A Selection Model of R&D Intensity and Market Structure of Spanish Firms

Joaquín Artés[1]

Department of Economics

University of California, Irvine

This version: October 18th 2006

Abstract

In this paper I provide econometric evidence using data of Spanish firms showing that monopoly power has a positive effect on the probability of a firm carrying R&D. In addition I show that conditioned on the firm undertaking positive R&D, market structure does not affect R&D intensity. The results are robust to several specifications and measures of monopoly power.

[1] I thank Professor Linda Cohen for encouragement and helpful comments. I also thank Professor David Brownstone for his help regarding many aspects of the paper. I gratefully acknowledge financial help from The Rafael del Pino Foundation and The SEPI Foundation for providing the data. I take responsibility for all remaining errors. E-mail address: jartes@uci.edu.

1. INTRODUCTION

The relation between R&D and market concentration has been widely studied in the Industrial Organization literature, perhaps because of the contradictory evidence found when using different data, variables, and econometric models. For this reason, a strong case should be made before embarking in a new study of R&D and market structure.

In this paper I present new evidence on such relation using a panel of Spanish manufacturing companies. Several facts make the effort worthwhile. I adopt a different approach from previous studies by distinguishing between long-run and short-run decisions of firms regarding R&D. I assume that the long-run or strategic decision is whether to conduct R&D or not and the short-run choice is how much to invest once the firm decides to be innovative. I argue that market structure affects long-run R&D decisions but does not affect short run ones. A Heckman-type selection model is used to test such a relation. The advantage of this model compared to the usual OLS or Tobit models is that it allows us to gain information from the fact that the decision of whether to innovate or not and how much to expend on R&D once the firms decides to innovate are different but, depend, to certain extent in related market and firm characteristics. Most of the previous literature in which a significant relation is found to exist between market power and innovation activities focus on a particular measure of market power. I include up to five different measures and show that including just one does not fully capture market power and could be misleading. Finally, our rich database makes it possible to control at the same time for many of the characteristics of companies and markets that previous studies have found to be important, including technological opportunities, appropriability and demand conditions, and industry characteristics.

In a paper that uses several measures of market power Geroski (1990) finds no evidence of the effect of these variables on innovation. My results show that concentration and other measures of monopoly power have a significant effect on the decision to innovate but not on the intensity of R&D. This result is robust to several specifications and measures of monopoly power.

In addition, my results provide insight in several policy issues: more monopoly power is associated with more firms choosing to innovate, but on the other hand, innovative firms could as a result be innovating less. If the decision to innovate depends in part on characteristics of the firm that do not affect R&D intensity, as is the case in our study, then, policy measures affecting those parameters would lead to more firms carrying out R&D without affecting R&D intensity of already innovative firms.

The remaining sections of the paper are organized as follows. In section 2, I review the relevant literature. In section 3, I present the econometric methods. Section 4 discusses data and measurement issues. In section 5, the main results are discussed. Section 6 is devoted to extensions and robustness checks. Section 7 concludes.

2. LITERATURE REVIEW

Modern research on the determinants of innovations by firms starts with the works of Schumpeter (1942). He argued that firms with greater monopoly power have a greater incentive to innovate because they can better appropriate the returns of their R&D investment. That is, competition reduces the expected returns to R&D spending. This simple argument has given rise to a vast literature both theoretical and empirical trying to assess the accuracy of Schumpeter's predictions.

The theoretical literature provides arguments to support both a negative effect of competition on R&D and a positive effect. The different conclusions come from different model assumptions regarding property rights (exclusive or non-exclusive), the type of innovation (cost reducing or demand enhancing), or the market structure in which firms operate before and after the innovation (monopoly, duopoly, perfect competition).

In an early paper, assuming that innovators enjoy perfect protection (monopoly power) of their property rights over the invention, Arrow (1962) argued that firms in competitive markets benefit more from process innovations than firms in monopolistic markets. Scherer (1980) and Porter (1990) from different perspectives, find that monopoly discourages innovation because firms don't need to innovate to stay in business.

In a model of patent races, Gilbert and Newbery (1982) maintain the hypotheses of exclusive property rights and cost-reducing innovations but consider a different market structure setting in which old and new firms race to obtain patents over new technologies. They find that monopolistic firms have greater incentives to invest on R&D than competitive ones because they can still enjoy duopolistic profits if they lose the race. A key assumption of the Gilbert and Newery (1982) model is that the winner of the race is the firm that invests the most on R&D. If uncertainty over invention is added to the duopoly model as in Reinganum (1983) firms in competitive markets have more incentives to innovate. Her result depends on the modeling of the discovery process but is robust to extending the model to a more complex market structure with more than two companies.

Dasgupta and Stiglitz (1980) consider the case of non-exclusive property rights but maintain the assumption that innovations are cost reducing. They find that competition is harmful to R&D because it results in redundant R&D expenditures. In particular they find that firms' incentives to invest on R&D depend on how competitive markets are before and after the invention.

In a model that integrates both positive and negative effects, Lowry (1979), building on Kamien and Schwartz (1972, 1976), assumes technological and market structure uncertainty, and focuses on the strategic interdependence of duopolistic firms regarding the optimal timing decision on when to invest. He finds an inverted U-shape relation between competition and R&D. In a recent model of growth, Aghion et al (2004) also show that more competition will initially cause greater R&D effort because of the higher incremental profits of innovation in competitive markets but innovative intensity will decrease above a certain level of competition due to the smaller incentive to innovate faced by laggards.

Given that the conclusions of theoretical papers differ and are strongly dependant on model assumptions, a vast array of empirical studies has tried to provide evidence on which of the two different competing effects (positive or negative) of market structure on innovation is predominant in real data. The evidence is somewhat disappointing because of its inconclusiveness[2]. Some papers find a positive relation between monopoly power and R&D -e.g. Kraft (1989), Crèpon et al. (1998), Blundell et al (1999))-, others find a negative relation -Geroski (1990), Harris, Rogers, Siouclis (2001), Okada (2005) or Nickell (1996)-, an even others, the inverted U-shape predicted in Aghion et al. (2004) as well as in some of the early studies -e.g. Aghion et al. (2004), Tingvall and Poldahl (2006). Most of them find that after controlling for demand, appropriability conditions and technological opportunities, the effect of market structure on innovation is very small or non-existent (e.g. Cohen, Levin and Mowery (1985)).

This inconclusiveness has been attributed to the different measures of innovation and market power and to problems with the estimation techniques and databases. To name only a few, innovation has been measured by R&D over sales (Cohen, Levin, Mowery (1985)) R&D employment (Scherer (1967)), R&D capital per worker (Crèpon et al (1998)), patent counts (Aghion (2004), Geroski (1990)) sales from innovative products (Mairesse and Mohenen (2001)), or innovation dummies (Polhmeier (1992) or Harris, Rogers and Siouclis (2001)). Market structure has been measured by price-cost margins (Okada (2005), Nickell (1996), Aghion (2004)), Herfindahl index (Tingval and Poldahl (2006)), concentration ratio (Scherer (1967), Culberson and Muller (1985), Cohen, Levin, Mowery (1995)), number of competitors (e.g. Kraft (1989)), number of new companies, market shares or market growth (Geroski (1990) includes all this three). Even papers that use the same measures of R&D and competition reach different conclusions.

There are very few papers however that use more than one variable to capture market structure. Data on market structure show that the

---

[2] See Cohen and Levin (1989) and Gilbert (2006) for a review of this literature.

correlation among the different possible measures of monopoly power (i.e. concentration ratio or profit margins) is low. Therefore the debate could benefit from a study that uses many of the possible variables. Geroski (1990) includes up to five different measures in the same paper and finds that monopoly power does not have an effect on the number of innovations.

In this paper I reach a different conclusion showing evidence that provides support to Schumpeter's arguments. I also include many different measures of monopoly power but adopt a different approach by distinguishing between long-run and short-run decisions of firms regarding R&D. I assume that the long-run or strategic decision is whether to conduct R&D or not and the short-run choice is how much to invest once the firm decides to be innovative. I argue that market structure affects long-run R&D decisions but does not affect short run ones. I adopt this focus based on the fact that firms' R&D policy is usually a long-run one due to the length and uncertainty of the innovative process; conducting R&D activities requires of high fixed costs that cannot be changed on the short run. Therefore, when a company decides to invest in R&D, it takes into account expectations over market conditions over a long period of time and only considers short-run fluctuations if they are sizable and likely to be permanent. Moreover, at the same time the decision of whether to conduct R&D is made, the "level" or average amount of innovative effort has also to be determined in order to compute if expected profits derived from this average level, given expected market structure, worthwhile the effort. As the usual data on R&D are observed over a short period of time, standard linear regression or time series analysis is not useful to measure the effect of stable or relatively permanent market conditions on firms R&D decisions.

I use a two-stage selection model. In the first-stage I use a panel probit analysis of the decision of companies of whether or not to undertake R&D. I combine it with a choice of covariates that captures market structure conditions that are relatively stable over time to measure the effect of "stable" market structure conditions on "stable" R&D decisions. The second stage captures the effect of market structure on short-run or non-permanent R&D decisions (R&D intensity). This approach is possible now due to the increasing availability of valuable innovation and business surveys with detailed information on both companies that perform R&D and those that do not.

Several papers before have used probit and selection models to study the determinants of innovation, and although their emphasis is not on the study of market structure, they all offer preliminary support to the hypothesis of a positive effect of market structure on the innovations dummies. In a related study Sanyal and Cohen (2001) use a similar selection model to study the impact of deregulation on US electricity market. Harris, Rogers and Siouclis (2001) estimate a panel probit model of the determinants of firm innovations in which the dependent variable is a dummy of innovations but the only regressor able to capture market power is price margin. Pohlmeier (1992) uses a probit model to study if concentration ratio and the probability of firm to report an innovation are simoultaneously determined. Martinez-Ros (2000) studies the difference between the determinants of product and process innovations and finds support for the U-shape influence of concentration on the probability of a company to report an innovation. In a more related paper, Crépon, Duguet and Mairesse (1998) use a selection model to estimate the relation between innovation and productivity and find in the first step a significant positive effect of monopoly power on innovation. Bound et al (1982) study the effect of total sales on R&D including a selection equation in the first step.

Similarly, Cohen, Levin and Mowery (1987) find that while the effect of size on R&D intensity vanishes after controlling for industry specific characteristics, the effect still holds if the probability of innovation is used as the dependent variable in the first stage of a selection two step model.

Given this preliminary evidence and based on the distinction between stable and variable R&D decisions a detailed study of the effect of market structure on innovative activities in this context seems useful. My first step probit regressions support Schumpeter's hypotheses even when controlling for industry characteristics, demand conditions, technological opportunities, appropriability conditions and some other controls that previous literature have found to be important. My second step regressions in which I estimate the decision of how much to innovate are in line with previous studies (i.e. Geroski (1990)) and thus do not find a strong significant effect of market structure on R&D intensity.

3. EMPIRICAL MODEL

I model firms' R&D decisions as a two-stage process. In the first stage firms decide either to be innovative or not. I assume that firms become R&D doers if the expected net profits from conducting R&D activities are positive. Let y_{it}* be expected net profits of firm *i* from conducting R&D on period *t*. Expected profits are unobservable and depend on a vector *X* of covariates representing present and future market conditions and firm characteristics. Thus we only observe positive R&D for companies with positive expected profits. Furthermore, let $RD_{it}=1$ denote firms that report positive R&D, and Pr(.) a density function that assigns a given probability to the argument inside the brackets. Then, F(.) is the corresponding distribution function, *B* parameters to be estimated and u_{it} the error term. Assuming $y_{it}^{*}= X'B+u_{it}$, then the model can be summarized as:

$$RD_{it} = 1 \text{ if } y_{it}^{*} = X'B + u_{it} > 0 \qquad [1]$$
$$= 0 \text{ if } y_{it}^{*} = X'B + u_{it} \leq 0$$

And, $Pr(RD_{it} = 1) = Pr(y_{it}^{*} > 0) = Pr(u_{it} > -X_{it}'B) = F(X'B)$ holds as long as F is symmetric. I assume that F(.) is a normal distribution function and thus estimate the vector of coefficients B using a probit model in a panel data context. In my preferred specification I assume a random effects structure in the error term so that $u_{it} = u_i + v_{it}$ where u_i and v_{it} are independently and identically distributed with zero means and σ_u^2 and σ_v^2 variances. The random effects specification requires us to assume that the error is uncorrelated with the covariates. Estimates would be inconsistent in case the restriction was not supported by the data. To solve this I follow Zabel (1992) approach and estimate the model including a vector of the group means of the time-variant regressors among the dependent variables[3]. This approach is also consistent with the hypothesis that the decision to carry R&D is a long run decision that does not depend on short run fluctuations if the fluctuations don't change the expectations of the firm about the future. In addition it allows for a long run interpretation of the means as the specific technological or managerial advantage of each firm.

In the second stage I estimate a random effects linear regression in which the amount of R&D (measured by R&D intensity), conditional on the company being an R&D doer, is the dependent variable. I include as a covariate a polynomial of the inverse Mills ratio computed after the first stage to account for the selection bias. The bias exists because we only observe R&D intensity of firms that do R&D but when they make the

[3] I specify the time invariant group-specific component of the error term as $u_i = \overline{X}'\delta + \varepsilon_i$ where δ is a vector of parameters and ε is distributed N~(0, σ_ε^2).

decision in the first stage they also take into account how much R&D spending they would make in case they performed technological activities[4].

I assume this second step to be a short run decision and thus dependant on actual market conditions. The distinction between short and long run decisions in this context is important for both theoretical and practical reasons. From a theoretical point of view it means that we assume that short run results do not affect the decision of whether to do R&D or not if they are not likely to be permanent. This is in accordance with the data as long as half of the firms in our sample do not change their R&D status over the years. The amount of R&D shows, on the other hand, more dependence on business cycle fluctuations. From a practical point of view this implies that some covariates should be different in the two stages, which is required for the estimation of the Heckman-type selection model to be stable[5].

The extension of the Heckman's (1978) two stage selection procedure to a panel data context requires stronger assumptions than in the cross-section case for the estimates of the parameters to be consistent. Vella and Verbeck (1999), Nijman and Verbeek (1992) or Ridder (1990) discuss these assumptions. The more problematic of them is that the error terms measuring the individuals specific random effects in both equations should not be correlated. This is not likely to be the case as both errors reflect individual non-observable heterogeneity. To test the robustness of the model to the violation of this assumption, I also estimate the model

[4] In the models of columns 1 to 4 of table 3 I include just a linear specification of the control term (inverse Mills ratio). Several other specifications including up to a four-order polynomial of the control term were ran leading to similar results. Higher order polynomials allow to account for non-linearities and to reduce the risk of collinearity without strict exclusion restrictions. Given that they do not change the results we report only the simplest specification.

[5] The inestability comes from the multicollinearity introduced by adding the inverse Mills ratio to the second stage. See Woolridge (2002:565).

using Woolridge's (1995) consistent estimator for panel data with sample selection. This method differs from the previous one in that a standard probit model, instead of a random effects, is estimated in the first step, and a pooled OLS in the second.

The variances of the second step of the two procedures need to be adjusted to account for the first step. Otherwise, the unadjusted covariance matrix would be biased downwards. While this is usually a problem because it leads to overestimate the statistical significance of the parameters of interest, in our case, given that our finding is that the parameters of interest are non-significant in the second stage, the uncorrected variance could be seen as the most conservative approach. However, given that both, corrected and uncorrected standard errors lead to the same conclusion regarding the market structure variables, I prefer to report the econometrically more accurate corrected standard errors. A feasible procedure to adjust the variances of the second step is to use minimum distance estimation on the second step after the unrestricted coefficients are obtained[6]. Here due to the problem of missing data in the database that is described in detail below, I follow a different procedure that is rooted in the multiple imputation literature (Brownstone (1991)) and that leads also to consistent estimates of the standard errors. The procedure consists in treating the inverse Mills ratio as a variable with missing values and can be summarized as follows:

1. Estimate the vector of parameters, B, of the probit model.
2. Draw m different samples of coefficients, B_i from a MVN~(B, V_p), where V_p is the variance covariance matrix of the probit estimated coefficients.

[6] See for example Dustman and Rochina-Barrachina (2000) for an application.

3. Compute the inverse Mills ratio *m* times (IM_i), using each time each of the different sampled coefficient vectors, B_i.
4. Estimate the second step random effects regressions *m* times including a different IM_i vector each time. Save the coefficients, b_i, and the covariance matrix, Ω_i, after each estimation.
5. The final estimates are obtained using Rubin's (1987) formulas to obtain consistent results. The formulas are the following:

$$\hat{b} = \frac{1}{m}\sum_{i=1}^{m} b_i \qquad [2]$$

$$\hat{\Omega} = U + (1 + \frac{1}{m})W \qquad [3]$$

Where:

$$U = \frac{1}{m}\Omega_i \quad \text{; and} \qquad [4]$$

$$W = \frac{1}{m-1}\cdot(b_i - \hat{b})'(b_i - \hat{b}) \qquad [5]$$

The method provides consistent estimates of the standard errors provided that we use a robust covariance matrix in the estimations of the second step as we did. The method is computationally intense but in our case, given that we already had to perform a multiple imputation procedure to solve the missing data problems of our database, computing the corrected standard errors this way actually saved time compared to alternative methods.

An alternative to these methods of estimating panel data sample selection models would have been to use conditional maximum likelihood estimation (as in Vella and Verbeeck (1999:246), integrating out the individual-specific error terms of the resulting likelihood and then maximizing it using maximum simulated likelihood. This approach is certainly to be preferred because it provides estimates that are both consistent and efficient. In our case, convergence problems of some the maximization algorithms left us with the easier to implement two-step procedures.

4. DATA, VARIABLES AND MEASUREMENT ISSUES

4.1 Data

I use the Spanish Survey of Entrepreneurs' Strategies (ESEE) conducted by the SEPI Foundation and the Spanish Ministry of Industry since 1991. The survey includes information on around 2000 companies every year. The survey is exhaustive for companies with more than 200 workers. For companies between 10 and 200 workers a random sample by industry and size interval (four intervals) is carried out, weighting companies by size intervals and industries according to the total population of Spanish manufacturing firms. A balanced panel of around 700 firms is included in the survey while others drop and enter the survey every year. In order to preserve the inference properties of the sample newly created companies are also incorporated every year. The final composition of the sample is supposed to be representative of Spanish industry, and, in fact, several studies have confirmed this fact (e.g. Fariñas and Jaumandreu (1999)).

4.2 Variables.

For both steps we need a measure of innovation on the left hand side, and several measures of market characteristics, controls for appropriability, and other market conditions on the right hand side. While some of the covariates are the same in both stages, others differ to reflect the different interpretations of the two steps. I now provide details on the choices made. I include a detailed definition of each variable in the appendix.

Innovation measures

In the first step I use two different measures of innovation. One of them captures innovative output and the other innovative input. In the main specification I include a dummy that takes value 1 if R&D expenditures over sales are greater than 0. However, not all the companies that report positive R&D, report product or process innovations, nor do all the companies that report product or process innovations report positive R&D. To account for this I reran the model using a dummy that captures innovative output and takes value 1 if the firm reports any product or process innovations and 0 otherwise. The conclusions derived from both models are similar and therefore I only report the first one, making the results of the second available under request.

In the second step I follow Crèpon et al (1998) and use R&D per worker (in real terms). An alternative is to use R&D to sales instead but this measure shows little time variance in our sample probably due to the fact that both R&D spending and sales present a similar pattern over time. For this reason the changes in both the numerator and denominator cancel each other and hide short run changes on innovative effort by firms[7]. The number of workers is more stable over time and therefore using R&D over workers facilitates identification of short run changes in R&D[8].

Market structure variables

7 Sales however are also included as a regressor in the second step.

8 Regressions ran using either of the two measures lead to the conclusion that market power does not have a significant effect on R&D intensity. Unreported results are available from author under request.

I use several indicators of product market competition: concentration ratio, price-cost margins, mean price-cost margin of the industrial sector, market share, a dummy that takes the value 1 if the firms reports that there are less than ten companies with significant market share in the main product market and a dummy that takes the value 1 if the firm reports to have a significant market share and 0 otherwise. None of the measures on its own fully captures the market power of the company. I follow Geroski (1990:591) and opt for the risk of excessive inclusion in exchange for a more comprehensive account of monopoly power[9]. Several regressions using only one or different subsets of the variables lead to similar results.

Appropriability, Opportunity and Demand Conditions

As noted above, several studies (e.g. Cohen et al. (1987), Levin et al. (1985)) have pointed out that the relation between R&D and market structure is weak after controlling for how easy is for the companies to appropriate the results of their R&D effort or for technological opportunities and heterogeneity across industries. I include industry dummies that capture technological opportunities and appropriability as well as other non-observable industry specific characteristics. Following Beneito (2001) I also include another variable to further control for appropriability that is defined as the total number of patents per industry over total number of companies that patent in that industry. The logic behind this measure is that more patents per innovative company should reflect that the results from innovating are easier to appropriate. Finally, demand conditions are captured by a dummy that takes value 1 if the firm

---

[9] Although separate regressions showed that all the market structure variables are correlated, correlation among them is not strong enough to create multicollinearity problems.

reports an increasing demand and 0 otherwise and by another dummy that takes value 1 if the firm reports to have increased its market share.

Other controls

I take into account several additional variables that influence the decision to undertake R&D. First I include the number of workers as a measure of size in the first equation and total sales in real terms in the second step. Many studies have tested the hypotheses that R&D intensity increases with firm size, finding little support (see Kamien and Schwartz (1982) or Cohen and Levin (1989) for a review). Studies that use probit models however, although usually focusing on other factors, unambiguously conclude that the probability of undertaking R&D activities increases with size (e.g. Cohen et al. (1987), Polhmeier (1992), Crèpon et al. (1998) or Cassiman and Veugelers (2002).

Second I control for financial constraints faced by firms by including the ratio of self-equity to total debt. In the second step, consistent with our interpretation of variations on R&D as a short run decision I prefer to use instead short run debt over total debt[10]. Several studies (e.g. Hall (1992) or Himmelberg and Petersen (1992)) find that the link between R&D and size or market structure is affected by the availability of funds for R&D investment. Theoretical reasons behind this relation are twofold. First, that R&D projects are difficult to fund with outside sources because of asymmetric information, exacerbated by the secrecy and uncertainty implied in innovative activities. Second, and related to the first issue, companies in more concentrated markets are supposed to be better able to self-finance from own profits.

[10] Short run debt over total debt is highly correlated with long run debt over total debt and self-equity ratio. Using long run debt or self-equity ratio instead in the regressions do not change the statistical significance nor the interpretation of the results.

Based on previous literature, I also include firm's age as determinant of R&D. Klepper (1996) presents a theoretical model in which the number of innovations per firm is affected by, and declines with age. Huergo and Jaumandreu (2004) test the implications of that model, and find positive support for the hypothesis that new firms are more innovative.

Finally, McDonald (1994) and Bertcheck (1995) find that exposure to external trade enhances R&D outcomes. Cassiman and Martinez-Ros (2005) obtain similar results for the Spanish case. I include export intensity to control for exposure to international markets trade.

4.2. Measurement problems

Data management

To avoid possible estimation problems caused by outliers I eliminated observations with price-cost margins greater than 1 and below –0.5. I also eliminated observations for firms that experienced mergers, excisions or other similar processes likely to affect the comparability of the data for the same firm across years. I only eliminated these firms from the sample for the years after the merger or acquisition. Finally in order to make the panel slightly more balanced and to allow for the inclusion of some dynamics I also eliminated firms for which there are less than two observations in the sample. All this procedures leaded to a final sample of 18,580 observations, which is more than 90 per cent of the original maximum sample size of 20,505. Table 1 includes summary statistics of the main variables in our sample.

Missing observations

The high rate of missing observations in one of the key variables creates a potential sample selection problem. Around 40 per cent of the firms do not report a concentration ratio for any of the markets in which they operate, presumably because some companies have problems identifying their market and rivals. The summary statistics provided in table 1 show that the companies that do not report a concentration ratio are smaller and operate, according to other indicators, in more competitive markets compared to the full sample. Unless the data for this variable were missing completely at random (very unlikely), the estimates of the sub-sample including Concentration Ratio as a regressor, would lead to inconsistent estimates of the population parameters[11].

In order to solve the problem I performed a multiple imputation procedure proposed by Rubin and Schenker (1986), and Rubin (1987) and applied in Brownstone and Steimetz (2005). The method provides consistent estimates when the missing observations are not missing completely at random. When missing observations are missing completely at random, it provides more efficient estimates than deleting the observations. The procedure consists of drawing the missing data from their conditional distribution. A linear regression model is used to estimate the conditional distribution. The dependent variable of this regression is concentration ratio, the variable with missing values. Let denote it by CR4. In order to obtain valid estimates I included as regressors of this regression all the other variables used in both the first and the second step described above. Previous studies using the same or similar databases use *ad hoc* or single

[11] See King et al (2001) for the differences between data missing completely at random, data missing at random and data not missing at random. Here I assume that data are missing at random, which means that the missingness is not endogenous and therefore can be explained with available information from our database. See also Rubin and Schenker (1986) and Rubin (1996).

imputation methods to impute the missing values. These methods lead to biased estimates and do not account for the estimation error of the imputation process. Another way of imputing missing values that has been used (i.e. Beneito (2003)) is to assign to each missing observation the mean of the concentration ratio of those firms within each industry that report the same number of competitors and operate in the same geographical market (regional, national, or international). I used this variable as one of the additional regressors in my imputation model and therefore estimated the following regression:

$$\frac{CR4}{(1-CR4)} = \alpha X + \delta Z + u \qquad [5]$$

Where, X is the matrix of covariates used in both the first and the second step described above; Z is the matrix of additional covariates from our database including the mean of concentration ratio as computed in Beneito (2003) and α and δ are vectors of parameters to be estimated. I use the logit transformation of *CR4* as the dependent variable in this model to avoid predicted values to be out of the [0,1] interval[12]. I obtained an adjusted R-squared of 0.83 in this regression showing a good fit of the imputation model to the data. Then I proceeded to obtain *m* different sets of imputations of *CR4* by the following procedure described in Steimetz and Brownstone (2005):

1. Draw σ^2_* by dividing the residual sum of squares from regression [5] by an independent draw from a χ^2 distribution with degrees of freedoms equal to the number of estimated parameters in regression [5].
2. Draw a vector of residuals u_* from a $N \sim (0, \sigma^2_* I_N)$ distribution.

[12] We transformed back the predictions once the different imputation were obtained.

3. Draw [θ_*]=[α_*, δ_*]', from a $N\sim(\theta, V)$ where θ is the estimated vector of coefficients and V is the estimated covariance matrix of regression 5.

4.Construct (CR4/1-CR4)=$\alpha X_* + \delta_* Z + u_*$

5. Undo the logit transformation to obtain the set of imputed values of CR4.

6. Repeat steps 1 to 5 until the desired number of imputations, *m*, is reached.

Once the *m* imputed vectors were obtained I estimated the first and second step (probit and RE regression) *m* times, using each time a different set of imputations. I then obtained the final parameters of the model using Rubin's rules [equations 2 to 5 above].

The vector of coefficients reported in the results sections below is therefore the average of the coefficient vectors obtained with each imputed vector. The variance covariance matrix Ω, is a weighted combination of matrices U and W. U is the average of the estimated covariance matrices obtained using the *m* different sets of imputations. W accounts for the variance between the *m* different vectors of parameters estimates, and reflects the measurement error due to the imputation process. The matrix Ω depends on the number of imputations, suggesting that more imputations would lead to reducing the error due to the imputation process. However, Rubin (1987:114) shows that between 3 and 10 imputations is usually enough to obtain valid results. Here I used m=20 to effectively minimize the error due to the imputation process.

Endogeneity

It has been argued that companies that perform R&D are more likely to obtain higher profits in the future, and market structure will shift accordingly. If innovation and market structure are both simultaneously determined, the estimations suffer from an endogeneity problem. Previous literature that tried to test endogeneity in the relation between market structure and innovation is contradictory (Pohlmeier (1992), Kraft (1989)).

The problem is difficult to address because there are lags between successful innovations, and the accumulation of profits able to affect the monopoly power of the company. Furthermore, it seems plausible that the lags vary from firm to firm[13]. In our sample, firms are observed during a short period of time (five years on average) over which market conditions are unlikely to change in response to firms' innovative activities. In that sense our panel being short should alleviate the possible endogeneity. However we performed several instrumental variable tests using similar instruments to those proposed in previous literature (Kraft (1989:334)) showing no evidence for endogeneity.

5. RESULTS

First step

Table 2 shows the result for the first-step probit models. As a benchmark case I report the results of the standard pooled probit model in the first

[13] In addition, although it seems plausible that more R&D spending will lead to better results in the future for companies that do R&D, it is not clear that carrying R&D is automatically translated into future higher profits. According to one of the hypotheses discussed above, firms decide to carry R&D activities if the profits to be obtained by doing so are greater than the profits of not carrying R&D. Therefore doing R&D is supposed to generate higher expected profits and to affect market structure for firms that decide to do R&D but would lead to worst results for firms that decide not to become R&D doers. But this endogeneity/selection problem is automatically corrected by our sample selection procedure that accounts for the fact that we only observe R&D of firms that decide to do R&D. And for those firms, endogeneity of the relation can be corrected by standard instrumental variable methods if an adequate instrument is available.

two columns. The regression in column 1 includes only the observations that report concentration ratio while model 2 is estimated including all the observations after the multiple imputation. In columns 3 and 4 I show the results of the random effects model in the two cases, with and without the multiple imputation.

The results in columns 1 to 4 provide support for the hypothesis that monopoly power increases the probability that firms become R&D doers. All the monopoly variables have the expected signs and are generally significant. The probability of conducting R&D activities increases with the price margin of the company and with an increase in the market share. Price margins are significant in all the models indicating that actual price margins are likely to affect the expectations the company regarding future monopoly power. The coefficient estimates of the Market Share dummy and of the Number of Competitors dummy have negative values indicating that more competition as reflected by more competitors in the market or by a non-singnificant market share, have, as expected, a negative impact on the probability of conducting R&D activities. The coefficient estimate of the dummy that measures an increase of market share however is not significant in our preferred specification of column 4 (it is significant in the other three though). This result is in accordance with the interpretation of the first step as a long-run decision. An increasing market share may be reflecting short run fluctuations that are not as likely to affect the decision to conduct R&D as much as other long term characteristics of the market.

Concentration ratio has a significant inverted U-shape relationship to the probability of doing R&D. This result shows that the well-known U-shape shown in other papers in which R&D intensity (Cohen, Mowery (1985) or the number of patents (Aghion (2005) is the dependent variable also appears in a limited dependent variable specification.

The results suggest that using a unique measure of market structure can be misleading. The correlation among the five variables used here is relatively small indicating that they are not measuring the same thing. Moreover, including just one measure can lead to different conclusions (the effect of Concentration Ratio is different from the other variables) when interpreted separately from the other variables that reflect market power. For this reason the approach followed here including up to five different variables seems to better capture what is the overall effect of monopoly power on innovation activities. In the models presented in table 2 the overall effect is positive, corroborating Shumpeter predictions and contradicting most of the previous studies on the topic in which a negative or an insignificant effect was found. A more careful look at the magnitude of the combined effects of the different variables is provided below.

The analysis of the control variables included in table 2 shows that demand conditions, and participating in international markets increases the probability of doing R&D at statistically significant confidence levels. Industry characteristics, as reflected by industry dummies are also significant in all the specifications. The coefficient of Age is also significant and has a positive sign reflecting that older companies are more likely to perform R&D. This result contradicts previous literature in which newer companies are found to be more innovative (Klepper (1996), Martinez-Ros (2005)). A possible explanation for this unexpected result is that age might be here reflecting that companies that perform R&D activities are more likely to survive to market competition than non-innovative ones.

The estimate of the financial constrains as measured by self-equity ratio has the expected sign although the coefficient is not statistically significant in some of the specifications. Finally, our measure of appropriability does not seem to be a determinant of the probability of performing R&D as reflected by the low statistical significance of the estimates and the change in signs in the different specifications.

Second step

The second step regression equations include a correction term to account for the selection bias caused by the fact that we only observe R&D intensity of firms that carry R&D activities. In this table we see that after controlling for appropriability, industry characteristics and demand conditions, the relation between R&D intensity and market structure vanishes. All the variables that reflect monopoly power have either a negative effect on R&D intensity (CR4) or non-significant t-statistics except for the dummy for the number of competitors[14]. Therefore our results are in accordance with most of the previous studies and specifically with Geroski (1990) who finds no empirical evidence of a positive relation between monopoly power and R&D intensity once heterogeneity among industries is taken into account.

Most of the control variables in the regression equation have the expected signs. Appropriability, R&D stock, and size measured by total sales have an unambiguous positive and significant effect on R&D per worker after controlling for industry characteristics. The coefficients of Export intensity and Age have also positive signs but are not significant in the model of column 3. The variable that captures demand conditions has the expected positive sign in most of the specifications while the coefficient of the variable that measures short run financial constraints presents an

[14] Several specifications in which logs of the dependent and independent variables and quadratic terms of the continuous covariates where included leading to the same conclusion but to worse overall fit of the models.

unexpected positive sign in one of the models (column 3) and is not significant in columns 1 and 4. Finally, the terms that control for the selection obtained in the first step regressions are always strongly significant reflecting the usefulness of the correction terms.

To sum up the results provide support for the Shumpeterian hypotheses that monopoly power provides better than perfect competition for firms to invest in R&D if the decision of whether to do R&D is the dependent variable, but do not support the hypothesis when the decision is how much to spend.

In a paper that follows a similar approach including also a variety of different measures of monopoly power Geroski (1990) finds no effect of the monopoly variables on innovation. The results in table 2 show that when a probit model of the probability of conducting R&D is used to estimate the effect of monopoly power, Shumpeter's predictions hold. However my results do not contradict Geroski (1990) because the second stage regressions in which the "amount" of innovative activity is estimated, the effect of the monopoly variables vanishes leading to the same conclusion as Geroski (1990). Therefore the results suggest that a more careful look at the first-step decision of companies regarding whether to do R&D seems helpful to provide some light to the nowadays inconclusive results of the literature on market structure and innovation. The results also highlight the importance of including different measures of monopoly power to capture market structure. Related to this, our results support Sherer (1967) Levin, Cohen and Mowery (1985), or Aghion`s (2005) inverted U-relationship between R&D and competition if concentration ratio is the only relevant variable to measure market structure but show that using just one measure could be misleading to infer what the overall effect of market structure on innovation is. Finally the results also point out that although the conclusions regarding the monopoly variables do not change using the multiply imputed full sample or the sub-sample of non imputed observations, the magnitude, signs and significance of some of the control variables do change.

Magnitude of the effects

An implication of the results is that the policy maker could increase the number of firms that conduct R&D without diminishing the amount of R&D of already innovative firms. In this section I provide simple estimates of the increase of the probability of firms conducting R&D that can be expected from different plausible scenarios.

In column 1 of table 3 I present the marginal effects of the relevant market structure variables from the probit model of column 4 in table 1 (my preferred specification). Just looking at the marginal effects does not give enough information about the relevance of the effects because of the different units of the variables and the joint effects. For this reason I provide the total effect on probability of three different plausible scenarios.

In the first one, let suppose a competitive market in which many firms compete and appropriability of inventions is low due to weak patent laws or because it is easy for rivals to invent around patents. In such a case we can assume that increasing the patenting rights of the innovator could lead to higher profit margins of the innovator. If the new gained market power means that others cannot produce similar products, we can also assume that innovators would increase their market share because now they would be the only ones to offer the product. In column two I provide the total marginal effect of such a policy on the probability of becoming an innovator. I suppose an increase in profit margin of 0.01 after the new policy and an increasing market share (dummy changing from 0 to 1). In such a case, the probability of reporting positive R&D would increase in 0.01 units or 2.6% of the actual probability of conducting R&D (0.38 is the sample frequency).

In the other two columns I provide the results of a different scenario in which competition authorities are thinking about authorizing a merger of two companies leading to a higher concentration ratio in a given market. Due to the quadratic influence of concentration ratio on the probability of conducting R&D, the results change depending on what is the starting market structure before the merger. In column two I consider the case of authorizing a merger in a competitive market. If the increase in concentration ratio is the only result of the merger (I assume profits for other companies do not change) and concentration increases in just 5 points (say from 10 to 15 per cent) the probability of carrying R&D would increase in 0.011 or 2.9 per cent. If previous to the merger no company had a significant market share (plausible if the starting point is a competitive market) the increase rises to 0.05 probability units or 13%.

Finally, if the starting point before the merger is an already concentrated market (assume concentration moves from 65 to 70 percent) the results change drastically and the negative effect of concentration ratio would lead to a decrease of probability of 0.007 units, or almost a 2%. This is an extreme case in which none of the other market structure variables vary and therefore cannot counterbalance the effect of more concentration. This case shows that an inverted U shape total effect, as reported by some papers can also be obtained given our data in certain scenarios.

6. EXTENSIONS AND ROBUSTNESS

In this section I explore further the first step decision of firms by allowing for some dynamics and state dependence. Our summary statistics show that firms that perform R&D persistently do so. Roughly 50% of firms in our database do not change their R&D status over the observed period. This shows evidence of a strong state dependence on firm's research policy that has not been taken into account so far in our model. I test now the effect of state dependence in the first step probit estimates following the approach proposed by Woolridge (2005). I estimate the following Conditional Maximum Likelihood dynamic random effects probit model:

$$\Pr(y_{it}^{*}>0) = \Pr(X'B+a_0+a_1y_{i1}+a_2y_{it-1}+u_i+v_{it}>0) \qquad [8]$$

Where I have substituted u_{it} in equation [1] by $u_{it}=a_1y_{i1}+a_2y_{it-1}+u_i+v_{it}$. The model therefore deals with state dependence in two ways. First, I allow for previous state to affect current state, and second, I deal with the initial conditions problem by including the R&D status in the initial observed period as an additional regressor. Not including the initial state would consider it as exogenous and uncorrelated with the disturbance u_{it}. This is not a reasonable assumption given that the first observed period corresponds to a different period in each firm's life and given that it would mean that the initial R&D status is not related with their idiosyncratic factor u_i. Ignoring the unobserved persistence would therefore overstate the effect of previous R&D status on current R&D decisions. This is corrected here by including the initial state as an additional regressor. Woolridge (2005) provides a detailed description of the properties of this estimator.

I estimated the dynamic model of equation [8] on a balanced sub-sample of the database. Given that the balanced sub-sample reduces our total number of observations to 5490 our results cannot be considered representative of the whole population as the previous ones were. However they are still useful as a robustness check. In table 5 I show the results of the estimation. I carried the same multiple imputation procedure described above to estimate this model. Therefore the estimates of table 5 are computed using formulas [2] to [5] above.

The results show that current R&D status is strongly dependant on previous R&D as reflected by the high significance level of the coefficient of the initial status. However, the conclusions reported in previous sections regarding the effect of market structure on the probability of reporting positive R&D do not change with the introduction of dynamics to the model. All the market structure variables have the same signs and similar significance levels (except for the dummy that captures the number of competitors that is now non-significant). Moreover the magnitude of the coefficients is also similar. Therefore, although a more careful look at the effects of dynamics is certainly worthwhile and left for future work, the results offered here show that the positive effect market structure on the probability of performing R&D are robust to the inclusion of some dynamics and state persistence to the model.

7. CONCLUDING REMARKS

Theoretical models predict that more monopoly power has both positive and negative effects on R&D. Depending on model assumptions, one or the other effect is stronger. Empirical papers have tried to provide evidence on what of the two effects dominates in real data. Many papers have found that there is no evidence of a significant effect after controlling for industry characteristics, appropriability, demand conditions and technological opportunities.

I find the opposite conclusion by focusing on the decision of firms of whether to perform R&D activities or not. I argue that the decision of whether to do R&D or not and decision of how much to spend on R&D are related and depend on similar factors. When firms decide to perform R&D, they make a careful evaluation of what are the returns to be expected from the different possible levels of R&D spending given expected market structure. Due to the long-run nature of innovation activities, that involve uncertainty over the invention and high fixed costs, the decision on how much to spend on R&D is not very likely to experience big changes with short-run changes on market structure. Therefore, I use a two-step procedure in which in the first equation I estimate the effect of different market structure variables on the probability of a firm performing R&D activities and in the second equation I estimate the effect of market structure on R&D intensity given a firm decides to undertake R&D. When controls for industry and market characteristics and technological opportunities are taken into account, R&D intensity is not affected by monopoly power, but the probability of a firm being innovative increases with it.

The results have interesting policy implications. Policy measures that increase market power of already innovative firms are not likely to increase overall innovative effort. However, changes in market structure can create incentives for new companies to perform R&D activities. This conclusion comes from the fact that market structure variables have different effect in the decision to undertake R&D than on how much to spend.

Finally, while the debate on the effect of market structure on R&D is far from being solved, the results suggest that more attention should be paid at the effect of market structure on the decision of firms to perform R&D or not. A more careful look to which policy channels can create more incentives for new innovators without distorting the activities of already innovative firms is also suggested by the results.

8. REFERENCES

Aghion, P.; Bloom, N.; Blundell, Griffith, R.; Howitt, P. (2005): "Competition and Innovation: An inverted U-Relationship", *Quarterly Journal of Economics*, 701- 728.

Angelmar, R. (1985): "Market Structure and Research Intensity in High Technological Opportunity Industries", *Journal of Industrial Economics*, 34, pp. 69-79.

Arrow, K. (1962): "Economic Welfare and the Allocation of Resources to Invention", in Nelson, R. R. (ed), *The Rate and Direction of Economic Activity*, Princeton University Press, N.Y.

Bertschek, I. (1995): "Product and Process Innovation as response to increasing imports and foreign direct investment", *Journal of Industrial Eocnomics*, 43, 341-357.

Blundell, R.; Griffith, R .; Van Reenen, J. (1999): Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms, *Review of Economic Studies*, vol. 66. 529-554.

Bound, J. (1984): "Who does R&D and Who Patents? In Griliches, Z. (ed), R&D Patents and Productivity", University of Chicago Press.

Brownstone, D. (1991): Multiple Imputations for Linear Regression Models, Mathematical Behavioral Science, Technical Report Series, MBS 91-37, Irvine, California.

Cassiman, B.; Martinez-Ros, E. (2005): "Innovation and Exports: Evidence from Spanish Manufacturing", *mimeo.*

Cohen, W. M.; Levin, R. C. (1989): Empirical Studies of Innovation and Market Structure, in Schmalanse, R.; Willing, R.D. (eds), *Handbook of Industrial Organization*, vol II.

Cohen, W. M.; Levinthal, D.A.(1989): "Innovation and Learning: The Two Faces of R&D", *The Economic Journal*, 99, pp. 569-596.

Cohen, W.M.; Klepper, S. (1996): "A reprise of size and R&D", *Economic Journal*, 106, 925-995.

Cohen, W.M.; Levin, R. C.; Mowery, D. C. (1987): "Firm Size and R&D intensity: A re-examination", *Journal of Industrial Economics*, 35, pp. 543-563.

Crépon, B.; Duguet, E.; Mairesse, J. (1998): "Research, Innovation, and productivity: An econometric Analysis at the firm level", *NBER Working Paper* 6696.

Culbertson, J.D.; Mueller, M. F. (1985): "The influence of Market Structure o Technological Performance in the Food-Manufacturing Industries", *Review of Industrial Organization*, vol. 2, pp. 40-54.

Dasgupta, P.; Stiglitz (1980): "Industrial Structure and the Nature of Innovative Activity", *Economic Journal*, vol 90, pp 266- 293.

Dustman, C.; Rochina-Barrachina, M.E. (2000): "Selection Correction in Panel Data Models: An Application to Labour Supply and Wages", *IZA Discussion Papers*, No. 162.

Fundenberg, D.; Gilbert, R. Stiglitz, J.; Tirole, J. (1983): "Preemption, Leapfrogging, and Competition in Patent Races", *European Economic Review* 22, pp. 3-31.

Geroski, P. A. (1990): "Innovation, Technological Opportunity, and Market Structure", *Oxford Economic Papers* 42, pp. 586-602.

Gilbert , R.(2006): "Market Structure and Innovation, Where are We in the Competition-Innovation Debate", *NBER, Innovation and the Economy 6, pp. 159-215*.

Gilbert, R. Newbery, D. (1982): "Preentive Patenting and the Persistence of Monopoly", *American Economic Review*, vol 72, pp. 514-526.

Harris, M.N.; Rogers, M.; Siouclis, A. (2001): “Modelling Firm Innovation using Panel Probit Estimators”, *Melbourne Institute Working Paper* No. 20/01.

Himmelberg, C.P.; Petersen, B.C. (1994): “R&D and internal finance: A panel study of small firms in high tech industries”, *Review of Economics and Statistics*, 76, pp. 38-51.

Horowitz (1962): “Firm size and research activity”, *Southern Economic Journal*, 28: 298-301.

Huergo, E.; Jaumandreu, J. (2003): “How Does Probability of Innovation Change with Firm Age?”, *Small Bussiness Economics* 22, 193-207.

Jaffe, A. B. (1986): “Technological Opportunity and Spillovers of R&D”, *American Economic Review*, 76, pp. 984-1001.

Kamien, M.; Schwartz, N. (1972): *“Timing of Innovations under Rivalry, Econometrica”*, 90, pp. 43-59.

Kamien, M.; Schwartz, N. (1976): “On the degree of Rivalry for Maximum Innovative Activity”, *Quarterly Journal of Economics*, 90,pp. 245-260.

King, G.; Honaker, J.; Joseph, A.; Scheve, K. (2001): "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation," *American Political Science Review*, Vol. 95, No. 1 (March, 2001): pp. 49-69.

Levin, R. C.; Reiss, P. C. (1984): Test of a Schumpeterian Model of R&D and Market Structure, in Griliches, Z. (ed), *R&D Patents and Productivity*, University of Chicago Press.

Levin, R.C.; Cohen, W.M.; Mowery, D. C.(1985): “R&D appropriability, opportunity and market structure: New Evidence of Some Schumpeterian Hypothesis”, *American Economic Review*, *Papers and Proceedings*, 75, 20-24.

Loury, G. C (1979): “Market Structure and Innovation”, *Quarterly Journal of Economics*, vol 93, pp. 395-410.

Lunn, J. Martín, S. (1986): “Market Structure, Firm Structure and Research and Development”, *Quarterly Review of Economics and Business*, 26, pp. 31-44.

MacDonald, J.M. (1994): “Does Import Competition Force Efficient Production?”, *Review of Economic and Statistics*, vol. 76, pp. 721-727.

Mansfield, E. (1964): “Industrial Research and Development expenditures: Determinants, Prospects, and Relation of Size of Firm and Inventive Output”, *Journal of Political Economy*, 72, pp. 319-340.

Martinez-Ros, E. (2000): “Explaining the decisions to carry out product and process innovations: The Spanish case”, *The Journal of High Technology Management Research*, 10, 223-242.

Nickel, S.J. (1996): “Competition and Corporate Performance”, *Journal of Political Economy*, vol. 104, pp. 724-746.

Okada, Y. (2004): “Competition, R&D and Productivity in Japanese Manufacturing Industries”, NBER Working Paper No. 11540.

Pohlmeier, W. (1992): “On the Simultaneity of Innovations and Market Structure”, *Empirical Economics* 17, 253-272.

Reinganum, J. F. (1983): “Innovation and Industry Evolution”, *Quarterly Journal of Economics*, vol. 100, pp. 81-99.

Reinganum, J. F. (1985): “Innovation and Industry Evolution”, *Quarterly Journal of Economics* 100, pp. 81-99.

Ridder, G. (1990): Attrition in multi-wave panel data, in Hartog, J., Ridder, G.; and Theeuwes, eds, *Panel data and Labor Market Studies*, North Holland, Ámsterdam, pp 45-67.

Romer, P. M. (1990): Endogenous Technological Change, *Journal of Political Economy*, 98(5), 71-102.

Rubin, D. (1987): Multiple Imputation for Non-response in Surveys, Wiley, New York.

Rubin, D. Schenker, N. (1986): “Multiple imputations for interval estimation form simple random samples with ignorable nonresponse”, *Journal of the American Statistical* Association 81, 366-374.

Sanyal, P. (2001): *Essays on Research & Development*, Ph.D. dissertation, University of California, Irvine.

Sanyal, P. and Cohen, L. R. (2004): "Deregulation, Restructuring and Changing R&D Paradigms in the US Electric Utility Industry", Working Paper available at Social Science Research Network (SSRN): http://ssrn.com/abstract=556633.

Scherer, F. M. (1967):"Market Structure and the Employment of Scientists and Engineers", *American Economic Review* 57, pp. 524-531.

Scherer, F.M.(1992): Schumpeter and Plausible Capitalism, *Journal of Economic Literature*, 30, pp. 1416-1433.

Schumpeter, J. A. (1942): *Capitalism Socialism and Democracy*, New York: Harper and Brothers. (Harper Colon edition, 1976).

Scott, J. T.(1984): Firm versus Industry Variability in R&D intensity, in Griliches, Z. (ed), *R&D Patents and Productivity*, University of Chicago Press.

Smolny, W. (2003): "Determinants of Innovation Behavior and investment Estimates for West-German Manufacturing Firms", *Economics of Innovation and New Technology*, 12, 449-463.

Steimetz, S.S.C.; Brownstone, D. (2005): "Estimating commuters "value of time" with noisy data: a multiple imputation approach", *Transportation Research Part B*, 39, pp. 865-889.

Syrneonidis, G. (1996): "Innovation. Firm Size and Market Structure: Schumpeterian Hypotheses and Some New Themes", *OECD Studies* No. 27.

Tingvall, P.G.; Poldahl, A. (2006); "Is there really an inverted U-shape relation between competition and innovation?", *Economics of Innovation and New Technology* 15, pp. 101-118.

Vella, F.; Verbeek, M. (1999): "Two-step estimation of Panel Data Models with Censored Endogenous Variables and Selection Bias", *Journal of Econometrics*, 90, pp 239-264.

Verbeek, M.; Nijman, Th.E. (1992): "Testing for Selectivity Bias in panel data models", *International Economics* Review 33, pp 681-703.

Woolridge, J.M. (1995): "Selection Corrections for panel data models under conditional mean independence assumptions", *Journal of Econometrics*, 68, 115-132.

Woolridge, J.M. (2005): “Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity”, *Journal of Applied Econometrics*, 20, 39-54.

Zabel, J. (1992): “Estimating Fixed and Random Effects Models with Selectivity”, *Economics Letters* 40, 269-272.

APPENDIX 1. VARIABLE DEFINITION

Age	Number of years since the company was created
Appropriability	Total number of patents over the total number of firms in the sector
Concentration	Sum of market share of main four industries in the product markets were the company operates weighted by the share of the sales in this markets over total sales of the company
Concentration squared	Square of Concentration
Demand conditions	Dummy variable that takes the value 1 if the firm reports that the demand for the market is increasing
Equity-debt ratio	Sum of total own funds of the firm over total debt
Export dummy	Dummy variable that takes the value 1 if firm operates in international markets
Export intensity	Total exports over total sales
Market share dummy	Takes the value 1 if the firm reports to have a non-significant market share
Market share growth	Takes the value 1 if the firm reports that its

dummy	market share has increased over the year
Number of competitors dummy	Takes value 1 if the firm reports that there are no firms in the market with significant market power
Number of workers	Total number of workers at the end of each year
Price Margin	Total sales plus stock variation plus other revenues minus intermediate consumption minus labor costs
Sales	Total sales of the firm in a given year real terms
Short run debt	Ratio of short run debt over total debt
Stock R&D dummy	Dummy variable that takes the value 1 if the firm reports to have obtained an innovation (patent) in the last four years

Table 1. Summary Statistics: Mean (Std. Dev.)

Variables	All firms	R&D doers only	Non R&D doers only	Balanced subsample	Non-Missing CR4	Missing CR4
Positive R&D	0.37 (0.48)			0.33 (0.47)	0.39 (0.49)	0.32 (0.47)
Age	32.1 (23.88)	39.6 (26.11)	27.64 (21.22)	35.01 (23.42)	32.27 (23.54)	31.8 (24.48)
Appropriability	0.53 (0.92)	0.62 (0.87)	0.47 (0.94)	0.54 (1.05)	0.54 (0.92)	0.51 (0.92)
Concentration		44.78 (27.7)	35.5 (33.86)		37.4 (32.32)	
Demand conditions	0.29 (0.45)	0.35 (0.48)	0.25 (0.43)	0.28 (0.45)	0.29 (0.46)	0.27 (0.45)
Equity-debt ratio	0.43 (0.23)	0.45 (0.21)	0.42 (0.25)	0.47 (0.23)	0.43 (0.23)	0.42 (0.23)
Export dummy	0.21 (0.41)	0.36 (0.48)	0.13 (0.34)	0.21 (0.41)	0.23 (0.42)	0.19 (0.39)
Export intensity	0.17 (0.24)	0.27 (0.27)	0.1 (0.21)	0.17 (0.25)	0.17 (0.25)	0.16 (0.24)
Market share dummy	0.48 (0.5)	0.28 (0.45)	0.6 (0.49)	0.51 (0.5)	0.44 (0.5)	0.56 (0.5)
Market share growth dummy	0.28 (0.45)	0.33 (0.47)	0.26 (0.44)	0.26 (0.44)	0.29 (0.45)	0.28 (0.45)
Number of competitors dummy	0.23 (0.42)	0.13 (0.34)	0.29 (0.45)	0.2 (0.4)	0.35 (0.48)	0 0
Number of workers	261 (819.3)	531.8 (1251.24)	100.66 (270.046)	149.269 (269.24)	286.81 (871.57)	212.6 (708.89)
Price Margin	0.09 (0.12)	0.1 (0.11)	0.09 (0.12)	0.1 (0.11)	0.10 (0.11)	0.09 (0.11)
Sales	4.24+07 (2.03+08)	9.12+07 (3.14+08)	1.36+07 (7.08+07)	1.91+07 (4.64+07)	4.81+07 (2.20+08)	3.18+07 (1.64+08)
Short run debt	0.45 (0.22)	0.44 (0.19)	0.45 (0.23)	0.42 (0.21)	0.45 (0.22)	0.46 (0.22)
Stock R&D dummy	0.28 (0.45)	0.45 (0.5)	0.18 (0.38)	0.29 (0.45)	0.29 (0.45)	0.26 (0.44)
Number of observations	18580	6434	10865	6624	6017	11282

Table 2. First step. Probit estimates. Dependent variable: R&D

	1. Pooled Probit	2.Multiple Imputation Pooled Probit	3. Random Effects Probit	4.Multiple Imputation Random Effects Probit
Regressors				
Constant	-0.2697** (0.1308)	-0.4605* (0.2484)	-0.6366** (0.2801)	-1.7188* (0.918)
Market structure variables				
Concentration Ratio	0.0121*** (0.0023)	0.0082*** (0.0020)	0.0095*** (0.0044)	0.0072** (0.0037)
Square of Concentration Ratio	-0.0001*** (0.00002)	-0.0001*** (0.00002)	-0.0002*** (0.00004)	-0.0001*** (0.00003)
Price Margin	0.462*** (0.1243)	0.5384*** (0.0973)	0.3781* (0.2201)	0.2883* (0.1687)
Market share growth dummy	0.0821** (0.0344)	0.1240*** (0.0266)	0.1006* (0.0572)	0.0402 (0.0453)
Market share dummy	-0.3919*** (0.0423)	-0.4684** (0.0279)	-0.3487*** (0.0728)	-0.3383*** (0.05)
Number of competititors dummy	-0.2223*** (0.0534)	-0.1189*** (0.0389)	-0.3148*** (0.0976)	-0.1219* (0.073)
Controls				
Size	0.0008*** (0.0002)	0.0007** (0.0003)	0.0001 (0.0002)	0.0004*** (0.0001)
Appropriability	-0.0141 (0.0221)	-0.0004 (0.0161)	-0.0492 (0.0426)	0.0065 (0.0297)
Demand Conditions Dummy	0.145*** (0.0345)	0.1364*** (0.0267)	0.1542*** (0.0571)	0.1065** (0.0449)
Self-equity ratio	0.0767 (0.0610)	0.1278*** (0.0467)	0.0867 (0.1731)	0.1222 (0.1311)
Export dummy	0.4971*** (0.0383)	0.4563*** (0.0289)	0.4314*** (0.0657)	0.4054*** (0.0547)
Age	0.0063*** (0.0008)	0.0083*** (0.0005)	0.0224*** (0.0018)	0.0225*** (0.0014)
Group means included	no	no	yes	yes
Industry dummies	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes
Number of observations	10651	18580	10651	18580

Notes: *** Significant at 99%, ** Significant at 95%, Significant at 90%
Standard errors in parentheses

Table 3. Second stage regression: dependent variable R&D over workers

	1. Pooled OLS	2.Multiple Imputation Pooled OLS	3. Random Effects	4.Multiple Imputation Random Effects
Regressors				
Constant	1331.32***	1075.79**	2628.91***	763.16
	(648.79)	(469.32)	(320.19)	(1473.12)
Market structure variables				
Concentration Ratio	-12.50***	-30.39***	-6.83	5.12
	(3.82)	(7.98)	(4.83)	(7.51)
Square of Concentration Ratio	0.1185***	0.2963***	0.0565	-0.0244
	(0.0342)	(0.0815)	(0.0441)	(0.0759)
Price Margin	403.84	496.73	202.03	-102.71
	(255.97)	(489.11)	(288.83)	(413.05)
Increase in market share dummy	-22.01	125.29	-38.54	62.23
	(48.06)	(107.80)	(57.19)	(85.27)
Non Significant market share dummy	33.98	-10.89	95.32	164.53
	(95.83)	(151.48)	(82.21)	(124.30)
Number of competititors dummy	-375.41***	-459.52***	-306.23***	-354.19**
	(92.49)	(151.59)	(102.66)	(160.49)
Controls				
Industry dummies	yes***	yes***	yes***	yes***
year dummies	yes***	yes***	yes***	yes***
Sales	3.24e-06***	1.58e-06***	5.11e-06***	1.91e-06***
	(1.51e-06)	(1.14e-07)	(5.84e-07)	(2.79e-07)
Appropriability	290.21***	141.78**	282.91***	164.39***
	(80.24)	(68.66)	(50.48)	(65.58)
Stock RD dummy	193.21***	733.17***	181.03***	717.89***
	(45.45)	(86.69)	(64.50)	(169.59)
Short run debt	-62.19	-509.81**	48.07	-363.16
	(125.63)	(216.05)	(207.55)	(281.05)
Export intensity	151.65	489.53**	203.09	400.69
	(96.18)	(197.03)	(225.31)	(303.58)
Age	2.47**	3.58*	2.53	10.73**
	(1.19)	-2.19	(1.93)	-4.35
Demand increase dummy	97.02	192.89***	55.52	51.24
	(67.99)	(24.93)	(57.96)	(87.31)
IMR polinomial	yes***	yes***	yes***	yes***
Number of observations	4250	6684	4250	6684

Notes: *** Significant at 99%, ** Significant at 95%, Significant at 90%
Standard errors corrected for first step probit in parentheses

Table 4. Magnitude of the Effects

Variables	Marginal Effect	Variables increase		
		Scenario 1	Scenario 2	Scenario 3
WeCR4 (from 10% to 15%)	0.0022	0	5	0
WeCR4 (from 65% to 70%)	-0.0014	0	0	5
Price Margin	0.0843	0.01	0	0
Number of competitors dummy	0.0374	0	1	0
Market Share dummy	0.1055	0	0	0
Mktshare increase dummy	0.0127	1	0	0
Total increase in probability		0.013	0.067	-0.007

Note: Total increase in probability is computed as the result of adding the marginal effect times the variable increase.

Table 4. Dynamic Probit estimates. Dependent variable: R&D

	Random Effects Dynamic Probit
Regressors	
Constant	-2.32***
	(0.506)
R&D in period 1	3.01***
	-0.27
Lag of R&D	0.0737
	(0.0864)
Market structure variables	
Concentration Ratio	0.0170***
	(0.0073)
Square of Concentration Ratio	-0.0001***
	(0.00007)
Price Margin	0.5862**
	(0.3444)
Market share growth dummy	0.0912***
	(0.0282)
Market share dummy	-0.1002
	(0.0958)
Number of competititors dummy	-0.0866
	(0.1333)
Controls	
Size	0.0022***
	(0.0006)
Appropriability	0.0310
	(0.0515)
Demand Conditions Dummy	-0.0115*
	(0.0817)
Self-equity ratio	-6.47e-06
	(0.2399)
Export dummy	0.3357***
	(0.1189)
Age	0.0069***
	(0.0033)
Group means included	yes
Industry dummies	yes
Year dummies	yes
Number of observations	5490

Notes: *** Significant at 99%, ** Significant at 95%, Significant at 90%
Standard errors in parentheses