Economics of Innovation and New Technology

ISSN: 1043-8599 (Print) 1476-8364 (Online) Journal homepage: http://tandfonline.com/loi/gein20

Profit differentials and innovation

Elena Cefis & Matteo Ciccarelli

Econ. Innov. New Techn., 2005, Vol. 14(1–2), January–March, pp. 43–61

Taylor & Francis
Taylor & Francis Group

PROFIT DIFFERENTIALS AND INNOVATION

ELENA CEFIS[a,b,*] and MATTEO CICCARELLI[c,†]

[a]*University of Bergamo, Bergamo, Italy;* [b]*Utrecht School of Economics, Utrecht University, Utrecht, The Netherlands;* [c]*European Central Bank, DG Research, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany*

(Received 16 May 2002; Revised 29 August 2003; In final form 13 February 2004)

The article aims to investigate empirically the effects of innovative activities on corporate profitability, using a panel of 267 UK manufacturing firms, over the period 1988–1992. Using the Bayesian approach to, explicitly, consider heterogeneity among firms, we find: (i) a positive and well-determined effect of innovation on profits that smoothly decreases as time passes by; (ii) a difference in profitability between innovators and non-innovators, greater when the comparison is between persistent innovators and non-innovators; and (iii) a long-run persistence in profit differentials.

Keywords: Innovation; Profitability; Firms differentials; Bayesian estimation; Panel data models

JEL Classification: O31; D21; C11; C23

1 INTRODUCTION

The article aims to investigate empirically some of the issues related to the impact of innovative activities on firm performances.

There are at least two alternative views about the way innovative activities affect the performance of firms. In the traditional view, innovations have only a transitory effect on the firm profitability by altering its competitive position in the short-run. The introduction of an innovation gives to the firm a temporary monopoly power by increasing firm's market share, which, allows for higher profits until other firms can imitate the innovation (Aghion and Howitt, 1992; Klepper, 1997).[1]

A second approach points out that innovations intrinsically 'characterize' a firm in that it creates a structural difference between innovating and non-innovating firms. According to this view, each firm owns different technological competencies that are firm specific, cumulative and emerge from the various learning processes the firm has passed through. These internal competencies, together with specific behavioral patterns, enable the firm to face better changes in the market in order to survive or even to obtain satisfactory profits over time (see Malerba and Orsenigo, 1995; Cohen and Levin, 1989; Dosi *et al.*, 1995).

*Corresponding author. E-mail: elena.cefis@unibg.it
† E-mail: matteo.ciccarelli@ecb.int

[1]This approach is common in the literature on the first mover advantage (e.g. Gorecki, 1986; Robinson *et al.*, 1994) and on the patent races (e.g. Tirole, 1988 (Ch. 10); Reinganum, 1989).

ISSN 1043-8599 print; ISSN 1476-8364 online
DOI: 10.1080/1043859042000232160

In order to discriminate these two alternative approaches, we pose three fundamental questions: (i) Are there differences in profitability between innovators and non-innovators? (ii) What are the effects of innovative activities on firm profits? (iii) Are these differences transitory or permanent? A clear answer to these three issues is important to shed light not only on the different patterns of innovation, but also on the understanding of the industrial dynamics.

Regarding the first two questions, the existence of correlation and the investigation of the channels of transmission between profit margins and innovations have been perennial concerns of industrial policy makers. At a first glance, there is nothing that prevents non-innovating firms to perform as well as innovating ones, either because non-innovators could develop different capabilities or because innovation may require a sunk-cost investment independent of innovation returns. On the other hand, answering the questions can be important for antitrust regulation to be able to discriminate among firms with relatively higher profits, between those who exploit a monopoly position and those who own capabilities and competencies that make them systematically better than the others. Previous studies (Geroski *et al.*, 1993; Geroski *et al.*, 1997) have found positive, though not well-determined direct effects of innovations on profitability in the short-run and large indirect effects due to the relative insensitiveness of innovating firms to adverse macroeconomic shocks.

Answering the last question is crucial. If, indeed, data show a certain relation between innovation and profitability, it is important to test whether differences in profitability are due to the production of specific innovations that alter temporarily the competitive position of the innovating firms, or whether they are due to the fact that innovating firms have higher competencies that allow them to face the challenges of the market better than the non-innovating firms.

In analyzing these issues, we want to take into account directly the heterogeneity among firms. Several empirical studies have shown that there is an evidence of heterogeneity among firms due to industrial and size classification, the access to capital market, etc. Furthermore, technological and organizational firm specific characteristics can cause heterogeneity at the firm level. Previous studies on profitability have not tackled the heterogeneity problem directly, but only indirectly considering an exogenous clusterization of firms by size or industrial sectors. Instead, we use the Bayesian approach which allows us to control general sources of firm heterogeneity in the population structure, acknowledging in this way that every firm can be generally different from the others.

Our results show that there is a positive and well-determined effect of innovation on the firm profit margin. The effect is larger for innovation patented 2–3 years before and then decreases smoothly as time passes by. There exists a difference between innovators and non-innovators. In the long run, we found that innovators and non-innovator converge (with different speeds) to different profitability steady states, being the innovators steady state higher than the non-innovator one. The effect of innovation on profitability seems to be permanent over time. Moreover, it is greater when the more persistent firms innovate. These results support the view that innovating firms have developed internal competencies and behavioral pattern that allow them to face the challenges of the market better than the non-innovating firms.

The article has the following structure: in Section 2, we describe data; in Sections 3, 4 and 5, we answer to the questions posed above, discussing both the methodology and the results. Section 6 concludes.

2 THE DATA

The analysis is performed on a sample of 267 UK manufacturing firms. The data set contains economic variables at the firm level taken from *DataStream* for quoted firms and from ICC for non-quoted firms. The economic variables at the industry level were taken from *Business*

Monitor: Report on the Census of Production: Summary Volume (London, 1988–1992). As a proxy of innovative activities carried out inside the firm, we use the number of patents requested by the individual firms to the European Patent Office in each year from 1978 to 1991.[2] If a firm has requested for at least one patent from 1978 to 1991, it is defined as an innovator, otherwise as a non-innovator.[3]

The sample is not random and was constructed as follows. We extract a random sample of 400 firms from the population of the non-innovating firms and a random sample of 400 firms from the population of the innovating firms for which DataStream and ICC supply the economic data. After cleaning the data, we remain with 168 innovating firms (62.9%) and 99 non-innovating firms (37.1%).

The data constitute a balanced panel of 5 years, from 1988 to 1992, as concerning the economic variables, while it is a balanced panel of 14 years for the patent data.

A preliminary descriptive analysis illustrates certain clear features of the data.[4]

Dividing the sample into five-dimensional classes according to the number of employees (small firms, from 1 to 99 employees; small-medium firms, from 100 to 199; medium firms, from 200 to 499; medium-large firms, from 500 to 999; and large firms, at least 1000 employees), large firms prevail (47.6%), followed by medium-large firms (12.7%) and medium firms (12%).

When we divide the sample in innovators and non-innovators, it is worth noting that non-innovators are more concentrated in the mechanical engineering (27.3%) and in the electrical and electronic sector (14.1%), whereas the innovators are more concentrated in the chemical sector (14.3%), in the other metal goods (10.1%), in the mechanical engineering (19%) and in the electrical and electronic sector (18.5%). These figures are not surprising, since it is commonly recognized that these sectors have the highest propensity to patent. Furthermore, when both size and innovation are considered, data show that innovating firms are generally larger than the non-innovating ones: 60.1% of innovators *vs.* 26.3% of non-innovators are concentrated in the large-firm class, whereas 40.4% of non-innovators *vs.* 7.7% of innovators are concentrated in the small-firm class. Therefore, the majority of firms in the sample, i.e., the survivors who maintained a fairly clear identity over the period, are large firms that carry out innovative activities inside the firm.[5]

The mean of the operating profit margin (OPM) decreases along time.[6] It is not surprising if we consider that the years 1990, 1991 and half of the 1992 were the years of economic recession at the world level. However, the mean of the OPM decreases more sharply for the innovator group (−40.72%) than for the non-innovator group (−13.02%). The same observation can be done for the size of the firm measured by the number of employees. For the pooled sample the

[2] The data of the European Patent Office data-bank on patent applications were kindly provided by CESPRI of the Bocconi University, Milan. We would like to thank Franco Malerba and Luigi Orsenigo, who allowed us to use these data.

[3] Geroski *et al*. (1993) used as a proxy of innovative activities the number of major innovations produced by each firm during the period 1945–1983. This information is taken from the SPRU Innovation data set that identifies over 4300 major innovation, defined as 'the successful commercial introduction of new or improved products, process or materials' introduced in Britain in 1945–1983. We think that this measure can be a little tautological in the sense that 'successful commercial' innovations are expected to have a positive effect on the firm profitability by definition. Patents, instead, are a more 'neutral' measure, since patents, on one hand, can vary in their economic value but, on the other, are always an indicator that innovative activities are carried out inside the firm.

[4] See the descriptive statistics of the data reported in Table I of the extended version of the article by the same authors in the Working Paper series (2004) of the T.C. Koopmans Institute, Utrecht School of Economics, Utrecht University (www.koopmansinstitute.uu.nl).

[5] Indeed, working with a balanced panel, we have a sample selection bias arising from the possibility that non-innovators are less likely to survive.

[6] See the descriptive statistics of the data reported in Table I of the extended version of the paper by the same authors in the Working Paper series (2004) of the T.C. Koopmans Institute, Utrecht School of Economics, Utrecht University (www.koopmansinstitute.uu.nl).

mean decreases along time, whereas for the non-innovator group the mean increases for the 1989 and 1990, and then decreases to reach approximately the same level as in 1988 (with a slight increase of 0.13% with respect to 1988). Indeed, the null hypothesis (that the two sub-samples, innovators and non-innovators, were taken from populations with the same mean when the variances are unknown and not equal) cannot be accepted for OPM, for size (SIZE) and for the variable indicating the average capital intensity (CAP), while it cannot be rejected for the other variables.[7] In other words, the means of the two groups are statistically different as long as we consider those variables that define the specific characteristics of the firm (profitability, size, market shares and capital intensity), whereas they are statistically indistinguishable for the variables defining the structure of the industry in which the firms operate (the concentration degree and the average scale of the firm in the industry). The average wage bill variables have not statistically different means either, suggesting that innovative firms do not pay higher salaries than the non-innovative ones.

The distributions of the number of patents requested each year by firms are very skewed with a large mass at zero patents. Indeed, 37.1% of firms never applied for a patent during the period 1978–1991 and 25.5% requested only one patent. Among those firms who applied for at least one patent during this period, on average, 76.15% of them did not request for a patent, 11.13% applied for one patent and 12.72% requested even as much as one patent per year. The mean of the number of patents requested each year grows with time. This is due to a legal-institutional event, which is the entry into functioning of the European Patent Office in 1978. For the first few years, firms were offered a new product/service by a quite new institution (Europe-wide patent protection obtained through a centralized procedure, less expensive than multiple national patents). As time went by, both the European patent system and the procedures to ask for a patent became more and more known among firms. In fact, the first 5 years of existence of the office show the highest growth rates of patent applications. This is due to a twofold circumstance. On the one side, more European and third country firms began to patent following this procedure. On the other side, the attractiveness of the European procedure has induced many firms to apply for more patents each year. Consequently, not only the mean of patent applications increases with time, but also its standard deviation.

Finally, when cross-correlations are computed among all variables, two features are worth noting.[8] On the one hand, we found (not surprisingly) high positive significant autocorrelation coefficients up to the fifth lag. This is especially true for variables like profits, patents, size and market shares. The strong inertia that characterizes the behavior of such variables is well-known and has been tested several times (as for persistence of patents and profits distributions see, Cefis, 2003a). On the other hand, data only show high (0.50–0.60) and significant correlation between the number of employees and the number of patents. This latter feature seems to support the Schumpeterian hypothesis that innovation increases with size.

3 ARE THERE DIFFERENCES IN PROFITABILITY BETWEEN INNOVATORS AND NON-INNOVATORS?

Our starting point is to study whether innovative firms are different in profitability from non-innovative ones. In fact, non-innovators could develop competencies and capabilities (other than those requested for innovating) that enable them to earn profits which are as high as those of innovators. The requirement to invest in sunk costs, which are independent of the

[7] Not reported and available on request.

[8] Cross-correlations are not reported and are available on request.

returns to innovation, and to fund product development and market diffusion activities (which some estimates place at 10–20 times initial R&D costs) may be such that there is an inevitable negative impact on retained profits. Therefore, non-innovating firms could perform as well as innovating ones. The next issue analyzes precisely whether differences between innovators and non-innovators in profitability exist, i.e., whether innovation is the source that generates different performances.

A simple way to do this is to compare the empirical distributions of the proxy for firm performance of the two groups. We proceed as follows:

1. Among different alternatives, we choose *OPMs* (operating profits – after depreciation and before taxes – divided by total sales) as a proxy for firm performance.
2. We construct the following variable: $opm_{it} = OPM_{it} - \left(\overline{OPM_{it}}\right)$, where OPM_{it} is the *OPM* of each firm at time t; and $\left(\overline{OPM_{it}}\right)$ is the mean of the *OPM* of the industry (at two digit) to which the firm belongs. Therefore, our new variable, opm_{it}, represents the deviation from the industry mean of the *OPM* of each firm. There are several reasons to implement this variable instead of the simple *OPM*: (i) this new variable controls the differentials in profitability among industries; (ii) it gets rid of more general common factors in the economy as business cycle and inflation; (iii) it has the advantage which can also be used to characterize distributions whenever the number of firm changes over time; and (iv) it provides an easy way of comparing distributions with different numbers of observations.
3. We test for stability along time of the empirical distributions of opm_{it} in the sample, applying a Kolmogorov–Smirnov test between the profit distribution in 1988 and each profit distribution of the following years.
4. In case of stability along time, we construct 'pooled' distributions (for non-innovators and innovators) using all the observations over the years of the sample for each group.
5. We test, whether the distributions of the two groups are equal, by means of Kolmogorov–Smirnov goodness-of-fit test.

The previous steps are then repeated also using a different clusterization. In the comparison between the two groups, instead of considering innovators, we take persistent innovators. The latter are defined as firms who do not remain for more than 2 years without asking for a patent.

3.1 Results

After the construction of the proxy for the firm's profitability in terms of deviations from the mean, we tested for the stability along time of the distributions of firm's profits. The distributions show stability along the years of the sample (the results of the test are shown in Table I(A)), allowing us to 'pool' together all the observations regardless to time. This procedure is useful because we want to compare the distributions of the three different groups of firms, and in order to achieve this we need distributions with a sufficient number of observations.

We then tested whether the profit distribution of non-innovating firms is equal to the distribution of the innovating firms. The test rejects the null hypothesis (see Table I(B)). Furthermore, the same null hypothesis was tested and rejected (not surprisingly) between non-innovators and persistent innovators, but it was also rejected between innovators and persistent innovators.

Indeed, the persistent innovators group has a higher mean, median and maximum value of the profit variable than the non-innovator group. The innovator group is between the two.

These results suggest that non-innovating firms generally perform worse than innovating ones and that among the innovators the more persistent ones are those who achieve higher profits. Therefore, innovations seem to be the source (or at least one of the major sources) of profit differentials.

TABLE I Kolmogorov–Smirnov Test.

	K–S statistics	*p-Value*
(A) *The stability of profit distributions*		
opm_{1988} *vs*. opm_{1989}	0.0568	0.7384
opm_{1988} *vs*. opm_{1990}	0.0871	0.2375
opm_{1988} *vs*. opm_{1991}	0.0682	0.5213
opm_{1988} *vs*. opm_{1992}	0.0758	0.3911
opm_{1989} *vs*. opm_{1990}	0.0644	0.5926
opm_{1990} *vs*. opm_{1991}	0.0606	0.6658
opm_{1991} *vs*. opm_{1992}	0.0341	0.9941
(B) *Differences in profit distributions*		
non-innovators *vs*. innovators	0.1753	0.0000
non-innovators *vs*. persistent innovators	0.1656	0.0000
persistent innovators *vs*. innovators	0.1763	0.0029

Note: H_0, the two sets of observations come from the same distribution.

4 WHAT ARE THE EFFECTS OF INNOVATIVE ACTIVITIES ON FIRM'S PROFITS?

In this section, we study the effects of innovation on firm's performance using a model specification, which is standard in the literature. The empirical analysis is based on the estimation of a dynamic panel data model. Our contribution at this step is the fact that we exploit the heterogeneity at the firm level allowing each parameter to be firm specific.

4.1 The Empirical Specification

Following the conventional structure-performance foundations given in the literature (see Mueller, 1990; Machin and Van Reenen, 1993; Geroski *et al*., 1993), we model the profit margins as a function of cost considerations at the firm level (capital intensity and average labour costs), firm specific characteristics (firm size and market shares) and market structure (industry concentration, interaction between shares and concentration, and scale of industry).

To capture the effects of innovative activities on profitability, we use the number of patent applications of each innovating firm in all the previous years and we allow this variable to have effects on profitability for up to 8 years. [For the goodness of patents as indicator of innovative activities, see Acs and Audretsch, 1989; and Griliches, 1990.]

In the literature, lagged profit margins are usually introduced. The motivation for using the lagged dependent variable, as an explanatory variable is that profits may have a positive feedback and, in this way, may influence the innovative activities (more profits imply more investment in R&D and therefore more innovations). The lagged profit variable can be interpreted as reflecting cash flow influences on innovations which feed through the innovative activities and affect the profitability within a year. Introducing lagged profits as an explanatory variable increases the goodness-of-fit of the regression and subtracts significance to the other variables. This is especially due to the high persistence of profits (Cefis, 2003b). In fact, their dynamics would be well explained by an autoregressive model (see Sec. 6). For this reason, we decided to use two different empirical specifications: one introducing the lagged dependent variable among the explanatory variables and the other excluding it. The former is estimated for the sake of comparison with related studies, whereas the latter allows focusing our attention on the effects of the other relevant variables.

Using i subscripts for firms and t subscripts for time, the model specification is given by:

$$\begin{aligned} OPM_{it} &= \alpha_i + \beta_{i1} OPM_{it-1} + \beta_{i2} CAP_{it} + \beta_{i3} AVL_{it} + \beta_{i4} SIZE_{it} + \beta_{i5} MKS_{it} \\ &\quad + \beta_{i6} CON_{it} + \beta_{i7}\,(MKS_{it} * CON_{it}) + \beta_{i8} SCA_{it} + \sum_{q=1}^{8} \gamma_{iq} PAT_{it-q} + \varepsilon_{it}, \end{aligned} \tag{1}$$

where OPM_{it} is the profit margin (operating profits before interest charges and tax divided by total sales); α_i is the firm specific fixed effect accounting for individual technological competencies; CAP_{it} is the inverse relative capital intensity (the inverse of total capital divided by total sales); AVL_{it} is the per capita wage bill (total remuneration divided by the number of employees); $SIZE_{it}$ is the total number of employees; MKS_{it} is the firm market share at the two-digit industry level (firm total sales divided by industry total sales); CON_{it} is the five firm concentration ratios at the three-digit industry level (total sales of the five largest firms divided by industry total sales); SCA_{it} is a scale measure at the two-digit industry level (industry total sales divided by the number of enterprises in the industry); OPM_{it-1} is the lagged OPM; PAT_{it-q} is the total number of patent applications of each firm for each year; γ_{iq} measures the short-run cumulative impact of innovation (patent demand) on profit margins, while the long-run impact of each innovation is given by $\sum_q \gamma_{iq}/(1-\beta_{i1})$; ε_{it} are the disturbances of the model which are assumed to be normally, identically and independently distributed.

4.2 The Estimation Method

One of the contributions of our work certainly relies on the econometric specification. Generally, the literature has focused on dynamic panel data models using a fixed or a random effects estimator at the firm level and common slopes. We allow for heterogeneity also in the slope coefficients. Hence, our specification is fairly more general than the previous ones. Two reasons justify this important assumption: first, we believe that even in very restricted industrial sectors (for example five-digit SIC classification) firms might not be homogeneous as the previous literature is constrained to assume; second, in a dynamic panel data framework with possible heterogeneity, the common slope assumption introduces a bias in the estimation results if heterogeneity is indeed present (Hsiao *et al.*, 1999).

To briefly describe the methodology, let us rewrite the model specification in a more compact form:

$$y_i = X_i\theta_i + \varepsilon_i, \tag{2}$$

where $y_i' = (y_{i1}, \dots, y_{iT})$ and $X_i = (X_{i1}, \dots, X_{iT})'$. X_{it} is the k-vector of independent variables (including the lagged dependent variable), with $X_{i1} = 1$ for all i and θ_i is the $k \times 1$ parameter vector, $\theta_i = (\alpha_i, \beta_{i1}, \dots, \beta_{i8}, \gamma_{i1}, \dots, \gamma_{i8})'$.

Since coefficients vary across firms, estimation of the parameters is impossible without imposing some restrictions. However, instead of constraining the parameter vector to be the same across units, we assume they are random and a prior distribution on θ_i is introduced. The following assumptions on the population structure are made:

$$\begin{aligned} &\theta_i \sim N(\bar{\theta}, \Sigma_\theta)\ \ \forall i, \quad \operatorname{cov}(\theta_i, \theta_j) = 0 \\ &\varepsilon_i \sim N(0, \sigma_i^2 I_T)\ \ \forall i, \quad E(\varepsilon_i \varepsilon_j') = 0\ \text{ if } i \neq j \end{aligned}$$

It is worth noting that the generality of the restriction on θ_i relies on the fact that the unit specific parameters come from a common distribution across firms. This assumption, called in the literature *exchangeability assumption*, cannot be relaxed in our particular case without incurring in estimation problems. As a matter of fact, without imposing an exchangeable prior on θ_i, the estimation would be impossible because the number of parameters for each individual equation exceeds the number of time point observations. Also, notice that, for the same reason, a non-informative prior on θ_i (i.e., $\Sigma_\theta^{-1} = 0$) cannot be imposed. In fact, with such a prior the posterior mean of θ_i would just be the OLS estimation, unit by unit, which is unfeasible for the reason mentioned above, namely, the lack of degrees of freedom.[9]

A full Bayesian analysis requires the specification of a prior for the hyperparameters. Assuming independence, as it is customary, we take:

$$p(\bar{\theta}, \Sigma_\theta^{-1}, \sigma_i^{-2}) = p(\bar{\theta})p(\Sigma_\theta^{-1})p(\sigma_i^{-2})$$

to have a normal-wishart-inverted gamma form, i.e.

$$\bar{\theta} \sim N(\mu, C), \quad \Sigma_\theta^{-1} \sim W((mS)^{-1}, m), \quad \sigma_i^2 \sim IG\left(\frac{v_o}{2}, \frac{v_o \tau_o^2}{2}\right)$$

A few remarks are needed. First of all, for finite T we must be more specific about the initial conditions y_{i0}. We assume that the initial conditions are fixed and derive the results conditional on them. Hsiao *et al.* (1999) showed that even with a sample size of $T = 5$ (as it is the case here), the small sample bias is fairly reduced using the Bayesian approach compared with other estimation methods. A more proper treatment of the initial conditions would require some extensions on the line of Sims (1998) that will be reported in our future work.

Secondly, to obtain the joint posterior density of our parameters we have to combine the prior information previously described with the (conditional) likelihood. The problem is that in order to get the marginal posterior of each of the parameter vectors, we must integrate out the others and the integration is analytically intractable. Hence, we rely on a numerical integration using the Gibbs sampling algorithm as described in Gelfand *et al.* (1990) and Hsiao *et al.* (1999), among others. The Gibbs sampler is a recursive simulation method requiring only knowledge of the conditional posterior distribution of the parameters of interest on MCMC methods. Given the conditional posterior distributions of the parameters of interest, the Gibbs sampler produces an approximation to the joint posterior density. Convergence of the Gibbs sampler to the true invariant distribution in our case is subject to standard, mild conditions since model (2) is a simple panel data model with uncorrelated and heteroskedastic errors (e.g., Hsiao *et al.*, 1999). Marginal posterior densities are then obtained by integrating out of these joint posteriors numerically within the Gibbs sampler. The relevant conditional distributions that are needed to implement the Gibbs sampler in our case are easily obtained from the joint posterior density.[10]

In general, inference on any continuous function of the parameters of interest, $G(\psi)$, can be constructed using the output of the Gibbs sampler and the ergodic theorem.

For example,

$$E(G(\psi)) = \int G(\psi)p(\psi|Y)\,\mathrm{d}\,\psi$$

can be approximated using:

$$\frac{1}{\bar{L}}\left[\sum_{\ell=L+1}^{L+\bar{L}} G(\psi^\ell)^{-1}\right]^{-1}$$

[9] See Appendix B of the extended version of the paper by the same authors.

[10] These distributions are reported in Appendix A of the extended version of the paper by the same authors in the Working Paper-series (2004) of the T.C. Koopmans Institute, USE, Utrecht University (www.koopmansinstitute.uu.nl).

where ψ^{ℓ} is the ℓth draws of vector $\psi = (\theta_i, \bar{\theta}, \Sigma_{\theta}^{-1}, \sigma_i^2)$, $(L + \bar{L}\ )$ is the total number of iterations in the Gibbs sampler and L is the number of discarded iterations.

In this article, the ergodic mean of the marginal posterior distributions for the parameters of interest are used to discuss the estimates of these parameters.

4.3 The Results

We estimated four different empirical specifications following the Bayesian approach previously described as well as three 'classical' estimation methods, reported for the sake of comparison with the previous literature. Results are presented in Table II. Model I is the Eq. (1) presented in Section 3. Model II is the same empirical specification, but with the cumulated variable, $PATCUM_{it}$, instead of the innovation variables disaggregated by years. $PATCUM_{it}$ is the sum of all the previous patents requested by firm i up to the time t.[11] If the disaggregated number of patents can be interpreted as a flow variable, the cumulated number of patents can be interpreted as a stock variable; it mainly captures the stock of knowledge and competencies that the firm has cumulated up to the present time. Model III is the empirical specification described in Eq. (1) excluding the lagged profit margins and with the disaggregated innovation variables. Finally, Model IV is the same as Model III but with the cumulated innovation variable instead of the disaggregated ones.

For each model specification we compare the Bayesian estimates with those obtained using traditional approaches, namely, random effect (RE), fixed effect (FE) and OLS. The Bayesian estimates reported in the table represent the means of the posterior distributions of each parameter. Notice that, the classical estimates are realization of random variables, whereas the Bayesian estimates are pure number. Therefore, there are not standard errors attached to these estimates. In the table, a star (*) denotes classical significance for estimates obtained with RE, FE and OLS, while it denotes the fact that the value zero is not included in the 50% central part of the posterior distribution for the Bayesian estimates. For the parameters of interest in our analysis, i.e., $\gamma_1, \ldots, \gamma_8$, we also report the entire empirical posterior distribution in form of a box plot, where the box contains the 50% central part. The central line represents the median. Though it is not entirely exact, we assess the significance of the Bayesian estimates by checking, whether the box contains the zero or not.

The estimation deserves several comments.

From a statistical side, notice that the estimation methods provide sometimes very different estimates of some crucial parameters, which would lead the analyst to derive different conclusions from their economic interpretation. While the classical approaches are very similar among them, except perhaps the FE, the Bayesian estimates sometimes differ importantly from the others. We believe that this is mostly due to the different treatment of the heterogeneity in the sample. For instance, even though the lagged dependent variable shows a positive well-determined effect on current profitability (β_1) regardless of the estimation approaches (Models I and II), the estimate of β_1 is considerably smaller in the Bayesian than in the traditional estimation methods. This result is not surprising since the latter, ignoring (possible) slope heterogeneity, overestimates the parameter of the lagged dependent variable (Hsiao *et al.*, 1999). On the other hand, the economic interpretation of this result does not contradict the findings of several studies on profitability at the firm level (e.g., Mueller, 1990; Cefis, 2003b), which have shown that profits display a strong persistence along time, that decreases only if we consider a very long time-lag (10 years). The higher inertia is due to the fact that these studies focus on time series approach mainly based on the analysis of simple AR(1) models without considering other explanatory variables, as we also do in Section 5.

[11] Its coefficient in Table II is denoted by γ.

TABLE II Estimation of the General Model. Bayes, Random Effect, Fixed Effect and OLS Estimates.

	Model I				*Model II*				*Model III*				*Model IV*			
	Bayes	*RE*	*FE*	*OLS*	*Bayes*	*RE*	*FE*	*OLS*	*Bayes*	*RE*	*FE*	*OLS*	*Bayes*	*RE*	*FE*	*OLS*
α	4.7798*	3.1928*		2.4793*	3.1647*	2.8899*		2.4156*	11.3056*	9.9679*		9.6485*	9.1516*	9.8721*		9.6068*
β_1 (OPM)	0.2276*	0.5896*	0.6899*	0.6957*	0.1853*	0.6272*	0.6921*	0.6975*								
β_2 (CAP)	0.0802	−0.0006	−0.0040	−0.0029	0.0999	−0.0018	−0.0039	−0.0031	0.0765	0.0256*	−0.01812*	−0.0155*	0.0792*	0.0192*	0.0182*	−0.01578*
β_3 (AVL)	−0.8362*	−0.0241	0.0265	−0.0167	−0.7300*	−0.0158	−0.0316	−0.01177	−0.7210*	−0.3068*	0.03025	−0.0955*	−0.6808*	−0.2791*	0.04389	−0.0898*
β_4 (SIZE)	−0.00383	0.000014*	0.00002*	0.000011*	0.00067*	0.000012*	0.00003*	0.000001*	0.00094*	0.00002*	0.00002*	0.00002*	0.00016*	0.00002*	0.00004*	0.00002*
β_5 (MKS)	0.6544	0.6234	−0.2570	−0.3243	0.7947	0.7607	−0.4517	0.0749	4.8686*	4.4705	6.5086	5.9662*	4.7470*	5.5790*	5.60663	6.5206*
β_6 (CON)	1.4394*	−3.7497*	−2.9543*	−3.1701*	4.4033*	−3.5226*	−2.9564*	−3.1428*	0.8046*	−2.6207*	−5.0597*	−4.4275*	2.1386*	−3.0673	−5.1316*	−4.4337*
β_7 (MKS$*$CON)	−0.7523	−0.9266	−6.9219	0.3508	−0.7462	−1.1275	−7.5548	−0.2044	−5.3556*	−5.7276	−14.770*	−8.9302*	−8.2940*	−7.7553*	−15.394*	−9.7602*
β_8 (SCA)	0.00431*	0.0001*	0.00001*	0.0001*	0.00303*	0.0001*	0.0001*	0.0001*	0.000412	6.52E−05	0.00021*	0.00019*	0.00234	0.00011	0.00021*	0.00018*
γ (PATCUM)					0.01091	0.00209	0.00072	0.00205					0.0287	0.0027	0.0059	0.0058
γ_1 (PAT$_{t-1}$)	0.0282	−0.0217	−0.0414	−0.0253					0.4087*	−0.0385	−0.1226	−0.0870				
γ_2 (PAT$_{t-2}$)	0.2046	−0.0152	0.0764	0.0147					0.5918*	0.0818	0.1023	0.0102				
γ_3 (PAT$_{t-3}$)	0.2861	−0.0826	−0.0565	−0.0780					0.1131	−0.0220	0.1191	0.1130				
γ_4 (PAT$_{t-4}$)	−0.0582	0.0994	0.1082	0.0969					−0.0845	0.0284	0.1283	0.0871				
γ_5 (PAT$_{t-5}$)	0.2561	−0.0218	−0.0415	−0.0220					−0.0968	−0.0744	−0.1045	−0.0803				
γ_6 (PAT$_{t-6}$)	0.086	0.0932	0.0634	0.0899					0.1293	0.0463	0.0770	0.0967				
γ_7 (PAT$_{t-7}$)	−0.1691	−0.0378	−0.0356	−0.0369					−0.0933	−0.0218	0.0367	−0.0011				
γ_8 (PAT$_{t-8}$)	−0.2274	0.0186	0.0290	0.0156					0.0505	0.0709	−0.0335	−0.1274				
σ	0.7594	6.2095	6.4822	6.4972	0.7586	6.2095	6.4635	6.4810	1.3812	5.6603	8.6671	8.7528	1.2816	5.8266	8.6689	8.7464
Cumulative	0.4063	0.0625	0.1020	0.0549					1.0188	0.0707	0.2029	0.0113				
Longrun	0.5260	0.1523	0.3291	0.1805	0.013469	0.0056	0.0023	0.0066								
ΔP	37.96%	1.20%	3.57%	1.07%	0.07%	0.04%	0.02%	0.04%	8.29%	0.97%	2.39%	0.16%	0.19%	0.04%	0.07%	0.08%

Notes: (1) For the classical estimates, the star (*) means that the variable is significative at least at the 0.15 level. For the Bayes estimates, the star (*) means that the zero is not included in the 50% central part of the distribution. (2). Fixed and random effect estimates are obtained using the standard routine PREGRESS in RATS v. 5.1.

Also, notice that the standard error of the estimate (σ) is much lower in the Bayesian estimation than in the classical approaches. This has a simple explanation, being just the result of assuming in the Bayesian method that all the parameters are random variables, so that most of the variability of the error term is just ‘shifted’ onto them. Further, all estimation methods provide higher and more significative estimates when the autoregressive parameter is dropped from the analysis (Models III and IV). Finally, the constant in the FE estimation is not identified because it has been removed by subtracting the time means.[12]

From an economic side and given the important statistical consideration above, it is worth mentioning the following. The firm cost characteristics as defined by capital intensity (β_2) and average labour costs (β_3) appear to have very little, negative and not statistically relevant impact on firm profitability in the classical estimation, regardless of the empirical specification. The Bayes estimation, instead, shows a positive, albeit not significative, effect of capital intensity, but always a relevant negative impact of labor cost on profitability.

Regarding the firm specific characteristics, note that size as proxied by the number of employees (β_4), has almost always a positive and significant impact on firm profit as expected. The sign of the impact of the market share (β_5) and the concentration ratio (β_6), instead, is different in the four estimation methods. In the one we prefer (Bayesian), they have always a positive well-determined effect. There is less ambiguity across methods regarding the product between these two variables (β_7), which always display a negative (in some case significative) effect on profit margins. Finally, the scale measure (β_8), the last variable that characterizes the market structure, has a positive effect through all the specifications as well as the estimation methods considered. Broadly speaking, these results confirm the conventional wisdom that the firm’s profit margins are higher when its industry is highly concentrated and when its market share is considerable. Besides being quite robust, regardless of the specification used, the results are almost in line with the previous findings (e.g., Geroski *et al*., 1993).

Let us consider now the effects of innovation on profitability (γ_1 through γ_8, or γ). Here is where remarkable differences can be observed across the various estimation methods. The classical estimates have an odd behavior and also a very little cumulative impact (cumulative) overall. The single coefficients show alternating signs over the recent past of the firms and seem not to follow a clear pattern. These results are in line with those found by Geroski *et al*. (1993).

The Bayesian estimates reveal a reasonable pattern of the impact of innovations on profits as well as a greater cumulative and long-run impact. Note that the effect of having requested at least one patent in the previous years seems to increase up to the second/third lag and then smoothly decreases afterwards. To put this number into their right perspective, Figures 1 and 2 plot the entire empirical posterior distributions over the eight lags. Then, the number reported in the table is just the line inside the box of these figures. Certainly, the posterior distributions give a better picture both of the dynamics and of the relevance of being an innovator in this sample. Figure 1 refers to Model I and Figure 2 refers to Model III. Note that besides providing almost the same qualitative conclusions, they differ in shape according to the inclusion or exclusion of the lagged profit variable. Without the lagged dependent variable, the distributions obtained are more symmetric, have a lower variance and a higher average impact than those obtained in the model specification that includes an autoregressive term. The cumulated impact is greater as a consequence of not considering the feedback effects of lagged profits (see Sec. 3). However, they do show the same tendency in behavior along time, firstly moving up (until the second/third lag of the innovation variable) and then slowly moving down. We consider both this pattern and the cumulative impact implied as strong evidence in favor of the existence of a clear effect

[12] Both the fixed and the random effect estimates have been obtained using the standard routine PREGRESS in RATS (v 5.1) by allowing for time effect in the error term.

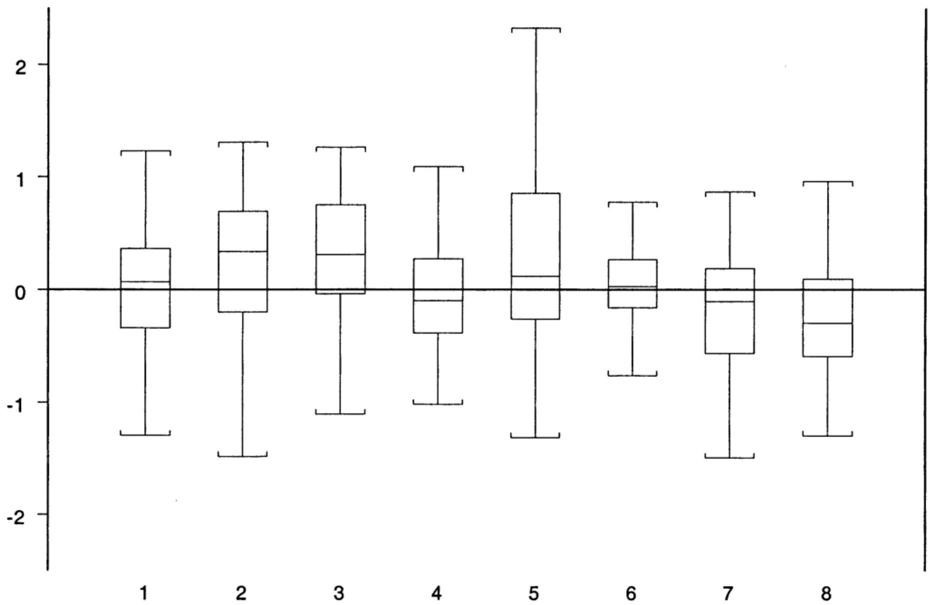

FIGURE 1 The effect of innovation over the years (Model I).

of innovation on profitability. This effect can be computed over the long run ('long run' in Table II) as $\sum_q \gamma_{iq}/(1 - \beta_{i8})$. As a consequence of the different estimates, the Bayesian, as opposed to the classical methods, provides a greater impact over the infinite future.

Finally, to see these numbers in their right perspective, the last row of Table II ($\Delta\mathbf{P}$) reports the result of a simple experiment, which computes the relative difference in the implied profit

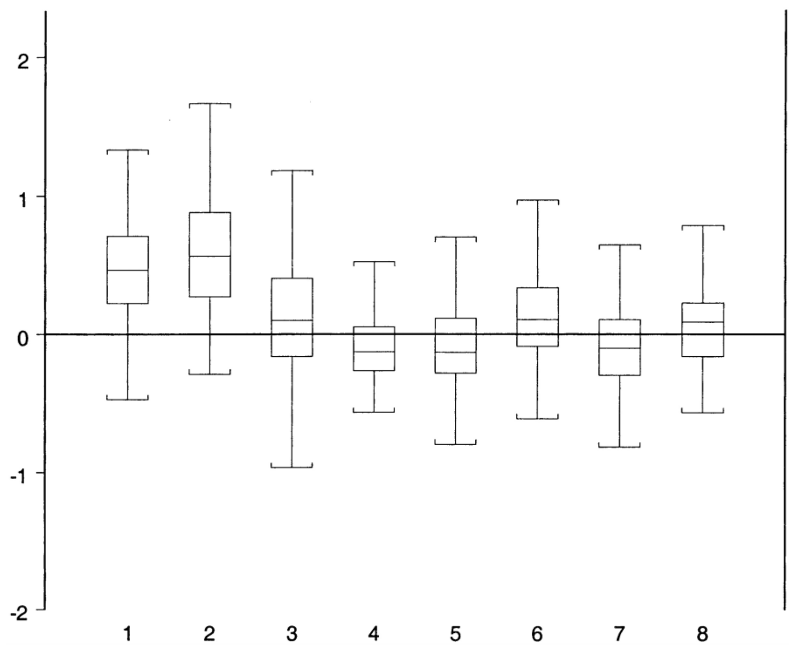

FIGURE 2 The effect of innovation over the years (Model III).

margins between the average firms which does and the average firms which does not innovate over the sample. Given the characteristics of the firm as implied in the estimated coefficient of the empirical mode, it is shown for instance, that the request of patents in the previous years rises profit margins by 37.96% relative to the mean in Model I. The same impact on average profits of innovations computed in RE, FE and OLS raises the profits only by 1.2%, 3.57% and 1.07%, respectively.

These results suggest that the latter methodologies dramatically underestimate the effect of innovation on profitability both in the long and in the short-run.[13]

5 ARE THE DIFFERENCES BETWEEN INNOVATORS AND NON-INNOVATORS TRANSITORY OR PERMANENT?

From the above sections, we conclude that innovations have a positive effect on profits after controlling for relevant characteristics at both sector and firm level. Unfortunately, the previous analysis does not allow to conclude that while innovating helps profitability of the firm, it also makes the firm permanently more profitable. In order to answer the last question, therefore, we need to change slightly the statistical framework and perform a long-run analysis by taking a more general parametric structure for the profit margins. Provided that several empirical studies have outlined the strong persistence in corporate profit margins (Mueller, 1990; Cubbin and Geroski, 1987; Cefis, 2003b), we model profits time series as an autoregressive process of order one (AR(1)):

$$OPM_{it} = \alpha_i + \rho_i OPM_{it-1} + \eta_{it} \tag{3}$$

The difference, with Eq. (1), is that here we neglect possible relevant firm or sector characteristics and relegate them to the error term. In a sense, all those characteristics are in $\alpha_i + \eta_{it}$. Heterogeneity in ρ_i is also needed to account for this feature.

This statistical model is flexible enough to formally test for persistence of differences between innovators and non-innovators. Given that we assume as heterogeneous the structural parameters, α_i and ρ_i, without loss of generality the residual η_{it} is assumed to have mean zero and to be independent across units and over time, namely, $\eta_{it} \sim N(0, \sigma_i^2)$. Also, notice that not by forcing the parameters to be the same across units, we can separately examine the issue of convergence to the steady states from questions regarding the limiting distributions of profit margins. It is well known that the steady state value of OPM_{it} is $S_i = \alpha_i/(1 - \rho_i)$, while $1 - \rho_i$ is the rate at which each unit converges to its own steady state. Our strategy has two steps. We first estimate in Eq. (3), and then we test whether the implied average steady state of innovators is different from the implied average steady state of non-innovators. As before we estimated Eq. (3) with different methods – pooled OLS, random effect, mean group[14] and the Bayesian estimates. A part from the pooled OLS, which we already know to be severely biased, the other methods provide qualitatively similar results, therefore we report only the Bayesian estimates. We preferred so also for other reasons.[15] First, it has been shown that with a dynamic heterogeneous panel data specification with a very short time dimension, the Bayes estimation has much lower bias as compared with the classical approaches and,

[13] The FE estimator seems to do better than RE and OLS, considering both the cumulative and the long run impact of innovation. Also the shape of the single coefficients in Model III is pretty much similar to an inverse U as the Bayesian estimate. Notwithstanding, its implied impact of innovations on profitability is still far lower.

[14] See Appendix B of the extended version of the paper by the same authors in the Working Paper-series (2004) of the T.C. Koopmans Institute, USE, Utrecht University (www.koopmansinstitute.uu.nl).

[15] Results based on the other methods are available on request.

therefore, is quantitatively more reliable (e.g., Hsiao *et al*., 1999). Second, after controlling for heterogeneity, the Bayes estimation provides an average estimation of the relevant parameters and of their posterior distributions as opposed to the classical estimation methods, which only provide point estimates. We will see in the discussion of the results, how important this feature becomes. Third, with the kind of generality implied by the assumed heterogeneity, there are too many parameters relative to the number of time series observations for each cross-sectional unit; a prior on the parameters in the form of a distribution to be combined with information contained in the data (likelihood) to obtain posterior estimates is fairly more general than the kind of restrictions imposed in the classical estimation. Finally, the procedure solves the small sample problem encountered by estimating separately using only the observations on unit i, since the Bayesian estimates are exact regardless of the sample size, and, at the same time, it does not require the stringent assumption of equality of the coefficients across units.[16]

Since, we are interested in testing for persistence of differences between innovators and non-innovators, our prior assumption is that the rate of convergence and the intercept of the model do not differ too much across units of the same group. More precisely, this means that we do not assume symmetry across all the firms, but we allow for a partition in two groups. In this way, each parameter vector $\theta_i = (\alpha_i, \rho_i)$ can be viewed as arising from a superpopulation (say G), which is a mixture of two populations (say G_1 and G_2), in some proportion π and $1 - \pi$, respectively, where π represents the proportion of innovators or persistent innovators. The prior distribution for the parameter vector satisfies:

$$\theta_i \sim N(\bar{\theta}^p, \Sigma_\theta^p) \quad p = 1, 2 \quad \text{cov}(\theta_i, \theta_j) = 0$$

Notice that the mean $\bar{\theta}^p = (\bar{\alpha}^p, \bar{\rho}^p)$ is the same in each group p. This is again the exchangeability assumption already encountered in Section 4, only modified to explicitly account for a possible prior difference between the two groups. How stringent is this prior clearly depends on the variance Σ_θ^p, which is generally taken as diffuse as possible.[17] Anyway, the results we obtain are robust to a change in the prior assumptions.

We further assume, as before, that for each group:

$$\bar{\theta} \sim N(\mu, C), \quad \Sigma_\theta^{-1} \sim W((mS)^{-1}, m)$$

and also that:

$$\sigma_i^2 \sim IG\left(\frac{v_o}{2}, \frac{v_o \tau_o^2}{2}\right)$$

With this specification, we can easily test whether the posterior distributions of the average parameters for the two groups are the same ($\bar{\theta}^1 = \bar{\theta}^2$). Also, we can test the null hypothesis $\bar{S}^1 = \bar{S}^2$ *vs.* the composite alternative hypothesis that they are different. Here, $\bar{S}^p$ is the posterior distribution of the average steady states. Clearly, $\bar{S}^p \neq \bar{\alpha}^p/(1 - \bar{\rho}^p)$, because of the non-linearity of the steady state, so we obtain it by computing the ergodic mean of the posterior distribution of the linearized S_i. Given that we dispose the empirical distributions of both $\bar{\theta}$ and $\bar{S}$ for the two groups, the null hypothesies can be easily tested using standard procedures to compare distributions (e.g., t-statistics on the mean, or Kolmogorov–Smirnov on the entire cdf).

Testing the null $\bar{S}^1 = \bar{S}^2$ is equivalent to test the unconditional difference in the long-run performance of the two groups. Whether it is rejected or not, it would also be interesting to

[16] See Canova and Marcet (1998) and Hsiao *et al*. (1999) for a more detailed discussion on this points.

[17] See Appendix B of the extended version of the paper by the same authors in the Working Paper-series (2004) of the T.C. Koopmans Institute, USE, Utrecht University (www.koopmansinstitute.uu.nl).

know what variables explain the cross-sectional dispersion of estimated steady states across unit S_i. For this reason, we run simple cross-sectional regressions of the type:

$$S_i = b_1 + b_2 INN_i + b_3 OPM_{i0} + X_i' b_4 + u_i \tag{4}$$

where X_i is a vector including some of the variables previously described, proxing for differences in market structure, size, capital intensity, etc.; *INN* is a dummy which assumes value 1 for innovators (or persistent innovators); OPM_{i0} is the profit margin initial condition for each unit. Notice that the usefulness of this cross-sectional specification is twofold. On the one hand, after conditioning on the variables contained in X_i the positiveness of the estimate of b_2 is crucial in determining whether innovating firms have long-run profit margins higher than non-innovating firms; on the other hand, the sign and the magnitude of the parameter b_3 is important to argue about the persistence of this difference.

5.1 Results

The estimation results of the AR(1) model for profit margins are presented in Table III.[18] Figure 3 reports the posterior distributions of the average steady states for the groups considered. Table IV reports the estimation of Eq. (4). Several comments are in the order.

The autoregressive parameter, ρ_i, is higher for innovators than for non-innovators, implying a stronger persistence in profits for innovators and a convergence speed to the steady state higher for non-innovators. Persistent innovators show even higher median values as compared to both groups, but a greater variance given the reduced number of firms in this group. The difference in the median values among the three groups is statistically relevant in that as shown in the table the central part of their distributions does not even overlap. A t-test on the mean of their posterior distributions rejects at 1% confidence level the null hypothesis of equal means.

Figure 3 shows the distributions of the implied steady states for the three groups. As expected, the posterior mean of the distribution for persistent innovators is greater than the posterior mean of the distribution for innovators that in turn is greater than the one for non-innovators. These differences are statistically relevant, in the sense that both a t-test and a Kolmogorov–Smirnov test reject the null hypotheses of equality of the means and of the entire distributions, respectively.[19]

TABLE III Bayesian Estimates of the AR(1) for Profit Margins.

	All		*Non-innovators*		*Innovators*		*Persistent innovators*	
	ρ	*ss*	ρ	*ss*	ρ	*ss*	ρ	*ss*
1st quart	*0.38*	*4.33*	*0.27*	*3.99*	*0.69*	*5.01*	*0.73*	*6.42*
Median	0.70	5.59	0.33	4.71	0.72	6.03	0.83	10.49
3d quart	*0.79*	*7.31*	*0.38*	*5.40*	*0.76*	*7.09*	*0.94*	*14.19*

Note: ρ, the estimate of the autoregressive parameter; *ss*, the estimate of the study state. The first quartile, the median, and the third quartile are reported for each distribution. Posterior distribution of the relevant parameters.

[18] The first column (all) reports the results obtained by imposing an exchangeability assumption without differencing between innovators and non-innovators, i.e., $\theta_i \sim N(\bar{\theta}, \Sigma_\theta)$, where $\bar{\theta}$ is common, *a priori*, to all firms.

[19] Test results are available on request.

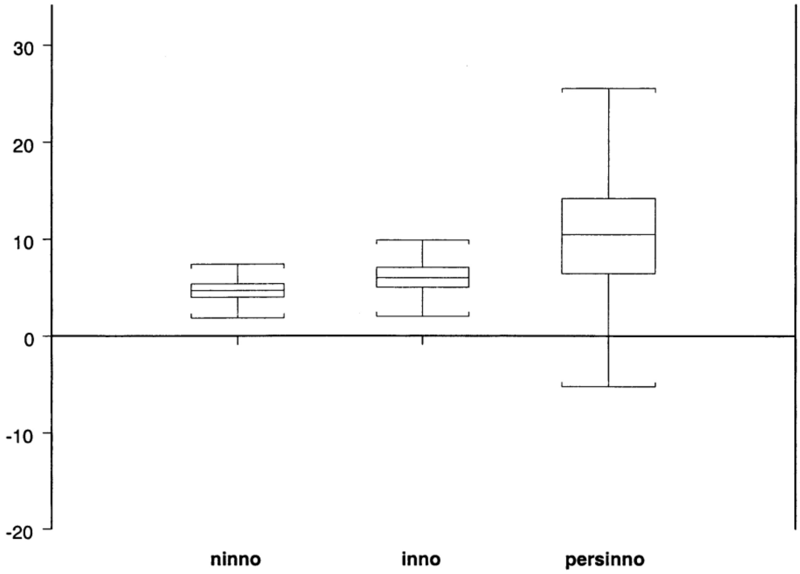

FIGURE 3 Posterior distribution of steady states.

TABLE IV Explaining the Steady State.

	A	*B*	*C*	*D*	*E*	*F*	*G*
Constant	**4.69***	**5.32***	**4.69***	**4.69***	**3.77***	**4.06***	**7.04***
	0.50	0.32	0.50	0.50	0.85	0.83	1.34
Innov	**1.457****		**1.04**	**1.04**	**0.79**		
	0.66		0.66	0.66	0.67		
Persist		**4.81***	**4.40***	**5.04***		**4.07***	**3.36****
		1.68	1.70	1.91		1.59	1.87
PATCUM				**−0.02**			
				0.016			
OPM_{88}					**0.16****	**0.15****	**0.13****
					0.09	0.09	0.08
SIZE							**1.4E − 05**
							0.000014
MKS							**−3.44**
							3.46
CONC							**−3.31****
							1.74
SCA							**3.2E − 05**
							0.000069
CAP							**−0.027***
							0.0080
AVL							**−0.019**
							0.09
R^2	0.020	0.050	0.050	0.060	0.080	0.110	0.170

Note: (*), (**), and (***) means significance at 0.01, 0.05 and 0.10 level, respectively. Estimates in bold, standard errors below.

The intuitive meaning of this is that when the average firm innovates in a persistent way, its profits are more persistent than those of a non-innovator and their long-run profits are also higher. In other words, these results suggest not only that innovations have a positive effect on profits, but also that this effect is permanent in the sense that innovating firms converge to higher steady states than non-innovators. Innovating firms, therefore, must develop competencies and capabilities that allow them to earn higher profits and to keep on maintaining the differential in profits as time goes by. The findings are then stronger the more persistently firms innovate. Notice, however, that the three distributions overlap for certain values and that the posterior distribution of the steady state of the average persistent innovator is far more spread than the distributions of the other two groups and it contains even negative values in the left tail. Overall, this means that though *a posteriori* it is more likely that innovators have higher steady states it cannot be discarded the possibility for non-innovators to be more profitable than some persistent innovators. This is not a contradiction and has three complementary explanations. First, the sample is very short; second, the number of persistent innovators in the sample is very limited (9% of total firms); third, given the previous two points, it is likely to find in the sample persistent innovators who may experiment negative profit values, perhaps just because of their innovations.[20] Notwithstanding, in the long run the overall picture is that persistent innovators overperform the other firms. These qualitative features can only be observed thanks to the estimation procedure. In a classical analysis, we would have ended up with just one point of these distributions (one for each group) and the comparison could not have been so precise given the small number of sample data points. As an example, the random effect estimation of Eq. (3) with individual effects in the error term implies a value of 7.79 for the steady state of the average persistent innovator and a value of 6.032 for the steady state of the rest. That is, this classical procedure would lead to the same overall conclusion than the Bayes estimation, but we would not have the possibility of qualifying the results as we have done by looking at the entire posterior distribution. On the other hand, in the case of pooled OLS (no heterogeneity considered whatsoever) the implied steady state estimate for persistent innovator would have been of 2.74 as opposed to a value of 4.89 for the rest of firms. In this case, we would have concluded that persistent innovators experiment much lower long-run profits than the non-innovators. Notice, however, that though difficult to justify on an economic ground, the result is not totally incompatible with the Bayes figures, at least in the sense that the classical point estimates in this case represent values that would be possible in the Bayes estimation, but which are just assigned lower probability mass *a posteriori*, as shown in Figure 3.

The structural differences we have found between persistent innovators and the other two groups suggest that there is a strong correlation between the persistence in innovating and the persistence of the positive differential in profitability. Now, the question is: what variables explain the cross-sectional dispersion of estimated steady states across units? Table IV provides an answer to this last question. First, it is easy to notice that the steady states of profit margins are associated with innovative activity. As a matter of fact, the coefficient of the dummy *INNOV* (1, if the firm has requested at least one patent in the sample; 0 otherwise) is positive and significant. Also, columns A, B and C show that being an innovator (defined as $INNOV = 1$) and especially being a persistent innovator ($PERSIST = 1$) explains (or is associated with) the long-run performance. Notice, however, that being a persistent innovator seems to matter more than just innovating sometimes over the sample, because when both dummies are accounted for, only the dummy for persistent innovators has a significant effect. Further, column D tells us that being an innovator matters more than the number of patents requested, because the coefficient of *PAT* (the average number of innovations per firm over the sample) is not statistically different

[20] For instance, this is the case in the sample considered for LAPORTE INDUSTRIES LTD, J.C. BAMFORD EXCAVATORS and MASSEY FERGUSON MANUFACTURING LTD.

from zero. Moreover, columns E and F lead to the conclusion that after controlling for initial conditions (*OPM_88*, the OPM in 1988) the dummy *INNOV* is not statistically relevant any longer, while it remains significant to be a persistent innovator. This evidence may have a simple interpretation: initial conditions matter more than being an (occasional) innovator, but less than being a persistent innovator. Hence, independent of initial conditions, a persistent innovator has on average higher long-run profit margins. Incidentally, note that initial condition does not matter for the non-innovators and that the innovators on average have higher initial conditions than the non-innovators. Finally, the other exogenous variables seem not to add much explanatory power, with the exception perhaps of concentration ratio (*CONC*) and capital intensity (*CAP*).

6 CONCLUSIONS

We have investigated empirically some issue concerning the impact of innovative activities on the firm performance applying a Bayesian approach that allows us to take fully into account the heterogeneity across firms. According to our results, the effects of innovation on firm profitability are positive and relevant. The average firm that applies for patents in the previous years has profit margins higher by 37.96% relative to the mean. The effect is larger for innovation patented 2–3 years before and then decreases smoothly. Results are not robust across different estimation methods. Our findings seem to suggest that classical methodologies underestimate the effect of innovation on profitability both in the long and in the short-run.

We also find that there exists a clear difference in profitability between innovators and non-innovators, and that this difference is sharper when we consider persistent innovators defined as firms who do not remain for more than 2 years without asking for a patent.

Finally, our results suggest that innovators and non-innovators converge to different profit margins steady states (with different speed), confirming that the distance between innovators and non-innovators is a permanent one and persists over time, probably due to the acquisition of distinct competencies and capabilities. Independent of initial conditions, a persistent innovator has on average higher long-run profit margins.

Acknowledgements

We wish to thank two anonymous referees, Giovanni Bruno, Giovanni Dosi, Luigi Orsenigo, the participants to the seminars at the University of Alicante, of Bergamo and of Strasbourg and to the ECIS Conference 2001, Eindhoven, for useful comments and suggestions. The financial supports of the University of Bergamo (grant ex 60% No. 60CEFI02, Department of Economics), Ministerio de Ciencia y Tecnología and FEDER (Project BEC2002-03097) and Generalitat Valenciana (Project CTIDIB/2002/175) are gratefully acknowledged. The views expressed in this article are exclusively those of the authors and not those of the European Central Bank.

References

Acs, Z.J. and Audretsch, D.B. (1989) Patents as a Measure of Innovative Activity. *Kyklos*, **42**(2), 171–180.

Aghion, P. and Howitt, P. (1992) A Model of Growth Through Creative Destruction. *Econometrica*, **60**(2), 323–351.

Canova, F. and Marcet, A. (1995) The Poor Stay Poor: Non-Convergence Across Countries and Regions. *CEPR Discussion Papers*, **1265**.

Cefis, E. (2003a) Is There any Persistence in Innovative Activities? *International Journal of Industrial Organization*, **21**(4), 489–515.

Cefis, E. (2003b) Persistence in Innovation and Profitability. *Rivista Internazionale di Scienze Sociali*, **110**(1), 19–37.
Cubbin, J. and Geroski, P. (1987) The Convergence of Profits in the Long Run: Inter-firm and Inter-industry Comparisons. *Journal of Industrial Economics*, **35**(4), 427–442.
Dosi, G., Marsili, O., Orenigo, L. and Salvatore, R. (1995) Learning, Market Selection and the Evolution of Market Structure. *Small Business Economics*, **7**(6), 411–436.
Gelfand, A.E., Hills, S.E., Racine-Poon, A. and Smith, A.F.M. (1990) Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *Journal of the American Statistical Association*, **85**(412), 972–985.
Gorecki, (1986) The Importance of Being First: The Case of Prescription Drugs in Canada. *International Journal of Industrial Organization*, **4**(4), 371–395.
Geroski, P., Machin, S. and Van Reenen, J. (1993) The Profitability of Innovating Firms. *RAND Journal of Economics*, **24**(2), 198–211.
Geroski, P., Van Reenen, J. and Walters, C.F. (1997) How Persistently do Firms Innovate? *Research Policy*, **26**(1), 33–48.
Griliches, Z. (1990) Patent Statistics as Economic Indicators: A Survey. *Journal of Economic Literature*, **28**, 1661–1707.
Hsiao, C., Pesaran, M.H. and Tahmiscioglu, A.K. (1999) Bayes Estimation of Short Run Coefficients in Dynamic Panel Data Models. In Hsiao et al. (eds) *Analysis of Panels and Limited Dependent Variable Models: In Honor of G.S. Maddala*. Cambridge: Cambridge University Press.
Klepper, S. (1997) Industry Life Cycles. *Industry and Corporate Change*, **6**(1), 145–181.
Machin, S. and Van Reenen, J. (1993) Profit Margins and the Business Cycle: Evidence from UK Manufacturing Firms, 1975–86. *Journal of Industrial Economics*, **41**(1), 29–50.
Malbera, F. and Orsenigo, L. (1995) Schumzeterian Patterns of Innovation. *Cambridge Journal of Economics*, **19**, 47–65.
Mueller, D.C. (ed.) (1990) *The Dynamics of Company Profits: An International Comparison*. Cambridge: Cambridge University Press.
Sims, C. (1998) Using a Likelihood Perspective to Sharpen the Econometric Discourse: Three Examples. Mimeo.
Reinganum, J. (1989) The Timing of Innovation: Research, Development, and Diffusion. In *Handbook of Industrial Organization*. Vol. 1, pp. 849–908, *Handbooks in Economics*. Vol. 10, Amsterdam, Oxford and Tokyo: North-Holland.
Robinson, W., Kalyanaram, G. and Urban, G. (1994) First-mover Advantages from Pioneering New Markets: A Survey of Empirical Evidence. *Review of Industrial Organization*, **9**(1), 1–23.
Tirole, J. (1988) *The Theory of Industrial Organization*. Cambridge, MA and London: MIT Press.