligopoly framework.

6. References

Amemiya, T., 1985, Advanced Econometrics, Cambridge Mass. Harvard University Press.

Bresnahan, T.F. 1989, Empirical studies of industries with market power, in R. Schmalensee and R. Willig, eds., Handbook of Industrial Organization, Vol. 2 (Elsevier Science Club) 1011-1058.

Bresnahan, T.F. and P.C. Reiss, 1991, Entry and Competition in Concentrated Markets, *Journal of Political Economy*, 99, 977-1009.

Claycombe, R.J., 1991, Spatial retail markets, *International Journal of Industrial Organization*, 9, 303-313.

Claycombe, R.J., and T.E. Mahan, 1993, Spatial aspects of retail market structure: Beef pricing revisited, *International Journal of Industrial Organization*, 11, 283-291.

Collins, N.R., and L.E Preston, 1969, Price-cost margins and industry structure, *The Review of Economics and Statistics*, 51, 271-286.

Cotterill, R.W., 1986, Market power in the retail food industry: Evidence from Vermont, *Review of Economics and Statistics*, 68, 379-386.

Culbertson, W.P., and D. Bradford, 1991, The price of beer: Some evidence from interstate comparisons, *International Journal of Industrial Organisation*, 9, 275-289.

Dixit, A.K., 1986, Comparative Static's for Oligopoly, *International Economic Review*, 27, 107-122.

Eaton B.C., and R.G. Lipsey, 1989, Product differentiation, in R. Schmalensee and R. Willig, eds., Handbook of Industrial Organization, Vol. 1 (Elsevier Science Club), 723-768.

Fik, T.J., 1988, Spatial competition and price reporting in retail food markets, *Economic Geography*, 64, 29-44.

Geithman, F.E., Marvel, H.P., and L.W. Weiss, 1981, Concentration, price, and critical, concentration ratios, *Review of Economics and Statistics*, 63, 346-353.

Greene, W.H., 1993, Econometric Analysis, New York; Macmillan.

Haining, R., 1984, Testing a spatial interacting-markets hypothesis, *Review of Economics and Statistics*, 4, 576-583.

Hay, D.A., and D.J. Morris, 1991, Industrial economics and organisation. Theory and evidence, Oxford University Press, Oxford.

Heckman, J., 1979, Sample selection bias as a specification error, *Econometrica*, 47, 153-161.

Johnson, R.N., and A. Parkman, 1983, Spatial monopoly, non-zero profits and entry deterrence: The case of cement, *Review of Economics and Statistics*, 65, 431-439.

Kaldor, N., 1935, Market imperfections and excess capacity, *Economica*, 2, 33-50.

Koller II, R.H., and L.W. Weiss, 1989, Price levels and seller concentration: The case of Portland Cement, in L.W. Weiss ed., Concentration and price, MIT Press, 17-40.

Martin, S., 1993, Advanced Industrial Economics, Blackwell, Cambridge, Mass., Oxford

Marvel, H.P., 1978, Competition and price levels in the retail gasoline market, *Review of Economics and Statistics*, 60, 252-258.

Marvel, H.P., 1989, Concentration and price in gasoline retailing, in L.W. Weiss ed., Concentration and price, MIT Press, 179-185.

McDonald, J., and R. Moffit, 1980, The use of Tobit analysis, *Review of Economics and Statistics*, 62, 318-321.

Schmalensee, R., 1989, Inter-industry studies of structure and performance, in R.Schmalensee and R. Willig, eds., Handbook of Industrial Organization, Vol. 1 (Elsevier Science Club), 951-1009.

Slade, M.E., 1986, Exogeneity tests of market boundaries applied to petroleum products, *Journal of Industrial Economics*, 34, 291-303.

Startz, R., 1983, Computation of linear hypothesis tests for two stage least squares, *Economics Letters*, 11, 139-141.

Sutton, J., 1991, Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration, The MIT Press, Cambridge Mass. and London.

Tirole, Jean 1992, The theory of industrial organization, MIT Press, London, England.

Weiss, L.W. 1989, Concentration and price, MIT Press, Cambridge Mass. and London.

Appendix

Data employed in statistical analysis

1. *Market variables*

1993 data on the total population of age 16-24 across municipals are collected from the Association of Local Authorities. Source: Kommunförbundet. Municipal data on the average number of students in the gymnasium for the academic years of 92/95 were collected from the National Agency for Education. Source: Skolverket, 1995, Skolan - Jämförelsetal för skolhuvudmän, Skolverkets rapport nr 73, table 2, Solna.

Municipal data on yearly average personal wage income in 1993 was collected from Statistics Sweden. Source: Statistiska Centralbyrån, 1995, Inkomst- och skattestatistik 1993, Statistiska meddelanden, Be20 SM 9501, table 14. We also used municipal data on average personal capital income in 1993, which was collected from the same source. The variable did not turn out significant in any of our regressions, and we therefore excluded it.

In addition to the variables measuring the distance to the closest markets conditional on there being at least one firm, we also measured the distance to the three closest markets unconditional on there being any firms. However, these variables are not included in our final regressions.

2. *Firms*

The firms were initially located from local phone books (yellow pages). A member list from the main interest organisation of the companies, the Swedish Association of Driving Schools, completed our set. Source: Sveriges Trafikskolors Riksförbund, 1995, Matrikel. Phone calls to all firms were made from the beginning of June to the middle of August in 1995. We reached about 95% of the firms within the first two weeks. The remaining firms were harder to reach, hence explaining the relatively long period of data collection. Our sample of 543 firms includes a few firms that are branches of firms in other nearby markets. A branch is counted as a separate entity if it uses different cars and teachers than the main school. If the same cars and teachers are used, the school is placed in the municipal where, according to the company, the principal operation is located, unless it is possible to divide the entity into separate entities on the basis of share of demand. Hence, a firm operating in two separate markets with three cars and three employed teachers altogether, is divided into two separate entities containing two and one car(s)/teacher(s) respectively if the firm stated that one of the markets accounted for approximately 67% of its total demand. It should be noted that this question of definition is relevant to less than ten firms.

3. *Costs*

i) Wages

There are no data available on wages for teachers employed in driving schools. Regional wage differences may exist, but is of less importance when Stockholm, Gothenburg, and Malmoe are excluded. Through the central agreement between the employer and employee unions, wage differentials are recommended on basis of education and the number of years of experience of the employee. It states that the wage should be increased by approximately SEK 300 every third year of employment. Data on skills levels in each firm has not been collected in this study. There are no clauses on regional wage differences. County level data on monthly salaries for skilled employees in the trade and commerce sector in 1990 were used as a proxy for the wages of the employed teachers. This is the last year for which a complete survey of wages and employment in the private sector was undertaken. Data was collected from Statistics Sweden. Source: Statistiska Centralbyrån, 1991, Löner och sysselsättning i den privata sektorn 1990, table 9.

Another possibility would be to use the income variable *PINC* as a wage proxy. However, there are reasons to believe that the correlation between these variables is less strong due to the fact that markets differ in the composition of industries. Some municipals are heavily dependent on a few industries (e.g. mining and forest industries in the northern part of Sweden). Furthermore, using the income variable as a wage proxy at the same time as it is used to predict consumer demand is less suitable, since a high income simultaneously corresponds to high costs and high demand, which affects the number of firms in the market in an ambiguous way.

ii) office and parking space

County level data on the average assessed value of non-residential floor space per square metre in 1994 was collected from Statistics Sweden. This includes space used not only for driving schools. Source: Statistiska Centralbyrån, 1995, Rikets Fastigheter 1994 (1), Statistiska meddelanden, Bo 38 SM 9501, table 7a.

iii) cars

Firms use different cars and one would therefore expect cost differences among firms. We collected data on the cars each firm used, including model and vintage. We ordered them in three classes according to their cost or present value, as stated by the Swedish variant of the "Blue Book", published by the Swedish Automobile Association. The three classes were used as dummy variables in our regressions, but did not have any explanatory power. The interest organisation of the firms claim that these cost differences are of minor importance, which was another reason for why we decided to drop them from our final regressions.

iv) traffic insurance

The amount of traffic insurance paid for cars differs depending on how expensive the car is and whether the firm is located in a large town or not. Insurance companies also provide bonus systems, giving a higher bonus to companies who have a low degree of injury. Information from a major insurance company indicate that possible differences among the firms are negligible, mainly because fees are very low (they vary from SEK 1 200 to 2 300 per year and car). We therefore omitted the variable.

v) Gasoline

Gasoline cost vary across the country and over time. It is generally cheaper in large cities than in non urban areas. However, the difference in list prices of a major chain is at most 0.02 per litre. Since during a lesson of 40 minutes one rarely drives more than 20 kilometres, the difference between the firms per lesson is negligible. Even at a widely exaggerated price difference of 1 SEK per litre, the cost difference is only about 2 SEK out of a price of 240 SEK. We therefore chose to omit this cost variable from our regressions.

4. State regulation

Current regulation state that one can start practising driving at the age of 16 and receive licence not before the age of 18. Taking driving lessons in a school is not the only way driving can be practised. A person can take lessons privately from a person who is not an authorised teacher. The only requirement of this non authorised person is that he/she is of age 25 or older and has had the driving license for at least 5 years. These lessons are quite common and are thereby a source of competition. The share of these lessons vary across the country, although it seems to be more common in large towns. The quality of these lessons are much lower compared to lessons offered by schools. This can be seen in statistics on driving exams provided by the Swedish National Road Administration, which show that the percentage failing is 30 percent on average by students taught in schools, while the corresponding number for students taught by non authorised persons is more than 60 percent. Source: Vägverket.

Table 4.– Regression Results

Variable	Probit *FIRMSPOS*	LS[a] *MPRICE*	2SLS[a] *MPRICE*	2SLS[a] *MPRICE*	2SLS[a] *MPRICE*
CONSTANT	7.03	2.45	1.34	1.87	1.34
	(7.11)	(1.71)	(2.35)	(2.42)	(2.38)
MONOPOL		0.311***	0.329***	0.262**	
		(0.111)	(0.121)	(0.105)	
DOUPOL		0.194*	0.196	0.164	
		(0.109)	(0.127)	(0.136)	
TRIOPOL		0.0915	0.105	0.120	
		(0.0960)	(0.0949)	(0.0932)	
QUADROPOL		0.0616	0.0568	0.0721	
		(0.121)	(0.131)	(0.137)	
HERFINDAHL					0.375***
					(0.145)
MPRICENEAR1		0.453***	0.468**	0.380	0.465**
		(0.0720)	(0.199)	(0.231)	(0.193)
MPRICENEAR2		0.267***	0.492**	0.581*	0.508**
		(0.0744)	(0.243)	(0.323)	(0.231)
DIST1	0.0114	0.959 E-02*	0.564 E-02	0.545 E-02	0.505 E-02
	(0.0279)	(0.538 E-02)	(0.546 E-02)	(0.590 E-02)	(0.527 E-02)
DIST1SQR	0.211E-04	-0.103 E-03**	-0.679 E-04	-0.588 E-04	-0.613 E-04
	0.301E-03	(0.470 E-04)	(0.511 E-04)	(0.555 E-04)	(0.470 E-04)
POP1624	0.0813**				
	(0.0366)				
STUDENTS	0.187**				
	(0.0804)				
PINC	4.65**	0.882*	0.781	1.02*	0.673
	(2.07)	(0.510)	(0.524)	(0.527)	(0.459)
WAGE	-1.12*	-0.175	-0.188	-0.258	-0.185
	(0.585)	(0.139)	(0.172)	(0.173)	(0.172)
OFFICEVAL	0.306	0.0334	0.0443	0.0658	0.0451
	(0.255)	(0.0584)	(0.0754)	(0.0787)	(0.0718)
IMR		-0.218	-0.197		-0.208
		(0.167)	(0.189)		(0.178)
Number of observations	250	181	181	181	181
adj. R^2		0.380	0.361	0.334	0.368
Log *L*	-81.3				
Percent correct predictions[b]	83.2				

Standard errors in parenthesis.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

a Numbers in parenthesis are heteroskedastic consistent standard errors.

b Percent positive observations in sample is 78.4.

Table 5.—Joint Tests on Market Structure Variables (p-values)

MON = DOUP	0.191	*MON = TRIOP*	0.058	*MON = QUAD*	0.093
DOUP = TRIOP	0.422	*DUOP = QUAD*	0.371	*TRIOP = QUAD*	0.756
MON = DUOP = TRIOP	0.051	*DUOP = TRIOP = QUAD*	0.302	*MON = DUOP = TRIOP = QUAD*	0.032

Table 6.—Regression Results for Subsamples

	2SLS			
Variable	*MPRICE* (FIRMS=1)	*MPRICE* (FIRMS=2)	*MPRICE* (FIRMS=3)	*MPRICE* (FIRMS≥4)
CONSTANT	0.907	0.440	-3.01	-1.19
	(1.54)	(2.22)	(1.35)	(1.60)
HERFINDAHL		-0.171	-0.590	-0.159
		(1.72)	(1.30)	(0.965)
MPRICENEAR1	0.407**	0.498**	0.536***	0.750***
	(0.181)	(0.203)	(0.0882)	(0.187)
MPRICENEAR2	0.113	0.819E-02	0.638***	0.131
	(0.260)	(0.164)	(0.117)	(0.208)
DISTI	0.0227	0.357E-02	0.0129	0.0165
	(0.0170)	(0.0148)	(0.0104)	(0.0179)
DISTISQR	-0.230E-03	-0.386E-04	-0.263E-03*	-0.287E-03
	0.185E-03	(0.114E-03)	(0.134E-03)	(0.256E-03)
PINC	1.15	2.48**	2.37***	1.05
	(0.783)	(1.10)	(0.912)	(0.939)
OFFICEVAL	0.0627	-0.171	-0.325***	0.103
	(0.0765)	(0.109)	(0.114)	(0.0847)
Number of observations	67	46	32	36
adj. R^2	0.298	0.129	0.774	0.504

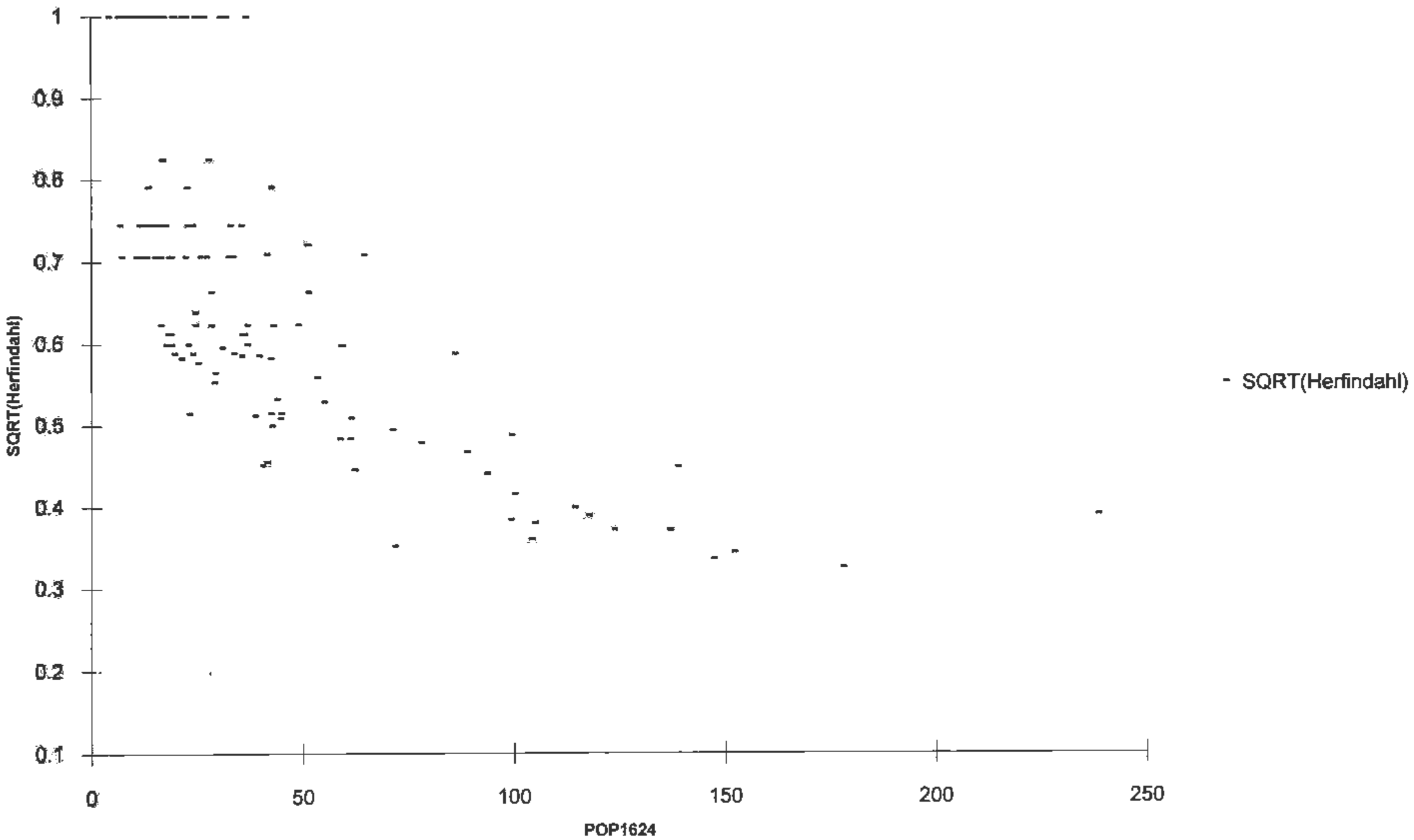
Numbers in parenthesis are heteroscedastic consistent standard errors.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

FIGURE 1.


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
SQRT(Herfindahl)
0
50
100
150
200
250
POP1624
SQRT(Herfindahl)

Estimating The Number of Firms and Capacity in Small Markets*

Marcus Asplund and Rickard Sandin

Department of Economics

Stockholm School of Economics

Box 6501, S-113 83 Stockholm, Sweden

e-mail: nema@hhs.se; ners@hhs.se

Abstract

Many oligopoly theories predict that there will be a positive correlation between market size and the equilibrium number of firms, and some also imply that competition is more intense in larger markets. We test these predictions with a sample of 535 driving schools in 249 markets. With an ordered Probit, a Tobit, and a Poisson model we estimate the relation between the number of firms, capacity, and market size. We find a strong positive correlation between market size and the number of firms. The results show that the per firm market size is increasing in the number of firms in the market. The market size per capacity unit is smaller in large markets. Since the industry produces a fairly homogenous good, we argue that this is evidence that competition is increasing in market size.

Key words: Industry structure; capacity; entry thresholds; count data; driving schools.

JEL specification: C24; C25; D43; L11; L13; L89; R32.

---

* We wish to thank Tore Ellingsen, Ulf Gerdtham, Per-Olov Johansson, Kasper Roszbach, seminars participants at the Stockholm School of Economics, and the Industrial Institute for Economic and Social Research (IUI) for helpful comments. We also express our thanks to Hans-Christian Ageman at the Swedish Association of Driving Schools. The first author acknowledges financial support from the Swedish Competition Authority, and Tom Hedelius and Jan Wallanders Foundation for Economic Research, and the second author from the Institute for Fiscal Policy Research.

1. Introduction

Many oligopoly theories predict that the number of firms is positively correlated with market size. If firms need to cover exogenous sunk costs with variable profits, then the demand in some small markets may be insufficient to support any firm. At some larger market size the demand will be sufficient for one firm, and for still larger markets there may be room for two or more firms. The exact relation between the number of firms (or concentration) and market size will in general depend on the magnitude of sunk costs, and the intensity of competition once firms have entered. Studies which compares concentration in a given industry between countries (≈markets) tend to find that concentration is higher in small countries, Caves (1989 p.1230-5), but see discussion in Curry and George (1983). Even though the negative relation between market size and concentration can be expected to hold for markets with exogenous sunk cost, a non-monotonic relation may be found in markets with differentiated products and endogenous sunk cost investments in advertising or R&D, Sutton (1991).

An econometric model that relates market size to the number of firms in homogenous goods industries is proposed by Bresnahan and Reiss in a series of papers (1988, 1990, and 1991). They suggest that one can draw inferences on the intensity of competition from the relation between the number of firms in the market and market size. The general idea is that if competition is increasing in the number of firms, then the minimum per firm market size, denoted per firm entry threshold, has to be increasing for firms to cover fixed cost. For example, if the smallest market size necessary to support one firm is equal to S, then the market must be greater than $2S$ to support two firms if competition reduces profits. Bresnahan and Reiss (1991), henceforth BR, find that estimated per firm entry thresholds are increasing in the number of firms for several retail and professional industries in the United States.

Our sample consists of information on 535 driving schools in 249 markets in Sweden. This is a "simple" industry where firms generally produce only one, relatively homogenous product, with the same technology. It is therefore reasonable to compare markets with respect to the number of firms, and to estimate the per firm entry thresholds, as in BR. In addition, the data includes information about firm capacity, measured by the number of cars in each firm. We extend the analysis of BR by estimating per capacity unit entry thresholds, defined as the minimum market size per car. The conjecture is that if competition is more intense, and prices are lower in larger markets, the market size per capacity unit will be smaller in those.

Several econometric techniques are used to estimate these thresholds. In general, we find that estimated entry thresholds are insensitive to the applied econometric specification. Results show that per firm entry thresholds are increasing in the number of competitors. Together with the finding that per capacity unit entry thresholds are decreasing in the number of capacity units (at least for small markets), we argue that competition intensifies with market size. Since we reached a similar conclusion in Asplund and Sandin (1995a), where we instead used market price as the dependent variable, we argue that information on the distribution of firms across markets, and their size distribution, can be used to predict competition, at least in homogenous goods industries.

2. The Data

Our study of the driving school industry consists of 535 firms located in 249 markets, where a market is defined as a municipal. The four largest cities (Stockholm, Gothenburg, Malmoe, and Uppsala) with surrounding suburb municipals are excluded from our sample because they are substantially larger than the rest, which makes it likely that each of them consists of many local markets. Definitions and summary statistics of our variables are presented in Table 1 and Table 2. We use total population of age 16-24, denoted *POP*, as our primary measure of market size, and also a dummy variable for whether the number of students in the gymnasium (≈ high-school) is greater than 50, *STUDENTSDUM*.[1] We focus on the 236 markets where *POP* < 100.[2] Yearly average personal wage income, *PINC* is used to represent the wealth level in a market. Competition is likely to be more intensive the closer markets are to each other, and we therefore include the distance variables *DIST1* and *DIST2*.

The cost structure in the industry will obviously affect the number of firms and capacity in markets. The principal costs are wages for teachers, the cost of office and parking space, cars, traffic insurance, and gasoline. We expect differences in costs of firms within a market to be small (since production functions are essentially the same), but there may be larger differences across markets. There exists no disaggregated data on wages for driving schools, and we therefore use the monthly wage of skilled employees in the trade sector, *WAGE* as proxy. Table 2 shows that the wage spread across counties is small, which is a common feature in Sweden with a generally compressed wage distribution across firms, industries, and regions. The cost of office and parking space is represented by *OFFICEVAL*. Remaining cost variables are excluded in our regressions because regional differences are small, as indicated by the main interest organisation of the firms.[3]

---

[1] We use a dummy variable because in some smaller markets there are no gymnasium, and students go to school in a nearby municipality, where it is likely that they take driving lessons. However, the number of students is highly correlated with *POP* for larger markets.

[2] Markets with *POP*>100 may for at least two reasons be different. First, larger municipals usually bave a university to which people in the age group of 18-24 move from smaller municipals, where they may already have taken their driver's licence. Second, large municipals typically have better public transport, which tend to reduce the demand for driving lessons.

[3] See Asplund and Sandin (1995a) for detailed motivations.

Table 1.–Variables Definitions

Variable	Definition
POP	Total population of age 16-24 in the municipal measured in hundreds. Data from 1993 was collected from the Swedish Association of Local Authorities.
STUDENTS, *STUDENTSDUM*	The average number of students in the gymnasium (high school) for the academic years 92/95 in the municipal, as stated by the National Agency for Education. From this we define the dummy variable to be one if there are 50 students or more, and zero otherwise. Source: Skolverket, 1995, Skolan - Jämförelsetal för skolhuvudmän, Skolverkets rapport nr 73, Solna.
PINC	Yearly average personal wage income in 1993 in the municipality, measured in hundreds of thousands of SEK.[4] Data was collected from Statistics Sweden. Source: Statistiska Centralbyrån, 1995, Inkomst- och skattestatistik 1993, Statistiska meddelanden, Be20 SM 9501, table 14.
DIST1, DIST2	Distance in kilometres to the nearest and the second nearest market where there exists at least on firm. Distances are measured as straight lines between central towns in each municipal.
WAGE	Average monthly wage measured in thousands of SEK in 1990 of skilled employees in the trade sector. County level data were collected from Statistics Sweden. Source: Statistiska Centralbyrån, 1991, Löner och sysselsättning i den privata sektorn 1990, table 9. Later statistics are not available.
OFFICEVAL	Average assessed value of office space per square metre in 1994, measured in thousands of SEK. This includes office space used not only by driving schools. County level data are provided by Statistics Sweden. Source: Statistiska Centralbyrån, 1995, Rikets Fastigheter 1995 (1), Statistiska meddelanden, Bo 38 SM 9501, table 7a.
FIRMS	Number of driving schools in the municipal. All operating driving schools were located from local phone books (yellow pages) and a member list from the main interest organisation, the Swedish Association of Driving Schools where about 95% of the firms are members. With this information we are confident that the data includes all active firms in the markets. See Asplund and Sandin (1995a) for the definition of a firm.
FCARS	Number of cars in the driving school. This information was collected by telephone interviews with each firm. All firms answered this question.
MCARS	Number of cars used by driving schools in the municipal.

The number of firms in a market is denoted *FIRMS*. We have information on the number of cars in each firm, *FCARS*, which gives the number of cars in the market, *MCARS*. The number

[4] In June 1, 1995, one SEK equalled 0.136 US dollars.

of cars is reasonable to use as a measure of capacity, since a car can be used by several teachers working different hours, and because teachers can be hired on a part time basis to cover high demand periods, whereas cars generally can not.[5] Conversely, in periods with low demand, cars are retained, but the hours worked by the teachers are adjusted. If one could assume that capacity utilisation were relatively constant over time and markets, *MCARS* would measure the quantity produced in a market. Even if this can not hold strictly when capacity is lumpy (as are cars), it provides a rough approximation to the produced quantity.

Table 2.– Sample Characteristics

Variable	mean	stand.dev.	min	max	median	no. of observations
POP	28.0	30.9	2.73	177	15.4	249
STUDENTSDUM	0.651	0.478	0	1	1	249
PINC	1.23	0.0836	1.03	1.53	1.22	249
DIST1	27.1	17.2	6.00	119	21.6	249
DIST2	37.6	22.0	9.60	144	32.4	249
WAGE	13.1	0.270	12.6	14.1	13.1	24
OFFICEVAL	3.03	0.592	2.23	5.81	2.87	24
FIRMS	2.16	2.39	0	13	1	249
MCARS	5.24	6.62	0	33	3	249
FCARS	2.44	1.36	1	9	2	535

Table 3 shows the distribution of the number of firms and average per firm capacity across markets. The span of market structures ranges from zero to thirteen firms, of which 207 out of 249 markets contain three firms or less. It reveals a strong positive correlation between the number of firms and market size.

[5] Cars used in driving schools have specific equipment, e.g. brakes that can be controlled by the teacher.

Table 3.–Distribution of Firms, Capacity and Market Size

FIRMS	*Number of Markets*	*Mean POP*	*Mean FCARS*
0	54	9.741	-
1	71	13.44	1.99
2	49	21.56	2.32
3	33	32.90	2.23
4	13	43.35	2.60
5	9	61.58	2.84
6	4	80.12	2.97
7	4	97.84	2.83
8	3	106.8	2.62
9	2	103.8	2.50
10	3	123.4	2.40
11	2	147.7	2.70
12	0	-	-
13	2	161.9	2.38

Table 3 further indicates that per firm capacity increases with the number of firms in the market for the 233 markets with less than seven firms, but for markets with more than seven firm there is a negative relation. The results from the least squares regression reported below confirms this.

$$FCARS = 1.78 + 0.261 FIRMS - 0.0171 FIRMS^2$$
$$(0.169) \; (0.0679) \qquad (0.00517)$$

Adjusted $R^2 = 0.029$ no. of obs. $= 535$

(Standard errors in parenthesis.)

As discussed below, we have no other explanation to why the average firm size is decreasing in markets with more than six firms, than that the estimates are based on few observations. Figure 1 gives a more complete picture on the size distribution of firms across markets with different number of firms. The important fact to note is that there are only a few large firms, indicating the absence of great economies of scale.

3. Econometric models

As noted in the introduction, Bresnahan and Reiss (1991) argue that if the per firm market size increases with the number of firms, it indicates intensified competition. This concave relation between the number of firms and market size implies an increase in average firm size, which is borne out in Table 3 (at least for markets with fewer than seven firms). More interestingly, it implies a convex relation between produced quantity and market size. That is, if a concave relation between the number of firms and markets size corresponds to lower prices in larger markets, then the per capita demand will be increasing in market size.[6] Now, let one car be the capacity unit, and let the entry threshold per capacity unit be the minimum market size necessary to support each car. As an alternative indicator of intensified competition, we then expect these entry thresholds to be decreasing in market size, since each capacity unit will be used to satisfy a smaller share of demand.

Similar methods are employed to estimate firm and capacity entry threshold levels. We begin by setting up an alternative specification of the econometric model in Bresnahan and Reiss (1991),[7] and suggest three other methods to estimate the relation between the number of firms and market size. In our specification it is assumed that each firm in market i, with N firms, has an unobservable profit function

$$\Pi_{i,N} = \frac{S(X_i^S, \beta^S)}{N} V_N(X_i^V, \alpha^V, \beta^V) - F_N(X_i^F, \alpha^F, \beta^F) + \varepsilon_i . \tag{1}$$

$S(.)$ is the size of the market, which is a function of a vector of demographic variables, X^S. $V_N(.)$ is the per capita variable profit, which is a function of market specific demand characteristics and

[6] A concave relation between the number of firms and market size need not necessarily correspond to lower prices, since it may also be associated with higher fixed costs per firm. For an industry such as driving schools, the latter explanation seems less likely, since there is little advertising, and because the supply of workers, cars, and office space tend to be quite elastic.

[7] The alternative specification is applied, since that used by BR has a different interpretation to the one they suggest, as shown in Asplund and Sandin (1995b) and discussed in footnote 8 below.

variable costs, defined by the vector X^V. $F_N(.)$ is the fixed costs function, determined by a vector of market specific fixed cost, X^F. The α's and β's are parameter vectors to be estimated. The error term ε is assumed to be normally distributed as

$$\varepsilon_i \in N(0,\sigma^2), \quad E\varepsilon_i X = 0, \qquad \text{and} \quad E\varepsilon_i\varepsilon_j = 0 \quad \text{if} \qquad i \neq j. \tag{2}$$

In an equilibrium with N identical firms in a market, the net profit of each these firms is non negative, while the profit would be strictly negative for at least one of $N+1$ firms. Given the assumptions about the error term, the probabilities that there are N firms in market i are

$$\begin{aligned}&\text{Prob}(FIRMS_i = N) = \\ &\begin{cases}\text{Prob}(\Pi_{i,1} < 0) = 1 - \Phi(\overline{\Pi}_{i,1}) & \text{if} \quad N = 0 \\ \text{Prob}(\Pi_{i,N} \geq 0 \quad \text{and} \quad \Pi_{i,N+1} < 0) = \Phi(\overline{\Pi}_{i,N}) - \Phi(\overline{\Pi}_{i,N+1}) & \text{if} \quad 0 < N < \hat{N} \\ \text{Prob}(\Pi_{i,\hat{N}} > 0) = \Phi(\overline{\Pi}_{i,\hat{N}}) & \text{if} \quad \hat{N} \leq N,\end{cases}\end{aligned} \tag{3}$$

where Φ(.) is the cumulative normal distribution function, and $\hat{N}$ is defined below. From the assumptions about the error term, the model is an ordered Probit, and can be estimated by maximum likelihood techniques. The per firm entry threshold for this specification (defined by $\overline{\Pi}_N = 0$) is

$$s_N = \frac{\overline{F}_N}{\overline{V}_N} \ . \tag{4}$$

The per firm entry threshold will depend upon values of the market specific variables in V_N and F_N. For comparison, these are evaluated at their sample means. We follow BR closely and assume the following functional forms of S, V_N, and F_N

$$S(X_i^S, \beta^S) = POP_i + \beta_1^S STUDENTSDUM_i \,, \tag{5}$$

$$V_N(X_i^V,\alpha^V,\beta^V)=\alpha_1^V-\sum_{n=2}^{N}\alpha_n^V+\beta_1^V PINC_i+\beta_2^V WAGE_i+ \\ \beta_3^V DIST1_i+\beta_4^V DIST1SQR_i, \tag{6}$$

and

$$F_N(X_i^F,\alpha^F,\beta^F)=\alpha_1^F+\sum_{n=2}^{N}\alpha_n^F+\beta_1^F OFFICEVAL_i. \tag{7}$$

We denote the dependent variable *FIRMS#* and define it as

$$FIRMS\#_i=\begin{cases} FIRMS_i & \text{if} \quad 0\le FIRMS_i<\hat{N} \\ \hat{N} & \text{if} \quad \hat{N}\le FIRMS_i. \end{cases} \tag{8}$$

In our sample we let $\hat{N}=5$, since there are insufficient number of markets with more than five firms in the sample to permit estimation of higher entry thresholds. Hence, all markets with five or more firms are pooled into one category.

The market size is normalised with the population in the age group 16-24 years, *POP*. The coefficient of *STUDENTSDUM* is expected to be positive in that a significant student population will increase market size and thereby gross profits. The coefficient of *PINC* is expected to be positive if it measures the wealth level of the consumers, and if the demand is less elastic in markets where the wealth level is high. *WAGE* is expected to have a negative coefficient, since it reduces the price-cost margin. It is likely that price-cost margins are lower if the distances to the closest markets are short, but the relation is not necessarily linear. Due to multicollinearity problems, we use only the distance to the closest market, *DIST1*, and its square term, *DIST1SQR*. Market specific fixed costs are represented by *OFFICEVAL*, which is expected to have a negative effect on profits, and therefore a positive coefficient. If per firm fixed costs are increasing with the number of competitors, then α_n^F, for $n=2,..,\hat{N}$, are positive. In $V_N(X_i^V,\alpha^V,\beta^V)$, α_n^V for

$n = 2,..,\hat{N}$, are positive if more firms lead to more intensive competition and lower price-cost margins.[8]

The results of the ordered Probit model are compared to a Tobit model and a Poisson model for count data. The Tobit model (see e.g. Amemiya (1985) or Greene (1993)) captures the fact that the dependent variable is non negative. The latent variable $FIRMS^*$ is assumed to be

$$\begin{aligned} FIRMS_i^* = \beta_0 + \beta_1 POP_i + \beta_2 POPSQR_i + \beta_3 STUDENTSDUM_i + \beta_4 PINC_i + \\ \beta_5 OFFICEVAL_i + \beta_6 WAGE_i + \beta_7 DIST1 + \beta_8 DIST1SQR_i + \varepsilon_i. \end{aligned} \tag{9}$$

In the data we observe

$$FIRMS_i = \begin{cases} FIRMS_i^* & \text{if} \quad FIRMS_i^* > 0 \\ 0 & \text{if} \quad FIRMS_i^* \leq 0. \end{cases} \tag{10}$$

The error term in (9) is allowed to be heteroskedastic[9] and normally distributed as $\varepsilon_i \in N(0,\sigma_i^2)$. We found that a good representation of the heteroskedasticity is obtained by setting

$$\sigma_i = \sigma \text{Exp}(\gamma_1 POP_i + \gamma_2 POPSQR_i + \gamma_3 STUDENTSDUM_i). \tag{11}$$

The square of the population variable, *POPSQR*, is included in (9) to capture the possibly non-linear relation between the number of firms and market size.[10] The drawback of the Tobit model is

---

[8] The specification used in BR is

$$\Pi_{i,N} = S(X_i^S, \beta^S) V_N(X_i^V, \alpha^V, \beta^V) - F_N(X_i^F, \alpha^F, \beta^F) + \varepsilon_i$$

The V function is then the *variable per capita profit of each firm*, which obviously will be decreasing in the number of firms, even if price-cost margins are the same for all market structures. It can therefore not be concluded, as in BR, that if the V function is decreasing in the number of firms, it implies lower variable profits *per capita*. However, the two specifications yield very similar estimates of the per firm entry thresholds; see Asplund and Sandin (1995b) for details.

[9] If not accounted for, heteroskedasticity may give inconsistent estimates of the coefficients in the standard Tobit model, although the direction and magnitude of the bias is ambiguous, see Greene (1993 p. 698-700).

[10] We also experimented with other transformations that can give convex or concave functions, but this formulation gave the best fit. However, the other functional forms generally supported a concave relationship between the number of firms and market size, similar to the results shown below.

that it does not use the information that *FIRMS* is a positive, integer variable, i.e. a count.[11] As a second alternative, we therefore apply the most commonly used count data model, the Poisson model. It is reasonable to use this distributional assumption on our sample given the data illustrated in Table 3, where the distribution of firms across markets roughly resembles a Poisson distribution. The probability that there are exactly N firms in market i is

$$\text{Prob}(FIRMS_i = N) = \frac{\text{Exp}(-\lambda_i)\lambda_i^N}{N!} \quad \text{for } N=0,1..,\infty\,, \tag{12}$$

where we specify λ by

$$\begin{aligned}\lambda_i = \text{Exp}(&\beta_0 + \beta_1 POPSQRT_i + \beta_2 POP_i + \beta_3 STUDENTSDUM_i + \beta_4 PINC_i + \\ &\beta_5 WAGE_i + \beta_6 OFFICEVAL_i + \beta_7 DIST1 + \beta_8 DIST1SQR_i).\end{aligned} \tag{13}$$

Note that we include the square root of *POP*, *POPSQRT* to allow for a concave relation between the number of firms and market size.[12] Apart from the specific functional form imposed on the distribution, a well known problem with the Poisson model is the implicit assumption that λ is both the conditional variance and the conditional mean. If this restriction is violated, the coefficients are consistent but their standard errors are inconsistent (see e.g. Cameron and Trivedi (1990)). As a final check of our results, we make a regression model of (13) by substituting *FIRMS* for λ and adding an error term. This model is estimated by non-linear least squares (NLS), ignoring the information that the dependent variable is a positive integer.

---

[11] It has been noted that this may lead to inconsistent estimates of the parameters, but the magnitude of the bias is unknown, see Stapleton and Young (1984). Since our primary concern is not the estimated values of any individual coefficient, but rather the predicted values, we use this model as one among other specifications. Moreover, it is in some instances better to have an inconsistent estimator with small variance, than a consistent estimator with large variance, see e.g. Greene (1993 p. 94-95).

[12] The motivation for not using *POP* together with *POPSQR* is that markets with many firms get very large weights, since λ is defined by $\lambda = \text{Exp}(\beta X)$.

To estimate market capacity we only use specifications (9) and (12), where we substitute *MCARS* for *FIRMS*, and use the same explanatory variables. The ordered Probit model described above is not directly applicable to the problem of estimating the market capacity.

Before turning to the results there are a few general points to note when we compare the models. First, if the specification of the ordered Probit model is correct, it provides valuable information about structural parameters in the profit function. If not, then the parameters of the misspecified model have no such interpretation. The other models do not run this risk, since these do not attempt to estimate a profit function, but merely to describe the relation between the number of firms and the exogenous variables. For example, in the ordered Probit specification it is implicitly assumed that the profit function for all firms *within* a market is the same, but that it differs *between* markets. [13] This is a restrictive assumption, which assumes away the possibility that firms within a market may have different profit functions. However, for an industry such as driving schools it may not be unreasonable, since firms are of roughly the same size, use the same technology, and because the product is relatively homogenous.[14] Second, the ordered Probit model is associated with a considerable information loss, since estimation required pooling of markets with five or more firms into one category. Finally, the number of firms in different markets is not only a ranking order, but also a count, that is, two firms are not only more than one, but it is exactly one more. The other models use this information and do not pool the markets into different groups, but the Tobit model may give inconsistent parameter estimates.[15]

[13] In Bresnahan and Reiss (1990), more elaborate specifications were applied to markets with one or two firms, where profit functions could vary across firms in a market. However, these become impossible to estimate when there are more than four firms in the market, since they require integration of multidimensional normal distributions.

[14] However, to use this as a general assumption seems too strong, since the size distribution of firms tend to be skewed (see e.g. Schmalensee (1989 p. 994)), and thus firms may have different profit functions even if they operate within the same market.

[15] See Cameron and Trivedi (1986) for more comparisons between count data models, least squares estimators, and ordered Probit models.

4. Results

The results reported in Table 4 are for the subset of 236 markets with *POP*<100, (the results for the full sample (249 markets) are shown in the Appendix). In the ordered Probit model, most of the coefficients have the expected signs. Higher wage levels and office values decrease profits, while higher income levels, and the distance to the closest markets increase profits. However, only a few of them are individually significant, which is similar to the findings in BR. Likelihood ratio tests of joint restrictions on the cost and distance variables are shown in Table 5. It is not possible to reject (the critical value at the 5% level is $\chi^2 = 5.99$) the hypothesis that costs differences have no effect on profits. Likewise, we can not reject the hypothesis that the profit is independent of the distance to the closest market. The α^F's are positive as expected, which suggests that fixed cost increase with the number of competitors. However, α_2^V is negative and significant, which suggests that duopoly firms have higher variable per capita profits than monopolies. Moreover, α_3^V and α_5^V, are negative (although insignificant). All these negative coefficients seem unreasonable, in particular when set in relation to the results in Asplund and Sandin (1995a), where estimated prices in duopoly markets were lower than in monopoly markets. The most obvious explanation is that the profit function (1) is too restrictive, such that there is little meaning in interpreting the coefficients as parameters in a profit function. Despite this question of interpretation, it is interesting to study the estimated per firm entry threshold levels. Before that we look at the alternative estimators.

Results reported for the Tobit model in Table 4 are robust to heteroskedasticity, which was present in a standard Tobit model. The most important finding is that the coefficient of *POP* is positive whereas *POPSQR* is negative, which implies a concave relation between the number of firms and market size. Both measures of costs, *WAGE* and *OFFICEVAL*, are negative as expected, but they are not jointly significant at the 5% level, as seen in Table 5. This is also true for the distance measures. As in the ordered Probit, markets with high average personal income tend to have a larger number of firms, but the effect is not significant.

Finally, the signs of the comparable coefficients in the Poisson and NLS models are similar to the ordered Probit and the Tobit model. As noted in Section 3, a well known problem with the Poisson model is the implicit assumption that the conditional mean and variance are equal, something that is often inconsistent with count data which usually display overdispersion. We tested for mean-variance equality using methods of Cameron and Trivedi (1990) and found *under*dispersion. Therefore our parameter estimates may be consistent, but their tests statistics are not, and we therefore refrain from discussing their significance levels.[16]

A point worth commenting on are the low significance of the cost variables, which potentially could be explained by the fact that these are measured at county level. The bias this may cause is probably of minor importance for the wage variable, since wage differences generally are small across regions in Sweden, as indicated in Table 2. However, we had expected to find *OFFICEVAL* more significant, since it has greater variation across regions.

Table 6 shows the estimated per firm entry thresholds. The market size necessary to support N firms in the Tobit model is defined by the value of *POP* that satisfies $FIRMS^{*}=N$, and for the Poisson and NLS by $\lambda=N$ and $FIRMS=N$ respectively, for $N=1,..,5$. The per firm entry threshold is obtained by dividing this value of *POP* with N. For all thresholds, the other variables are evaluated at their sample means, and for $STUDENTSDUM=1$. By inspection, the general conclusion must be that the different econometric techniques give similar estimates of the threshold values The differences between the models are small, e.g. for monopolies (duopolies) the implied per firm entry threshold varies between 895 and 1150 (1120 and 1200) people in the age group of 16 to 24 years old. It is interesting to note that all specifications (with the exception of s_5 for the ordered Probit and s_1 in the Tobit) support the hypothesis that the per firm minimum market size is increasing with the number of firms.

[16] We have also estimated an extended Poisson model (WZ-Poisson) of Mullahy (1986), which allows for a specific form of underdispersion. However, the specified form of underdispersion did not fit our data well, and the model was highly sensitive to specification, why we report the original Poisson model. Results from the WZ-Poisson model are available upon request.

The results for the regressions with *MCARS* as the dependent variable are shown in Table 7. As in Table 4, the most significant variables are those measuring market size, while the remaining variables generally are insignificant, and some even have a perverse sign. Table 8 shows the estimated per capacity unit entry thresholds. As expected, the market size of each capacity unit is decreasing in market size, at least for eight or fewer capacity units. For more capacity units the market size increases in the Poisson model, which is an effect of the non-monotonic relation found in Table 3. Up to this level, each capacity unit satisfies a smaller share of the consumer demand. This finding provides complementary evidence for that competition is increasing in market size. It is interesting to note that the capacity unit entry threshold for the first car in the Poisson (Tobit) model is slightly (much) lower than the monopoly entry threshold. One interpretation is that per firm entry thresholds for these models are for the average firm, which may not be the smallest possible firm. For example, the average monopoly has 1.99 cars, but there are several monopolies with only one car, and these may have an opportunity to enter at markets sizes smaller than those estimated.

The different econometric specifications gave broadly similar predicted values of entry thresholds for firms and capacity. As noted in section 3 all specifications have drawbacks, and the questions about model selection naturally arise. We weakly prefer the Poisson model to the alternatives. This is motivated by the count data nature of the dependent variables, which also had distributions that resembled the Poisson, and by the computational simplicity of the model. The ordered Probit model seemed to impose too much structure on our data, which gave unreasonable results. However, we still believe that it can work well on another sample.

5. Concluding remarks

For industries with homogenous goods and absence of large barriers to entry, it is not unreasonable to assume that the relation between market size and the number of firms provides information about how fast competition is increasing with the number of firms. When entry is relatively easy, new firms will establish as the market expands, and if this leads to intensified competition with lower profits, the per firm market size needs to be larger in order for firms to cover fixed costs. This study showed that per firm entry threshold is increasing in the number of firms, which is similar to the results in Bresnahan and Reiss (1991). This finding was remarkably robust to all our econometric specifications. In addition we showed that capacity unit entry thresholds are decreasing in market size. Together these results indicate that competition is more intense in larger markets. This conclusion is broadly consistent with the findings in Asplund and Sandin (1995a), where we used price information from firms in this industry, and showed that prices were significantly decreasing in the number of firms in the market.

Although the method of estimating the relation between market size and the number of firms, or capacity, is useful for homogeneous goods industries, it is less clear that it can be used for industries with differentiated products. For example, Sutton (1991) shows that the relation between market size and the number of firms (or more correctly - concentration) could be non-monotonic due to endogenous investments in R&D or advertising. Extending the analysis to deal with differentiated products, and to account for the fact that firms within a market can have significantly varying profit functions, seem to be important questions to address in future research.

6. References

Asplund, M., and R. Sandin, 1995a, Competition in interrelated markets: An empirical study, Working Paper Series in Economics and Finance, no 84, Stockholm School of Economics.

Asplund, M., and R. Sandin, 1995b, Entry and Competition in Concentrated Markets: Comment, in manuscript, Stockholm School of Economics.

Amemiya, T., 1985, Advanced Econometrics, Cambridge Mass.: Harvard University Press.

Bresnahan, T.F., and P.C. Reiss, 1988, Do entry conditions vary across markets?, *Brookings Papers on Economic Activity*, 3, 833-82.

Bresnahan, T.F., and P.C. Reiss, 1990, Entry in Monopoly Markets, *Review of Economic Studies*, 157, 531-53.

Bresnahan, T.F., and P.C. Reiss, 1991, Entry and Competition in Concentrated Markets, *Journal of Political Economy*, 99, 977-1009.

Cameron, A.C., and P Trivedi, 1986, Econometric models based on count data: Comparisons and applications of some estimators and tests, *Journal of Applied Econometrics*, 1, 29-53.

Cameron, A.C., and P. Trivedi, 1990, Regression based tests for overdispersion in the Poisson model, *Journal of Econometrics*, 46, 347-364.

Caves, R.E., 1989, International differences in industrial organization, in R. Schmalensee and R. Willig, Handbook of Industrial Organization, Vol. 2 (Elsevier Science Club), 1225-1250.

Curry, B., and K.D. George, Industrial concentration: A survey, *Journal of Industrial Economics*, 31, 203-255.

Greene, W.H., 1993, Econometric Analysis, New York; Macmillan.

Mullahy, J., 1986, Specification and testing of some modified count data models, *Journal of Econometrics*, 33, 341-365.

Schmalensee, R., 1989, Inter-industry studies of structure and performance, in R. Schmalensee and R. Willig, eds., Handbook of Industrial Organization, Vol. 2 (Elsevier Science Club), 951-1009.

Stapleton, D.C., and D.J. Young, 1984, Censored normal regression with measurement error on the dependent variable, *Econometrica*, 52, 737-760.

Sutton, J., 1991, Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration, The MIT Press, Cambridge Mass. and London.

Appendix

Table A1.–Regression Results with *FIRMS* as the Dependent Variable (249 markets)

	Ordered Probit	Tobit[a]	Poisson	NLS[b]
Variable	*FIRMS#*	*FIRMS*	*FIRMS*	*FIRMS*
CONSTANT		1.64	-0.678	-0.576
		(4.52)	(4.54)	(1.59)
POPSQRT			0.498***	0.364***
			(0.160)	(0.0731)
POP	≡1.00	0.0745***	-0.0164*	-0.828 E-02**
		(0.0113)	(0.995 E-02)	(0.410 E-02)
POPSQR		-0.461 E-04		
		(0.777 E-04)		
STUDENTSDUM	4.28*	0.571***	0.397*	0.511***
	(2.46)	(0.182)	(0.218)	(0.135)
PINC	0.0964	2.46**	0.571	0.485
	(0.0947)	(1.18)	(1.18)	(0.444)
OFFICEVAL	0.168	-0.133	-0.0758	-0.112**
	(0.160)	(0.163)	(0.158)	(0.0556)
WAGE	-0.0265	-0.365	-0.120	-0.0778
	(0.0303)	(0.381)	(0.383)	(0.113)
DISTI	0.191E-02	0.814 E-02	0.228 E-02	-0.857 E-03
	(0.133 E-02)	(0.0133)	(0.0120)	(0.426 E-02)
DISTISQR	-0.213 E-04	-0.864 E-04	-0.536 E-04	-0.432 E-04
	(0.139 E-05)	(0.131 E-03)	(0.125 E-03)	(0.430 E-04)
α_1^V	0.303			
	(0.393)			
α_2^V	-0.0938***			
	(0.0295)			
α_3^V	-0.0333			
	(0.0326)			
α_4^V	0.0133			
	(0.0420)			
α_5^V	-0.0511			
	(0.0445)			
α_1^F	0.432			
	(0.615)			
α_2^F	1.22***			
	(0.358)			
α_3^F	0.590*			
	(0.331)			
α_4^F	0.325			
	(0.434)			
α_5^F	0.704			
	(0.516)			
σ		0.982***		
		(0.160)		
Number of obs.	249	249	249	249
Log *L*	-271.2	-344.8	340.8	
adjR^2				0.84

Standard errors in parenthesis.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

[a] In (11) $\gamma_1 = 0.0152$ (0.639 E-02), $\gamma_2 = -0.706$ E-04 (0.451 E-04), and $\gamma_3 = -0.284$ (0.173).

[b] Heteroskedastic consistent standard errors.

Table 4.-Regression Results with *FIRMS* as the Dependent Variable (236 markets)

Variable	Ordered Probit *FIRMS#*	Tobit[a] *FIRMS*	Poisson *FIRMS*	NLS[b] *FIRMS*
CONSTANT		1.50	-1.73	-1.93
		(4.57)	(5.06)	(1.77)
POPSQRT			0.770***	0.630***
			(0.275)	(0.104)
POP	≡1.00	0.106***	-0.0394*	-0.0289***
		(0.0176)	(0.0209)	(0.745 E-02)
POPSQR		-0.466 E-03**		
		(0.189 E-03)		
STUDENTSDUM	4.99*	0.449**	0.297	0.366***
	(2.72)	(0.190)	(0.235)	(0.139)
PINC	0.103	2.47**	0.807	0.766
	(0.0929)	(1.19)	(1.22)	(0.490)
OFFICEVAL	0.135	-0.122	-0.0834	-0.111*
	(0.163)	(0.162)	(0.174)	(0.0601)
WAGE	-0.0141	-0.387	-0.115	-0.0608
	(0.0313)	(0.380)	(0.409)	(0.129)
DIST1	0.132 E-02	0.0133	0.614 E-02	0.641 E-02
	(0.131 E-02)	(0.0131)	(0.0127)	(0.513 E-02)
DIST1SQR	-0.162 E-04	-0.115 E-03	-0.751 E-04	-0.917 E-04*
	(0.136 E-04)	(0.128 E-03)	(0.127 E-03)	(0.525 E-04)
α_1^V	0.136			
	(0.401)			
α_2^V	-0.0928***			
	(0.0285)			
α_3^V	-0.0344			
	(0.0321)			
α_4^V	0.0128			
	(0.0428)			
α_5^V	-0.183*			
	(0.0945)			
α_1^F	0.469			
	(0.614)			
α_2^F	1.25***			
	(0.353)			
α_3^F	0.610*			
	(0.336)			
α_4^F	0.333			
	(0.448)			
α_5^F	1.98**			
	(0.956)			
σ		0.964***		
		(0.187)		
Number of obs.	236	236	236	236
Log *L*	-266.8	-316.4	-310.5	
adjR^2				0.69

Standard errors in parenthesis.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

[a] In (11) $\gamma_1 = 0.0169$ (0.0122), $\gamma_2 = -0.102$ E-03 (0.130 E-03), and $\gamma_3 = -0.298$ E-03 (0.183).

[b] Heteroskedastic consistent standard errors.

Table 5.–Likelihood Ratio Tests

Restriction	Ordered Probit	Tobit	Poisson
DIST1=DIST1SQR=0	5.63	0.58	0.92
WAGE=OFFICEVAL=0	2.29	4.01	2.00

Table 6.–Estimates of Per Firm Entry Thresholds

Per firm entry threshold	Ordered Probit	Tobit	Poisson	NLS
s_1	8.95	11.5	11.2	9.18
s_2	11.2	11.3	12.0	11.2
s_3	12.2	11.8	12.4	12.0
s_4	14.5	12.6	13.3	12.8
s_5	12.8	14.1	16.4	14.0

Table 7.–Regression Results with *MCARS* as the Dependent Variable

Variable	Tobit[a] *MCARS*	Poisson *MCARS*
CONSTANT	10.3	-0.338
	(10.1)	(1.59)
POPSQRT		1.08***
		(0.125)
POP	0.330***	-0.0564***
	(0.0477)	(0.987 E-02)
POPSQR	-0.150 E-02***	
	(0.468 E-03)	
STUDENTSDUM	0.476	0.345***
	(0.424)	(0.110)
PINC	3.32	-0.195
	(3.00)	(0.441)
OFFICEVAL	0.0495	0.840 E-02
	(0.363)	(0.0556)
WAGE	-1.34*	-0.165
	(0.789)	(0.120)
DIST1	0.0103	-0.101E-02
	(0.0342)	(0.547 E-02)
DIST1SQR	0.118 E-03	0.485 E-05
	(0.351 E-03)	(0.600 E-04)
σ	1.39***	
	(0.242)	
Number of obs.	236	236
Log *L*	-477.1	-474.7

Standard errors in parenthesis.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.
a $\gamma_1 = 0.0477$ (0.0136), $\gamma_2 = -0.407$ E-03 (0.156 E-03), and $\gamma_3 = -0.211$ (0.148).

Table 8.–Estimates of Per Capacity Unit Entry Threshold Levels

Per capacity unit entry threshold	Tobit	Poisson
s_1	6.41	9.30
s_2	6.00	6.65
s_3	5.55	5.60
s_4	5.20	5.12
s_5	4.94	4.86
s_6	4.75	4.71
s_7	4.58	4.62
s_8	4.46	4.61
s_9	4.37	4.62
s_{10}	4.31	4.66

igure 1. - Size Distribution of Firms