# European Research Joint Ventures and Innovation: a microeconometric analysis of RJV impact on firms' patenting activity

**Lucia Cusmano**
**(lucia.cusmano@uni-bocconi.it)**

CESPRI – Università L. Bocconi
Via S. Mansueto 4 – 20136 Milan (Italy)
Tel. +39-02-58363397
Fax. +39-02-58363399
www.cespri.uni-bocconi.it

Prepared for the Nelson and Winter DRUID Summer Conference, Aalborg, Denmark, June 12-15 2001

## Introduction

Co-operative technological activities have been extensively investigated by theoretical literature, in the broader framework of technology policy analysis (e.g. Stoneman and Vickers, 1988; Geroski, 1995; Llerena and Matt, 1999), game theoretical models (e.g. Spence, 1984; Katz, 1986; d'Aspremont and Jacquemin, 1988; Sinha and Cusumano, 1991; Kamien et al., 1992) or strategic management studies (e.g. Teece, 1986, 1992; Contractor and Lorange, 1988; Hagedoorn, 1993; Hagedoorn and Narula, 1996).

Despite the increasing attention to the rationales, benefits and costs of R&D agreements, there has been little corresponding empirical work on their efficacy. The empirical investigations have been conducted mostly at the level of case-studies or one-time surveys of participants, whereas the emphasis of technology policy on extensive collaborative programmes, covering a wide range of industrial research areas and involving very large numbers of organisations, has been barely matched by systematic investigations on large samples[1].

At European level, recent econometric analyses, based on comprehensive datasets concerning European Programmes, have focussed on the impact of RJV affiliation on firms' performance, as measured, for instance, by total factor productivity (e.g. Benfratello and Sambenelli, 2000; Bussoli, 2000). However, no extensive study about the impact of collaborative R&D activities, supported by EU or intergovernmental agreements, on firms' *innovative potential*, as measured by effective patenting output, has yet been produced.

The present work proposes a critical analysis of the relationship between participation to European RJVs and patenting activity, which elaborates a large panel data of European firms, the RJV_EPO dataset. The assessment exercise focuses on R&D consortia in two high tech areas, that have been assigned increasing attention and financial resources by EU institutions, and play a critical role for the development of a competitive European Innovation System: *Information and Communication Technology* (IT) and *Medical and Biotechnology* (MB).

The dataset covers the period 1992-96 and includes both participants to European RJVs and a counterfactual of non-participating firms. The latter is selected according to a stratified sampling methodology, which attempts to replicate in the "Control Group" the industry and country composition of RJV members.

The formalisation of the patenting process, presented in Section 2, follows the Knowledge Production Function Approach, which is extensively used in the literature about R&D

[1] In this respect, an interesting exception is represented by the work of Branstetter and Sakakibara (1998), who analyse the Japanese experience of government-supported R&D consortia and investigate the impact of participation on firms' patenting output.

productivity and spillovers (Griliches, 1984, 1995,1998; Jaffe, 1986). The production of knowledge is related to industry and individual characteristics, including the participation to collaborative R&D projects. Patent activity is assumed to reflect this process of knowledge generation, although propensity to patent may vary over industries and individual organisations.

Section 3 discusses the sampling methodology and describes the data. Two nested samples are selected: a smaller "balanced" sample and a larger "imputed" sample, where missing values of employment have been imputed using in and out of the sample information.

For both samples, the descriptive statistics reveal a high degree of heterogeneity between RJV participants and non-participants: firms which enter European R&D consortia are, on average, bigger and more innovative than non-participants. This first-level analysis points to the need for addressing the issue of self-selectivity of RJV members. That is, the positive correlation between patenting activity and participation may reflect the higher propensity to participate of more innovative agents, rather than a learning process favoured by R&D partnership.

The assessment of self-selectivity, proposed in section 4, follows the approach of recent empirical works in the area of international trade, which deal with the evidence of a positive correlation between firms' productivity and exporting activities (Bernard and Jensen, 1995; Clerides at al., 1998, Aw et al., 2000). The overall findings of the proposed empirical tests seem to be consistent with the self-selection hypothesis, but there is also evidence of learning processes taking place once the (more innovative) firms enter R&D consortia, so that the innovative gap in relation with non-participating firms widens.

Section 4 further discusses the use of appropriate control variables, such as size and knowledge capital, which may capture "exogenous" innovative propensity and allow to disentangle the self-selection of innovative firms from the learning effect induced by co-operative action.

The econometric assessment proposes the application of non-linear panel data models for exploring the relationship between RJV affiliation and patenting behaviour, emphasising the significant drawbacks of a linear approximation, when dealing with count data which exhibit a high frequency of zero-value observations and a greatly skewed distribution.

Section 5 reports the econometric analysis based on Poisson and Negative Binomial models, developed by Hausman, Hall and Griliches (1984), that generalise the basic assumptions of standard fixed and random effects models to the case of count data and allow to handle the integer property of the patent variable.

Controlling for cumulated innovative experience, size and industry effects, it appears that participation to European R&D consortia has positively affected patenting output in the MB area, whereas the impact seems negligible in the IT area. The above finding is robust to different specifications of the key explanatory variables and induce a thorough comparison of nature of RJV participation in the two areas.

Results about alternative specifications, investigating the impact of cumulated knowledge and the role of peculiar factors, such as firms' relation to large industrial groups or affiliation to both Framework and EUREKA programmes, are also discussed. In particular, in the last part of the work, we evaluate whether the relationship between patenting activity and RJV participation changes over size classes. The estimation results point again to interesting differences between technological areas. In both IT and MB fields, we estimate a negative impact of RJV participation on small firms' patenting activity, whereas the effect appears to differ, in the two areas, for other size classes. In the IT field, for medium-size firms the estimated impact is slightly significant and positive, whereas it is negligible for large firms. In the MB area the effect appears greatly significant and positive for medium-large firms.

The final section comments on the main results of the microeconometric analysis and underlines issues and aspects that may represent the object of further empirical research.

## 2- The analytical framework: knowledge production function approach

The theoretical and empirical literature investigating the complex chain of links and feedback leading to industrial innovation assumes, more or less explicitly, that innovativeness hinges on a process of knowledge creation and accumulation. Modelling efforts and empirical investigations in this area have focussed in particular on the knowledge accretion generated by R&D investments, and related these latter to firms' economic performance, or, more widely, to economic growth (Griliches, 1995).

The creation of knowledge represents the essential focus of the Knowledge Production Function (KPF) approach, which assumes that knowledge, like other tangible output, is the result of a production process, that involves heterogeneous inputs (Griliches, 1984). Hence, knowledge creation is modelled employing standard production functions, including R&D investments as a basic input of the process, and, since knowledge is not directly observable, empirical analyses are mostly based on the assumption that knowledge accumulation reflects into firms' performance.

This approach is extensively adopted in the literature investigating the relationship between R&D and productivity (Griliches 1984, 1995, 1998), or, as in the work by Jaffe (1986), the nature and effects of R&D spillovers, where a knowledge production function is postulated, generally in the form of a modified Cobb-Douglas, taking R&D investment as the main input of the process. Firms' patenting performance is then assumed to be proportional to such knowledge output, controlling for industry- and firm-specific effects.

Branstetter and Sakakibara (1998) extend the standard model to include co-operative R&D as a distinct input of the knowledge production process. Firms are assumed to generate new knowledge by investing in in-house R&D and by pooling resources and sharing risks with technological partners, i.e. participating to research consortia:

$$N_{it} = R_{it}^{\beta} e^{\gamma C_{it}} e^{\sum_d \delta_d D_{id}} e^{\varepsilon_{it}} \qquad \mathrm{i}=1,\dots,\mathrm{n} \qquad \mathrm{t}=1,\dots,\mathrm{T} \qquad (1)$$

where $N_{it}$ is the amount of knowledge produced by firm *i* at time *t*, $R_{it}$ is firm *i*'s in-house R&D expenditure at time *t*, $C_{it}$ is a measure of firm *i*'s intensity of participation in consortia at time *t* (i.e. the number of RJVs in which firm i is involved at time t), $D_{id}$'s are industry dummy variables that capture industry-level differences in “opportunity” conditions, $\varepsilon_{it}$ is a random component related to firm and time.

New knowledge is not directly observed but it reflects into patent applications ($P_{it}$) according to the following:

$$P_{it} = e^{\sum_d \omega_d D_{id}} e^{\eta_i} N_{it} \qquad (2)$$

where industry dummies now relate to industry-level differences in the propensity to patent, or “appropriability” conditions, and $\eta_i$ is a firm-specific component[2].

Substituting (1) into (2) and taking logs yields a simple log linear relationship:

$$p_{it} = \beta r_{it} + \gamma C_{it} + \sum_d \theta_d D_{id} + u_{it} \qquad (3)$$

[2] Accordingly, the ratio $P_{it}/N_{it}$ represents firm *i*'s propensity to patent at time t and is related to industry as well as firm specific characteristics.

where $\theta_d=(\omega_d+\ \delta_d)$ captures the effects of inter-industry differences in both opportunity and appropriability conditions, and the error term $u_{it}$ consists of unmeasured firm-specific effects and a truly random, *iid* error term component:

$$u_{it} = \eta_i + \varepsilon_{it}$$

At this stage, we slightly modify the model proposed by Branstetter and Sakakibara (1998) and assume firm's patenting activity draws on the stock of cumulated knowledge, which represents the "cognitive capital" that fuels technological advancements. In other terms, we assume research expenditure does not enter directly the relationship. Rather, R&D contributes, together with less formal activities, to the accumulation of knowledge capital. Current knowledge production depends on the *stock* of knowledge the firm has accumulated over time, rather than on the *flow* of R&D activities that take place simultaneously.

In addition, we assume knowledge creation is related to firm's size, as measured, for instance, by the number of employees. The generation of new knowledge by firm *i* is thus formalised as:

$$N_{it} = L_{it}^{\alpha} K_{it-1}^{\beta} e^{\gamma C_{it}} e^{\sum_d \delta_d D_{id}} e^{\varepsilon_{it}} \qquad (4)$$

where $L_{it}$ is the level of employment, $K_{it-1}$ is the firm's stock of knowledge cumulated up to time t-1, which fuels the generation of new ideas at time t. $C_{it}$, $D_{id}$ and $\varepsilon_{it}$ are defined as above. We assume the propensity to patent is defined by equation (2), and allow for a time specific effect, exp ($\phi_t$), in the relationship between $P_{it}$ and $N_{it}$. The term $\phi_t$ is meant to capture exogenous changes over time in the propensity to patent that are not explained by variations in the other factors. Substituting (4) in (2) and adding the time term, we get:

$$P_{it} = L_{it}^{\alpha} K_{it-1}^{\beta} e^{\gamma C_{it}} e^{\sum_d (\delta_d + \omega_d) D_{id}} e^{\varepsilon_{it} + \eta_i + \phi_t} \qquad (5)$$

Taking logs gives us a log linear relationship, that is close to the standard specification estimated in the works concerned with R&D and productivity:

$$p_{it} = \eta_i + \phi_t + \alpha l_{it} + \beta k_{it-1} + \gamma C_{it} + \sum_d \theta_d D_{id} + \varepsilon_{it} \qquad (6)$$

Standard panel data techniques can be used for the estimation of the above two-way error component model (Hsiao, 1986). However, the log linear specification requires that logarithmic values are defined, i.e. the dependent variable and the regressors take positive values. In the case of count data, such as patents, this may require a significant adjustment of variables, since the zero-value is a natural outcome of the process being modelled.

Pakes and Griliches (1984), Bound et al. (1984) and Branstetter and Sakakibara (1998) solve the "zero problem" setting the log of patents and the log of patent stock to zero for all zero patent and patent stock observations, and allowing those firms to have a separate intercept, i.e. generating dummy variables that flag recoded values.

The zero problem is particularly severe in our samples, where non innovative units account for about 80% of the firms. The high frequency of zeroes leads to search for different specification that may allow to handle the integer nature of the patent variable, for which we are forcing a rather questionable continuous approximation.

The problems related to the linear specification when modelling patenting behaviour have been extensively discussed by Hausman, Hall and Griliches (1984), who develop generalisations of count data statistical models, such as the Poisson and Negative Binomial distributions, focussing specifically on empirical investigations related to patents[3].

Section 5 presents the results of the application of these models to a sub-sample of the RJV_EPO database, identified in accordance with the technological areas of the projects to which firms participated. The non-linear models are employed to investigate the relationship between the patent variable and the regressors identified in the linear specifications. Section 4 discusses in more detail the role of these "control variables" for disentangling self-selection and learning effects.

## 3- Data and Sampling Methodology

The RJV_EPO dataset represents the main source of information for the current investigation. This dataset combines information about RJVs promoted within Framework Programmes and

[3] The Poisson process is a natural candidate for a better formalisation of the innovative process, as proxied by patenting activity, since the integer property of the outcome $P_{it}$ is handled directly and the zero value represents a natural outcome of such a distribution. The statistical specification proposed by Hausman, Hall and Griliches (1984) accounts for both individual specific effects and overdispersion, i.e. variance to mean ratio greater than one. The Negative Binomial model generalises the Poisson one, introducing randomness both across firms and across time.

EUREKA Programme in all technological areas with data about firms patenting activity retrieved from the European Patent Office[4].

The present analysis focuses on two large and fast growing research fields: *Information and Communication Technology* and *Medical and Biotechnology* (IT and MB hereafter). The interest in these areas is driven by their critical role for the development of a competitive European Innovation System and by the differences in both structure and degree of heterogeneity of collaborative patterns, and in the nature of the knowledge base underpinning their technological advancements.

The panel data analysis refers to a five-year time period, 1992-96, and to a limited set of firms comprised in the RJV_EPO dataset, so that RJV and patent data can be used in conjunction with sector and employment information extracted from AMADEUS database[5].

The analysis covers firms from the main industries investing in IT and MB cooperation, i.e. we have dropped out of the dataset "outlier sectors", which are present only occasionally in European RJVs in the two areas of interest.

The resulting "RJV sample" comprises 491 firms in the IT area and 199 firms in the MB area. However, for some of those units, the information provided by AMADEUS about employment level is not complete over the entire period 1992-96. Hence, we label this set "**incomplete RJV sample**"[6]. The complete or "**balanced RJV sample**", obtained from case-wise deletion of incomplete cases, is constituted by 228 firms in the IT area and 78 firms in the MB area.

---

[4] The dataset represents the output of a research work conducted within the EC TSER Programme's KNOW FOR INNOVATION Project (*Innovation-Related Knowledge Flows in European Industry: Extent, Mechanism, Implications*), Contract No. SOE1-CT98-1118, DGXII G4, co-odinated by Prof. Y. Caloghirou of the National Technical University of Athens (Greece).

In more detail, the RJV_EPO dataset results from the merge of three databases:

a) CORDIS (Community Research and Development Information Service), which contains details about the consortia established in Europe within the Framework Programmes (1984-2000), such as general programme type (e.g. 3rd FWP), specific programme (e.g. ESPRIT, RACE, etc), technical area, time-scale, number of participants, participants' type (e.g. firm, University, public organisation, etc.) and country;
b) EUREKA (European Research Co-ordinating Agency) central database, which gives information about projects and participants to EUREKA Programme (1985-2000), such as, similarly to CORDIS, area of research, technological development envisaged, project status (completed, in execution, or planned) and time-scale, participants' organisation type and country, contribution of each participant to the project;
c) EPO-CESPRI database, which provides full details about patents applied for and granted by the European Patent Office (1978-1996).

[5] AMADEUS gives longitudinal financial and sectoral information for approximately 200,000 firms in Europe over the time period 1992-96.

[6] Referring to "incomplete sample" seems more appropriate than using the standard notion of "unbalanced sample", which normally refers to missing values of the dependent variable or asymmetry in T over the observed units. The dependent variable of our estimation exercise is the number of a firm's patent applications in a given year. The EPO dataset is "complete" and gives no problem of missing values for any of the units in our sample.

Since the evaluation of RJV impact on firms patenting activity needs a counterfactual, we have constructed a "Control Group" (CG) of firms which have not participated to European RJVs. The Control Group has been extracted from AMADEUS by way of a stratified sampling methodology. In fact, the counterfactual mimics the RJV sample, in terms of both sectoral and country composition. In other terms, for each industry, the share of firms included in the Control Group is close to the share included in the RJV sample, and within each industry, the country composition of the Control Group reflects the country composition of RJV participants[7].

The resulting "**incomplete CG sample**", which constitutes the counterfactual for the "incomplete RJV sample", is composed of 552 and 167 firms from industries participating to IT and MB consortia respectively.

The "**balanced CG sample**" is obtained by discarding incomplete cases and includes 280 firms for the IT area and 113 firms for the MB area.

Estimations are first performed on balanced samples, composed of both RJVs participants and CG units, with a total number of firms equal to 508 (IT area) and 191 (MB area). The pooled balanced sample is composed of 658 firms[8].

However, the "case-wise" deletion, induced by employment missing values, implies a remarkable loss of information in comparison with the original dataset. Moreover, if missingness does not occur at random, but rather the pattern of missing values is related to other variables in the panel, then the econometric analysis that discards this information is affected by selectivity bias (Little and Rubin, 1985).

Therefore, we perform a second set of estimations on the larger dataset, after imputing the employment missing values, using an imputation technique based on sector- and country-specific average rate of growth of employment, with the addition of a stochastic component[9].

Table 1 summarises the composition of balanced and imputed pooled (ALL hereafter), IT and MB samples. Estimation results presented in the following sections mainly refer to the balanced samples. We report detailed estimates on the imputed samples when results

[7] Again, AMADEUS dataset presents sparse missing values for the employment level over the period 1992-96. Only units for which the data about employment is given for at least one year have been taken into account.

[8] Firms entering both IT and MB samples are counted only once in the pooled sample.

[9] The imputation draws on out of the sample information, since it exploits the whole AMADEUS database. In short, time *t* employment missing value for firm *i* of sector *k* and country *w* is imputed taking into account time series information about the firm employment (number of employees at time *t+1* or *t-1*) and extra-sample information about the average employment rate of growth at time *t* in sector *k* and country *w*. Moreover, in order to attenuate the problem of reduced sampling variability, which typically arises when adopting imputation techniques, we add to the sector and country specific rate of growth a stochastic component, normally distributed, with zero-mean and standard deviation equal to the estimated standard deviation of the sector and country specific rate of growth at time t.

significantly differ or we require a sufficiently large sample size in analysing specific subgroups.

**Table 1**

**Balanced and imputed samples: composition**

| **SAMPLE** | **Total number of firms** | | ***RJVs participants*** | | ***Control Group*** | |
|---|---|---|---|---|---|---|
| | Balanced | Imputed | *Balanced* | *Imputed* | *Balanced* | *Imputed* |
| ALL* | 658 | 1349 | *302* | *714* | *356* | *635* |
| IT | 508 | 1043 | *228* | *491* | *280* | *552* |
| MB | 191 | 366 | *78* | *199* | *113* | *167* |

* The number of firms in ALL does not result from the exact sum of firms in IT and MB, since units belonging to both subsamples are entered only once in ALL.

Table 2 describes the distribution of firms over industrial classes (2- and 3-digit US SIC Code). The industries included in the analysis are obviously mostly related to Information and Communication Technology and Chemical and Pharmaceutical areas. Two classification criteria are proposed, at different level of aggregation. At the lower level of aggregation we split computer (COMP_P), communication (COMM_P) and drugs (DRUGS_P) sectors, distinguishing between computers (COMP) and computer services (COMP_SER), communication (COMM_SER) and communication equipment (COMM), drugs (DRUGS) and medical instrument (MED_INS). We will mainly refer to the higher level of aggregation, unless estimates based on the lower level give significantly different results.

### 3.1- Samples Statistics

Figures 1 and 2 report the distribution of firms over size classes for the balanced and imputed samples, distinguishing the specific technological areas. In both samples, medium-size firms constitute the largest category, whereas small firms never account for more than 10% of the units.

Table 3 reports the country distribution with reference to the balanced samples. UK, France and Germany are the most represented countries. British firms represent by far the relative majority in the IT field, whereas in the MB area they are outnumbered by German and French firms. The Italian participation is not evenly distributed, rather it concentrates in the IT area.

**Table 2**
**COMPOSITION OF INDUSTRY CLASSES**

***Balanced Samples*** — ***Frequency***

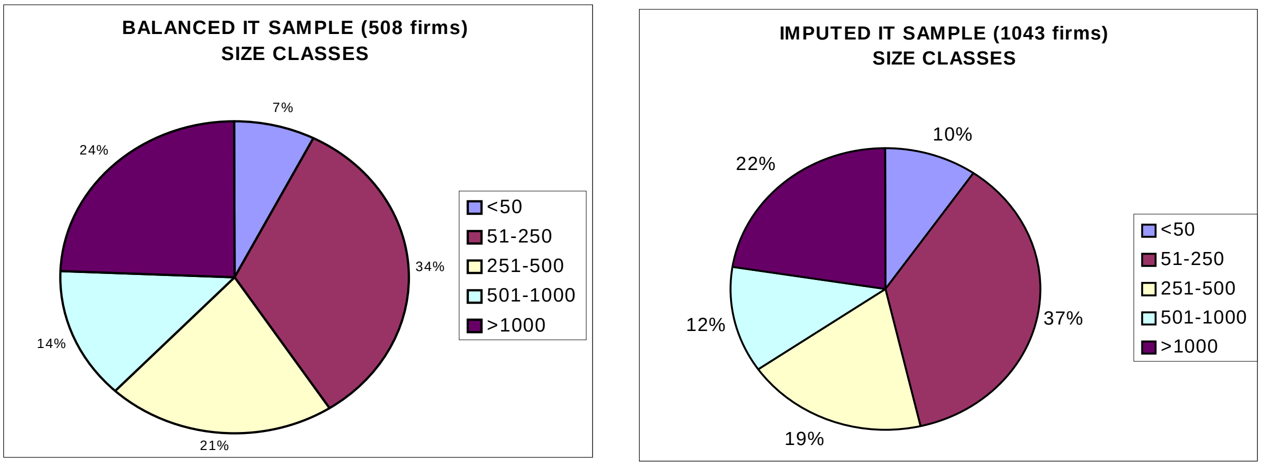
| Industry dummies | CLASSES | US SIC Code | ALL | IT | MB |
|---|---|---|---|---|---|
| FOOD | Food & Kindred Products | 20 | 53 | - | 53 |
| CHEM | Chemicals (excl. Drugs) | 28 (excl. 283) | 19 | - | 19 |
| DRUGS_P | Drugs & Medical Instruments | 283, 384 | 74 | 16 | 60 |
| *DRUGS* | *Drugs* | *283* | *(51)* | *(0)* | *(51)* |
| *MED_INS* | *Medical Instruments* | *384* | *(23)* | *(16)* | *(9)* |
| COMP_P | Computers & Computer Services | 357, 737 | 174 | 174 | - |
| *COMP* | *Office, Computers & Accounting Equipment* | *357* | *(53)* | *(53)* | - |
| *COMP_SER* | *Computer Services* | *737* | *(121)* | *(121)* | |
| E_EQUIP | Electric Equipment & Supplies | 36 (excl. 365, 366, 367) | 31 | 31 | - |
| COMM_P | Communication & Communication Equipment | 365, 366, 367, 48 | 87 | 87 | - |
| *COMM* | *Communication Equipment* | *365, 366, 367* | *(61)* | *(61)* | - |
| *COMM_SER* | *Communication* | *48* | *(26)* | *(26)* | |
| P_EQUIP | Professional & Scientific Equipment | 38 (excl. 384) | 41 | 40 | 15 |
| TRADE | Wholesale Trade Durable & Nondurable Goods | 504, 506, 508, 514 | 107 | 95 | 27 |
| ENGMAN | Engineering & Management Services | 871, 873 | 72 | 65 | 17 |
| **Total** | | | **658** | **508** | **191** |

***Imputed Samples*** — ***Frequency***

| Industry dummies | CLASSES | US SIC Code | ALL | IT | MB |
|---|---|---|---|---|---|
| FOOD | Food & Kindred Products | 20 | 89 | - | 89 |
| CHEM | Chemicals (excl. Drugs) | 28 (excl. 283) | 39 | - | 39 |
| DRUGS_P | Drugs & Medical Instruments | 283, 384 | 154 | 23 | 135 |
| *DRUGS* | *Drugs* | *283* | *(115)* | *(0)* | *(115)* |
| *MED_INS* | *Medical Instruments* | *384* | *(39)* | *(23)* | *(20)* |
| COMP_P | Computers & Computer Services | 357, 737 | 366 | 366 | - |
| *COMP* | *Office, Computers & Accounting Equipment* | *357* | *(111)* | *(111)* | - |
| *COMP_SER* | *Computer Services* | *737* | *(255)* | *(255)* | |
| E_EQUIP | Electric Equipment & Supplies | 36 (excl. 365, 366, 367) | 56 | 56 | - |
| COMM_P | Communication & Communication Equipment | 365, 366, 367, 48 | 201 | 201 | - |
| *COMM* | *Communication Equipment* | *365, 366, 367* | *(130)* | *(130)* | - |
| *COMM_SER* | *Communication* | *48* | *(71)* | *(71)* | |
| P_EQUIP | Professional & Scientific Equipment | 38 (excl. 384) | 75 | 73 | 20 |
| TRADE | Wholesale Trade Durable & Nondurable Goods | 504, 506, 508, 514 | 228 | 201 | 48 |
| ENGMAN | Engineering & Management Services | 871, 873 | 141 | 123 | 35 |
| **Total** | | | **1349** | **1043** | **366** |

## Figure 1

## IT Samples: distribution of firms by number of employees (average number 1992-96)




## Figure 2

## MB Samples: distribution of firms by number of employees (average number 1992-96)

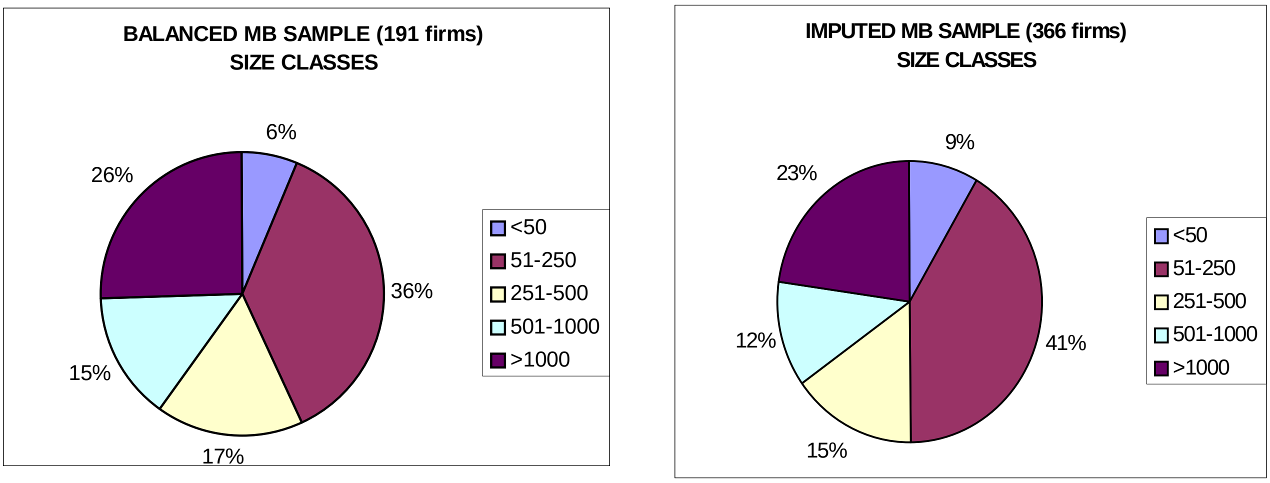



**Table 3**
**Balanced Samples - Country composition**

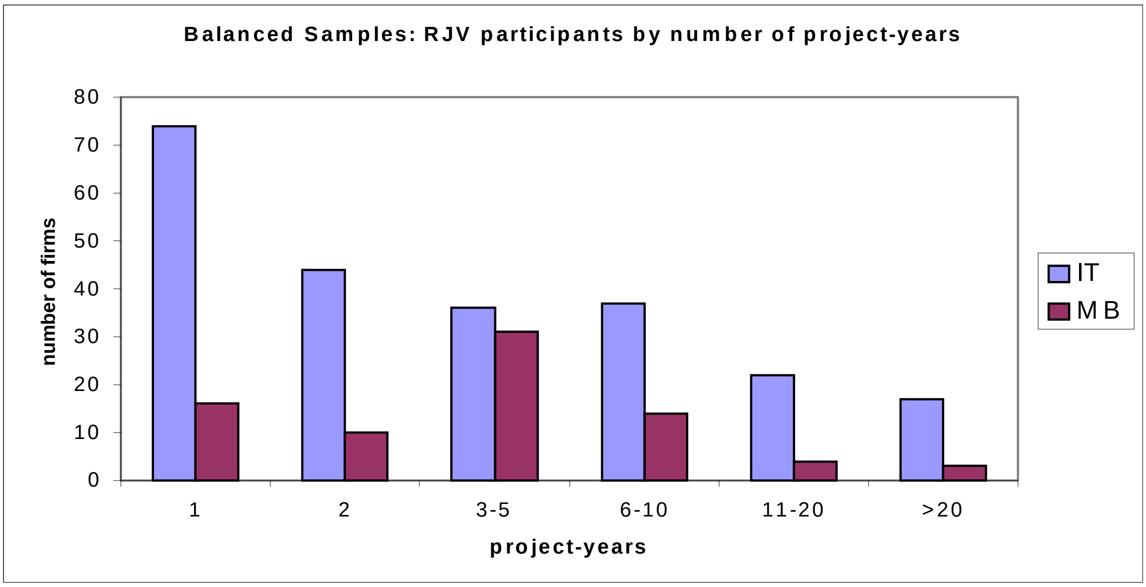
| Country | IT (N=508) | MB (N=191) |
|---|---|---|
| UK | 154 | 35 |
| France | 92 | 49 |
| Germany | 86 | 37 |
| The Netherlands | 46 | 31 |
| Italy | 41 | 8 |
| Denmark | 32 | 11 |
| Belgium | 17 | 8 |
| Sweden | 14 | 5 |
| Finland | 7 | 2 |
| Ireland | 8 | 1 |
| Portugal | 5 | 2 |
| Austria | 3 | - |
| Switzerland | 2 | 1 |
| Spain | - | 1 |
| Luxembourg | 1 | - |

Table 4 focuses on RJV members and give details about the "intensity" of their participation, as measured by the number of project-years, i.e. the number of years the firm has been affiliated to R&D consortia, where independent RJVs in the same year account for one project-year each. In the IT area, the frequency of participation is on average higher than in the MB area, although there is a much greater variability among participants. The greater variance in the IT sample is more evident in Figure 3, that compares the distribution over frequency classes of IT and MB firms in the balanced samples. Over the period 1992-96, for the great majority of IT members, participation has been limited to one year, whereas for MB firms the median is between 3 and 5 project years[10].

**Table 4**
**RJV participants: number of project-years**

| Sample | Firms | Mean | Std Dev | Min | Max |
|---|---|---|---|---|---|
| **Balanced** | | | | | |
| IT | 228 | 6.523 | 11.027 | 1 | 117 |
| MB | 78 | 5.217 | 5.971 | 1 | 41 |
| **Imputed** | | | | | |
| IT | 552 | 6.735 | 11.739 | 1 | 117 |
| MB | 168 | 4.738 | 4.853 | 1 | 41 |

**Figure 3**




Tables 5 and 6 provide the basic descriptive statistics about size and patenting activity, for the balanced and imputed samples respectively, distinguishing between RJV participants and non-participants. The simple comparison of average employment level and innovative output reveals a high degree of heterogeneity. Firms participating to RJVs are on average larger and more innovative than non-participants. The difference in the level of employment is very similar across technological areas, whereas the difference in average patenting activity is more marked in the MB area.

**Table 5**
**Balanced Samples - Descriptive statistics: employment and patenting**

| | Average Employment Level (1992-96) | | | | Patent Applications (yearly average, 1992-96) | | | |
|---|---|---|---|---|---|---|---|---|
| Sample | Mean | Std Dev | Min | Max | Mean | Std Dev | Min | Max |
| **IT** | | | | | | | | |
| Total | 1463.805 | 4532.615 | 4.8 | 70540 | 0.814 | 4.110 | 0 | 49 |
| RJV | 2514.572 | 6362.740 | 14 | 70540 | 1.307 | 5.407 | 0 | 49 |
| CG | 608.18 | 1660.859 | 4.8 | 23974 | 0.414 | 2.558 | 0 | 28 |
| **MB** | | | | | | | | |
| Total | 1592.305 | 3886.075 | 14.2 | 26915 | 0.638 | 6.120 | 0 | 84 |
| RJV | 2927.664 | 5337.671 | 14.2 | 26915 | 1.512 | 9.541 | 0 | 84 |
| CG | 670.552 | 1983.001 | 15 | 18051 | 0.035 | 0.264 | 0 | 2 |

**Table 6**
**Imputed Samples - Descriptive statistics: employment and patenting**

| | Average Employment Level (1992-96) | | | | Patent Applications (yearly average, 1992-96) | | | |
|---|---|---|---|---|---|---|---|---|
| Sample | Mean | Std Dev | Min | Max | Mean | Std Dev | Min | Max |
| **IT** | | | | | | | | |
| Total | 1485.341 | 5312.644 | 1.7 | 74098 | 0.735 | 5.016 | 0 | 141 |
| RJV | 2391.314 | 7077.469 | 1.7 | 74098 | 1.038 | 5.959 | 0 | 141 |
| CG | 466.812 | 1314.954 | 4.8 | 23974 | 0.394 | 3.648 | 0 | 98 |
| **MB** | | | | | | | | |
| Total | 1182.507 | 3034.667 | 7.2 | 26915 | 0.555 | 4.552 | 0 | 84 |
| RJV | 1887.667 | 4001.408 | 7.2 | 26915 | 1.070 | 6.667 | 0 | 84 |
| CG | 590.739 | 1667.819 | 10 | 18051 | 0.121 | 0.645 | 0 | 9 |

It is interesting to note that firms in the balanced samples exhibit an average size and a patenting level greater than entities in the imputed sample. The enlargement of the dataset produces the most relevant change in the MB area, where the average size of participating firms significantly decreases, and the gap between patenting activity of participants and non-participants reduces. In this sense, we may suspect of a selectivity bias. In fact, full employment data may be mostly available for larger firms. Thus, referring to this availability as a criterion for selecting the sample could bias its composition and, consequently, the estimation results.

[10] This evidence reflects the overall different patterns of participation to EU consortia and duration of projects in the two fields. In fact, in the IT field, firms tend to participate to a greater number of projects than in the MB area, but the average projects' time-scale is shorter.

Another kind of selection bias, related to the problem of endogeneity, is worth further investigation, since it may represent a serious concern for the interpretation of the econometric results. It regards the comparison of participants and counterfactual, which is at the very base of our exercise. In fact, even though firms of the Control Group result from a stratified sample selection procedure, which replicates in the counterfactual the sector and country composition of the corresponding RJV group, the remarkable differences in average size and level of patenting activity pose the important question of self-selectivity, that is, whether participation to RJVs and patenting activity are positively correlated simply because more innovative firms tend to participate more frequently to consortia. Therefore, before presenting the details of the statistical models and the results of the microeconometric investigation, we analyse the relationship between firms' patenting level and entry into European RJVs and assess whether the use of appropriate control variables can allow to disentangle the effects of learning by co-operating from the effects of self-selection.

## 4- Model specification and variables definition: controlling for self-selection

The econometric assessment of RJV impact on firms' patenting output requires, first of all, the identification of variables that *control* for significant factors influencing patenting behaviour, other than the affiliation to research consortia. The question of identifying appropriate control variables appears even more important once we consider the problem of self-selection. In fact, in case a positive statistical association was to be found, its interpretation would be made ambiguous by the problem of self-selection of RJV entrants. To what extent does the (possibly) higher innovativeness of RJV participants reflect the self-selection of innovative firms into RJVs? Is it possible to identify and distinguish the learning effects of co-operation from the level of innovativeness before RJV affiliation?

We hereby propose a statistical analysis of the problem and discuss the use of variables that should control for the relevant heterogeneity of RJV members and non-participants.

### 4.1 - Self-selection of innovative RJV participants and RJV learning effects

The issue of self-selectivity has not been explicitly addressed by the scarce econometric literature about R&D co-operation. Rather, our analysis follows the approach of very recent empirical works in the area of international trade, which deal with the evidence of a positive correlation between firms’ productivity and exporting activities. Bernard and Jensen (1995), Clerides et al. (1998), Aw et al. (2000) address the complex issue of whether exports play a causal role in generating higher productivity, since exporters gain knowledge and expertise in the export market (learning hypothesis), or the correlation simply reflects the decision to export by the most productive firms (self-selection hypothesis). The authors underline that both mechanisms are plausible, though their actual importance most likely varies across countries and industries, and propose some econometric tests for evaluating their relative importance.

There are clear conceptual analogies with the problem we are investigating in this paper. Hence, we adapt and implement a few tests developed in the international trade literature, with particular attention to the analysis of Aw et al. (2000).

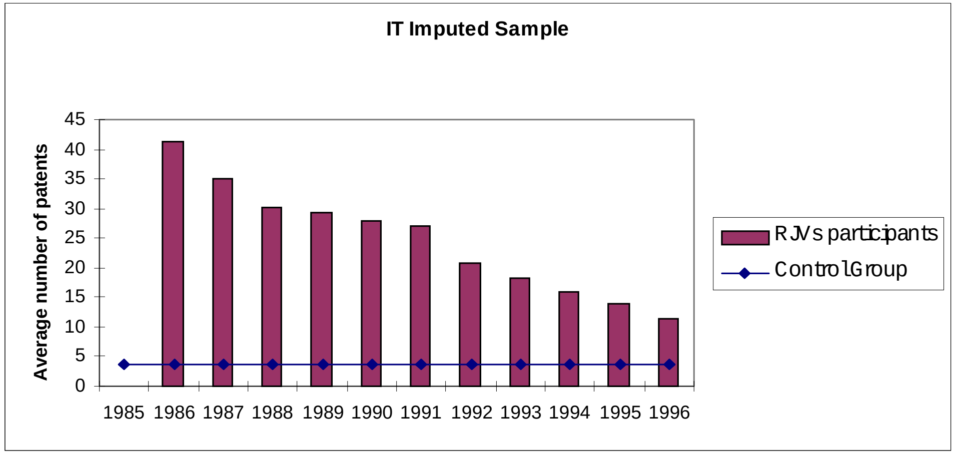
Firstly, we further characterise the innovativeness differential between RJV participants and non-participants in terms of patent stock. Figure 4 depicts for each year, from 1986 to 1996, the average number of patent applications, over the whole EPO horizon 1978-1996, by firms which entered at least one RJV by time t.

Except for the first years of operation of the MB programmes, the average number of participants' patent applications is remarkably higher that the average patent stock in the CG. In both IT and MB areas, there is clear evidence that early participants have been more innovative, on average, than later entrants. In fact, the average level of innovativeness of firms involved in European projects decreases over time, as if those programmes had attracted the most innovative entities in the early stages, and later extended to include less innovative firms. The decline is more evident in the IT area, whereas in the MB area, where very early "pioneers" have been barely innovative, the entry in the "RJV group", in 1996, of highly innovative firms has reversed the negative trend.

At this stage, we cannot draw any conclusion about causality, but can get further insight by considering the average level of innovativeness of RJV members *before* their first RJV participation, rather than up to 1996. Figure 5 depicts the average number of patent applications up to time t-1 by firms which entered for the first time a European RJV at time t. The cumulated innovative experience of new entrants is very heterogeneous over time. In the IT area, we can identify a few peaks, which roughly correspond to the starting years of new waves of projects, in particular 1986 for EUREKA and 1992 for the Third Framework Program, which has directed a conspicuous budget share to co-operative R&D in Information and Communication Technology. The last wave of MB programs has attracted highly innovative new members, whereas, except for the peak of entrants' innovativeness in 1989, early RJV participants did not significantly innovate *before* their first RJV experience.

**Figure 4**

**RJV participants' average level of innovativeness. Patent stock (1978-1996) of RJV participants by year t**



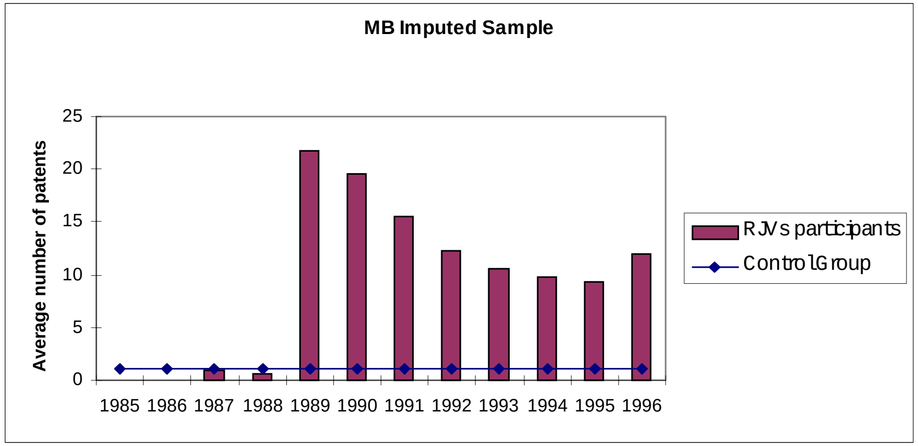
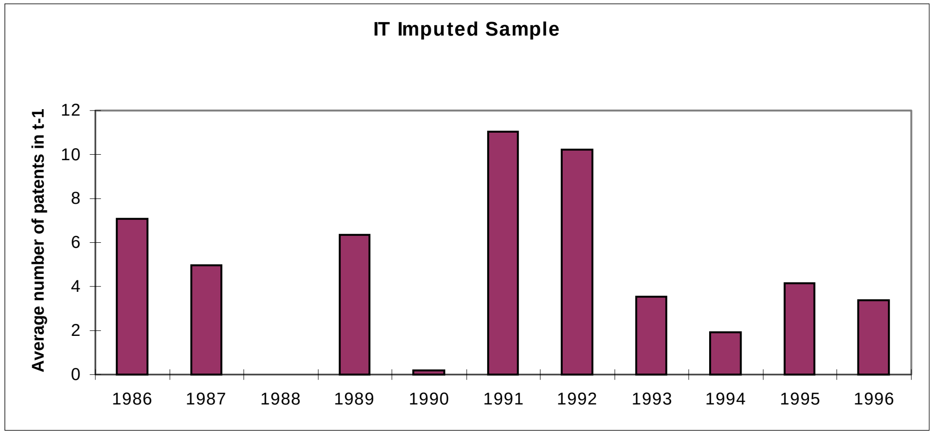
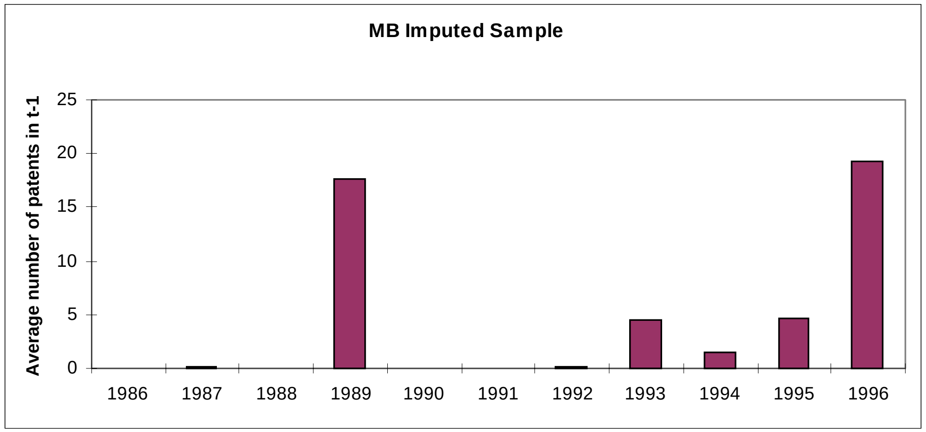




**Figure 5**

**RJV entrants' innovative experience. Patent stock at time t-1 of new RJV participants at time t**



In order to evaluate, in statistical terms, the relationship between the level of patenting activity and RJV participation, we fully exploit the time series dimension of our dataset and group firms into "transition groups" (Table 7). At each time t, we compare firms' status, taking into account their "backward (from t-1 to t) and "forward" (from t to t+1) transition.

**Table 7**

**Transition groups: RJV affiliation over time**

| | **t-1** | **t** | **t+1** |
|---|---|---|---|
| **CG** | 0 | 0 | 0 |
| **LATE*** | 0 | 0 | 0 |
| **EARLY**** | 0 | 0 | 0 |
| **ENTRY** | 0 | 0 | 1 |
| **INCUMBENT** | 1 | 1 | 1 |
| **FREQ** | 1 | 0 | 1 |
| **EXIT** | 1 | 1 | 0 |
| **NEW** | 0 | 1 | 1 |
| **OLD** | 1 | 0 | 0 |

Note: RJV membership=1, RJVs non-membership=0
* at least one participation after t+1
** at least one participation before t-1

The Control Group (CG) represents the benchmark, i.e. innovativeness (number of patent applications per year) in the other groups is defined in terms of differential from innovativeness of non-members. Therefore, we distinguish firms which never entered consortia (CG) from firms which are not affiliated in the three contiguous time periods but become members in a later time (LATE) or participated to a project for the last time before t-1 (EARLY). INCUMBENT refers to firms which are members in three contiguous years, while FREQ (that stands for "frequent"), flags firms which are not participating to a consortium at time t, but were and will be members in the adjacent years, hence have participated to at least two projects in a short time period. EXIT refers to firms which are in the final stage of some RJV project, NEW refers to firms which enter a consortium at time t and continue their joint investment in t+1, OLD refers to "newly old" members, whose participation has recently expired. Finally, ENTRY represents firms which enter a RJV for the first time in t+1.

If self-selection is at work, we expect the relative innovative level in period t to be positively correlated with RJV participation in t+1, i.e. with the variable ENTRY. Therefore, we run a regression of firms' innovative differential, with respect to the average CG level, at time t ($PAT\_DIFF_{it}$), on the variables describing firms' transition status at time t[11].

[11] At each time t, only one of the variables is equal to 1 for firm i, whereas all others equal 0. The estimated equation includes time dummies.

$$PAT\_DIFF_{it}= \mu + \eta_i + \beta_1\ ENTRY_{it}+ \beta_2\ INCUMBENT_{it}+ \beta_3\ NEW_{it}+ \beta_4\ EXIT_{it}+ \beta_5\ LATE_{it}+$$
$$+ \beta_6\ EARLY_{it}+ \beta_7\ OLD_{it}+ \beta_8\ FREQ_{it}+ \Sigma_t\ YEAR_t + \varepsilon_{it}$$

where

$$PAT\_DIFF_{it}= Pat_{it} – Mean^{CG}{}_t\ (Pat^{CG}{}_{it})$$

Table 8 presents the results of the balanced sample Random Effects estimation, where firms specific effects ($\eta_i$) are assumed to be random and uncorrelated with explanatory variables.

**Table 8**
**Self-selection and learning: average patenting differential and transition into or out of RJVs (Balanced Sample- Random Effects Model, 1985-1996)**

| Dep. Var. | ALL (N=658) | | | IT (N=508) | | | MB (N=191) | | |
|---|---|---|---|---|---|---|---|---|---|
| PAT_DIFF | Coeff | (z-value) | | Coeff | (z-value) | | Coeff | (z-value) | |
| ENTRY | 1.200 | (2.982) | ** | 0.477 | (1.321) | | 1.039 | *(1.484)* | |
| INCUMBENT | 0.574 | (3.770) | *** | 0.188 | (1.317) | | 1.351 | *(6.226)* | *** |
| NEW | 0.292 | (1.582) | | -0.035 | -(0.216) | | 0.379 | *(1.422)* | |
| EXIT | -0.102 | -(0.351) | | -0.142 | -(0.556) | | 0.348 | *(0.800)* | |
| LATE | 0.811 | (2.285) | | 0.882 | (2.705) | * | 0.898 | *(1.360)* | |
| EARLY | 0.117 | (0.403) | | 0.212 | (0.597) | | 0.242 | *(0.492)* | |
| OLD | 0.081 | (0.682) | | -0.093 | -(0.911) | | -0.145 | *-(0.812)* | |
| FREQ | 2.418 | (3.288) | *** | 3.530 | (4.679) | *** | -0.537 | *-(0.382)* | |
| CONS | -0.004 | -(0.114) | | -0.001 | -(0.007) | | -0.096 | *-(0.211)* | |
| Wald chi2 (Pr>chi2) | 63.080 | (0.000) | | 62.490 | (0.000) | | 62.580 | *(0.000)* | |
| Nr Obs | 6842 | | | 5003 | | | 1887 | | |

Note: all estimated equations include time effects
z-values in parenthesis
* significant at 10% level
** significant at 5% level
*** significant at 1% level

In the pooled sample (ALL) we detect statistical significance of the ENTRY (5% level), INCUMBENT (1% level) and FREQ (1% level) variables. Hence, the econometric analysis on the innovative differentials of transition groups points at a positive correlation between RJV entry and patenting level *prior* to entry.

The overall finding is consistent with the self-selection hypothesis, but the evidence is less clear when distinguishing between technological areas. In fact, the result does not hold, when we run the regression on separate samples. In the IT sample, the outcome is only slightly different, since we find a positive correlation between current patenting activity and RJV participation in a later time: ENTRY is non significant, but LATE is significant at 10% level. This result is consistent with the representation of Figure 5, where we detected a high level of patent stock of new RJV participants. FREQ is also significant in the IT sample, whereas, in the MB area only the INCUMBENT variable is significant (1% level) and has a high coefficient. In the MB sample, there is no significant evidence of more innovative firms at time t being more likely to enter a European RJV at time t+1. Rather, we find the innovative differential to be significantly correlated with a continuous involvement into co-operative R&D. This preliminary result for the MB sample is more consistent with the learning hypothesis than with the self-selection one.

However, in interpreting results, we should also take into account the limits of the patent variable in depicting innovative *potential*. That is, there may be cases in which the test rejects the hypothesis of self-selection because firms have not significantly patented before entry, but in fact there is self selection by firms with a high innovative potential. Other factors, such as prestige of researchers, number of publications, participation to meeting or conferences, could be considered as signals of innovative capacity, which is perceived by possible partners and is exploited during or after the participation to consortia.

We further investigate self-selection and learning assumptions by considering the change of innovative differential over time. We may interpret the evidence of a widening post-entry differential as favourable to the learning hypothesis. Therefore, we relate the average of firms' innovative differential (from the CG level) over the entire RJVs' period (1985-1996) to the pre-entry level of the differential and the post-entry change in the differential[12], where, for frequent participants, the threshold for distinguishing between "pre" and "post" is represented by the year of affiliation to the *first* RJV.

Table 9 reports the results of this cross-section regression with respect to the pooled sample and the two technological areas[13]. The evidence from the three analysis points to the same direction: the participants' higher average level of innovativeness (AV_DIFF) is mainly explained by a high innovative activity prior to the entry (DIFF_PRE). However, there also appears to be a significant contribution by the increase in differential taking place after the entry into the first RJV (D_POST)[14].

**Table 9**
**Self-selection and learning: patenting differential change**
**(Balanced Sample, 1985-1996)**

| Dep Var | ALL (N=302) | | IT (N=228) | | MB (N=78) | |
|---|---|---|---|---|---|---|
| AV_DIFF | Coeff | (z-value) | Coeff | (z-value) | Coeff | (z-value) |
| DIFF_PRE | 1.148 | (54.447) *** | 1.149 | (36.693) *** | 1.206 | (62.050) *** |
| D_POST | 0.361 | (38.965) *** | 0.340 | (19.685) *** | 0.372 | (61.530) *** |
| CONS | -0.008 | -(0.225) | -0.042 | -(0.896) | -0.043 | -(1.119) |

Note: AV_DIFF = $\text{Mean}_{1985\text{-}1996}\,[\text{Pat}_{it} - \text{Mean}_t^{CG}]$

[12] D_POST is the average change, from the entry date to 1996, of the differential in the number of patent applications per year.

[13] Results always refer to the balanced sample.

[14] When excluding from the analysis the highly innovative firms which entered in 1996 only (see Figure 4), results do not significantly differ. In the MB area only, the coefficient of DIFF_PRE strongly reduces, whereas D_POST coefficient increases, but still both variables are significant at 1% level.

$DIFF\_PRE = Mean_{1985\text{-}entry} [Pat_{it} - Mean_t^{CG}]$
$D\_POST = Mean_{entry\text{-}1996} [(Pat_{it} - Mean_t^{CG}) - DIFF\_PRE]$

## 4.2- Control variables for disentangling self-selection and learning

Self-selection does not rule out learning, therefore the real basic issue for our estimation is that of identifying control variables which account for the high innovative level of RJV members, independently of their RJV participation. Equation (6) includes two factors which should capture, to a relevant degree, this exogenous higher propensity to innovate. The first one is $k_{t-1}$ , the stock of knowledge that has been cumulated up to time t-1. We employ the stock of patent applications from time 0 (i.e. 1978, the first year of EPO operation) to time t-1 as a proxy of this "cognitive" capital, which fuels the generation of knowledge at time t. This variable represents patenting experience prior to time t, that could be the result of other co-operative experiences as well as of in-house R&D investments or informal activities. In any case, it allows to evaluate the effect of RJV participation separately from the potential benefits of an already extended knowledge base.

In this sense, we view the patent flow as a measure of output of knowledge-producing activities and the (lagged) patent stock as a measure of input in the same process. This aspect has been emphasised in the innovation literature since Arrow (1962) and has been specifically addressed by the literature about technological regimes (Nelson and Winter, 1982; Winter, 1984; Malerba and Orsenigo, 1993, 1997): the existing knowledge base represents the main input in the generation of new knowledge, i.e. the production of knowledge is a cumulative process.

The patent stock is calculated, as in Henderson and Cockburn (1996) using a perpetual inventory method and a 15% or, alternatively, a 20% depreciation rate[15]. The underlying assumption is that knowledge stock is subject to obsolescence and depreciation, so that investments in knowledge generating activity are continuously required to maintain, if not improving, a certain degree of innovative capacity.

The second factor which should allow to control for firms' propensity to patent, and the relevant heterogeneity between participants and non-participants, is size, as measured by the employment level.[16].

[15] Henderson and Cockburn (1996) employ a patent stock variable to measure knowledge capital accumulated by major R&D performing firms in the pharmaceutical industry, and find, over the period 1961-1988, a significant effect of such a variable on grants of "important" patents, i.e. patents that have been granted in two of the three main jurisdictions: Japan, Europe and the United States.

[16] Patent stock is also positively correlated with size, and may be considered an alternative control variable for size itself. However, the empirical literature investigating the relationship between size and innovation has not produced unambiguous evidence, or unambiguous interpretation of identified patterns.

Tables 10 and 11 report the Variance Analysis of the pre-entry differential (DIFF_PRE) and post-entry differential change (D_POST), respectively, when both are related to size[17]. This variable results significant (at 1% level) in the explanation of the pre-entry differential variance, whereas it is non significant for post-entry differential change. The overall result is robust to technological specification though statistical significance reduces. Therefore, size appears to be a relevant variable, to be included in the estimation in order to control for innovative propensity that is developed prior to entry. On the contrary, it does not seem that the level of employment explains much of the change in differential that takes place after the entry. In other words, there is no clear evidence that the increase in innovative gap between participants and non-participants depends on size.

The inclusion of size and other explanatory variables (e.g. group dummies) should allow to control for the prior-to- (or independent-from-) entry innovative level, so that the RJV variable captures the possible benefit of RJV participation, *given* firms' size and their level of priorly accumulated knowledge.

**Table 10**

**Pre-RJV patenting differential in relation with size**
**Analysis of variance (Balanced Sample, 1985-1996)**

**ALL** (N= 302)

| Dep Var DIFF_PRE | Partial SS | df | F | Pr>F |
|---|---|---|---|---|
| AV_EMP | 311.292 | 37 | 1.95 | 0.001 |
| Residual | 1134.724 | 263 | | |
| Total | 1446.016 | 300 | | |

**IT** (N=228)

| Dep Var DIFF_PRE | Partial SS | df | F | Pr>F |
|---|---|---|---|---|
| AV_EMP | 226.594 | 37 | 1.4 | 0.076 |
| Residual | 825.516 | 189 | | |
| Total | 1052.11 | 226 | | |

**MB** (N=78)

| Dep Var DIFF_PRE | Partial SS | df | F | Pr>F |
|---|---|---|---|---|
| AV_EMP | 160.193 | 24 | 1.57 | 0.0846 |
| Residual | 229.15 | 54 | | |
| Total | 389.343 | 78 | | |

**Table 11**

**Post-entry patenting differential change in relation with size**
**Analysis of variance (Balanced Sample, 1985-1996)**

**ALL** (N= 302)

| Dep Var DPOST | Partial SS | df | F | Pr>F |
|---|---|---|---|---|
| AV_EMP | 894.504 | 37 | 0.96 | 0.533 |
| Residual | 6595.842 | 263 | | |
| Total | 7490.346 | 300 | | |

**IT** (N=228)

| Dep Var DPOST | Partial SS | df | F | Pr>F |
|---|---|---|---|---|
| AV_EMP | 627.666 | 37 | 1.13 | 0.292 |
| Residual | 2835.345 | 189 | | |
| Total | 7490.346 | 226 | | |

**MB** (N=78)

| Dep Var DPOST | Partial SS | df | F | Pr>F |
|---|---|---|---|---|
| AV_EMP | 1010.194 | 24 | 0.75 | 0.775 |
| Residual | 3022.321 | 54 | | |
| Total | 4032.515 | 78 | | |

## 4.3 – Variables specification

[17] Size is here measured by the log of the average number of employees over the period 1992-1996.

Table 12 describes the variable used in the estimation, including all the factors whose significance is tested in alternative specifications.

**Table 12**

**Variables description**

| NAME | VARIABLE | DESCRIPTION | NOTE | SOURCE |
|---|---|---|---|---|
| $PAT_{it}$ | Patents | Number of firm i patent applications at EPO at time t | | EPO-CESPRI |
| $STOCK_{it}$ | Cumulated knowledge stock | Sum of firm i patent applications up to time t | | |
| $D15\_ST_{it}$ | 15% rate depreciated cumulated knowledge stock | Sum of firm i patent applications up to time t, 15% depreciation rate | Perpetual inventory method: $\Sigma_0^t\ PAT_{jk}\ (1+\delta)^{-k}$ $\delta$=0.15 | |
| $LAGST_{it}$ | One-period lagged 15% depreciated cumulated knowledge stock | Natural logarithm of $D15\_ST_{it-1}$ | | |
| $D20\_ST_{it}$ | 20% rate depreciated cumulated knowledge stock | Sum of firm i patent applications up to time t, 20% depreciation rate | Perpetual inventory method: $\Sigma_0^t\ PAT_{jk}\ (1+\delta)^{-k}$ $\delta$=0.20 | |
| $LAG2ST_{it}$ | One-period lagged 20% depreciated cumulated knowledge stock | Natural logarithm of $D20\_ST_{it-1}$ | | |
| $NOST_{it}$ | No-patent stock dummy variable | Dummy variable identifying adjusted patent stock observations | $NOST_{it}$=1 if $STOCK_{it}$=0 | |
| $LAGNOST_{it}$ | One-period lagged no-patent stock dummy variable | | | |
| $RJV_{it}$<br>$RJV\_1_{it}$<br>$RJV\_2_{it}$<br>$RJV\_3_{it}$ | RJV participation<br>One-period lagged RJV participation<br>Two-period lagged RJV participation<br>Three-period lagged RJV participation | Number of RJV projects firm i is involved at time t | | EU_RJV |
| $LEMP_{it}$ | Size | Natural logarithm of firm i number of employees at time t | | AMADEUS |
| $YEAR_t$ | Time dummy variable | Year index | | |
| $D_i$ | Industry dummy variable | Firm i 2 or 3-digit US SIC code | | AMADEUS |
| EU_EKA | EU-EUREKA dummy variable | Participation to both EU Framework projects and EUREKA Programs | | EU_RJV |
| G_P<br>G_S<br>G_I | Group dummy variables | Parent company<br>Subsidiary<br>Independent establishment | | Dun&Bradstreet |
| FR3<br>FR5<br>FR10 | Frequent_participant dummy variables | Dummy variables identifying firms whose project-yeras are equal or greater than the cutoff number (3, 5 or 10) | | EU_RJV |

Table 13 presents the Spearman correlation coefficients for the main quantitative variables, computed on the balanced samples. The correlation pattern offers a first idea about relevance of explanatory variables and possible multicollinearity. There appears to be a very strong positive relationship between current innovative flows and lagged stocks, whereas contemporaneous and lagged RJV participation seems to be more highly related to patents in the MB area rather than in the IT sector, except when considering RJV long lags.

The pattern of dependencies, in terms of ranking of observations, appears very “thick”. In the IT and sample the test of significance always strongly rejects the null hypothesis of independence. In the MB sample, strong rejection results only when comparing patent flows or stocks with a four period lagged measure of RJV participation.

**Table 13**

**Spearman correlation coefficients (Pr>|t|)**

| **IT Sample** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | PAT | D15_1 | D20_1 | EMP | RJV | RJV_1 | RJV_2 | RJV_3 | RJV_4 |
| PAT | 1.000 | | | | | | | | |
| D15_1 | 0.637 | 1.000 | | | | | | | |
| | *0.000* | | | | | | | | |
| D20_1 | 0.638 | 1.000 | 1.000 | | | | | | |
| | *0.000* | | | | | | | | |
| EMP | 0.214 | 0.274 | 0.274 | 1.000 | | | | | |
| | *0.000* | *0.000* | *0.000* | | | | | | |
| RJV | 0.108 | 0.136 | 0.135 | 0.287 | 1.000 | | | | |
| | *0.000* | *0.000* | *0.000* | *0.000* | | | | | |
| RJV_1 | 0.095 | 0.113 | 0.112 | 0.249 | 0.760 | 1.000 | | | |
| | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | | | | |
| RJV_2 | 0.081 | 0.094 | 0.093 | 0.224 | 0.636 | 0.771 | 1.000 | | |
| | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | | | |
| RJV_3 | 0.073 | 0.068 | 0.068 | 0.213 | 0.558 | 0.667 | 0.775 | 1.000 | |
| | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | | |
| RJV_4 | 0.061 | 0.047 | 0.047 | 0.184 | 0.494 | 0.589 | 0.672 | 0.749 | 1.000 |
| | *0.001* | *0.003* | *0.003* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | |
| **MB Sample** | | | | | | | | | |
| | PAT | D15_1 | D20_1 | EMP | RJV | RJV_1 | RJV_2 | RJV_3 | RJV_4 |
| PAT | 1.000 | | | | | | | | |
| D15_1 | 0.633 | 1.000 | | | | | | | |
| | *0.000* | | | | | | | | |
| D20_1 | 0.634 | 1.000 | 1.000 | | | | | | |
| | *0.000* | | | | | | | | |
| EMP | 0.149 | 0.238 | 0.238 | 1.000 | | | | | |
| | *0.000* | *0.000* | *0.000* | | | | | | |
| RJV | 0.140 | 0.149 | 0.149 | 0.281 | 1.000 | | | | |
| | *0.000* | *0.000* | *0.000* | *0.000* | | | | | |
| RJV_1 | 0.117 | 0.102 | 0.102 | 0.216 | 0.778 | 1.000 | | | |
| | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | | | | |
| RJV_2 | 0.091 | 0.077 | 0.077 | 0.196 | 0.605 | 0.802 | 1.000 | | |
| | *0.000* | *0.001* | *0.001* | *0.000* | *0.000* | *0.000* | | | |
| RJV_3 | 0.049 | 0.046 | 0.046 | 0.178 | 0.448 | 0.611 | 0.778 | 1.000 | |
| | *0.042* | *0.056* | *0.057* | *0.000* | *0.000* | *0.000* | *0.000* | | |
| RJV_4 | 0.000 | 0.017 | 0.017 | 0.167 | 0.320 | 0.427 | 0.553 | 0.738 | 1.000 |
| | *0.987* | *0.507* | *0.515* | *0.000* | *0.000* | *0.000* | *0.000* | *0.000* | |

The Spearman statistics give further support to the main findings outlined in the descriptive statistics section. We expect larger firms to be more innovative, both in terms of patents' current flow and cumulated stock, and more innovative firms to be more involved in RJV projects. Hence, we find further support to the idea that, when assessing RJV impact on current patenting activity, there is a strong need to control for both size and past innovative experience.

In the benchmark estimations we use the two-period lagged measure of RJV participation (RJV_2), i.e. we assume that RJV affiliation produces its effects with a short lag. As Branstetter and Sakakibara (1998) point out, a two-period lag appears a sensible estimation of the time required for the firm to absorb and process the knowledge which "spills out" of the consortia, and employ it for patenting purposes. In the case of Japanese RJVs, research personnel are typically rotated into R&D consortia and rotated back to parent firms on a two-year cycle, bringing a substantial amount of explicit and tacit knowledge about the new technology developed within consortia (Sakakibara, 1997).
It is sensible to assume that similar routines occur in firms which take part to European RJVs. Moreover, European projects have a medium-term time-scale, with most of the consortia lasting between two and three years and only very few, concentrated in EUREKA Programme, more than four years. Using a two-period lag amounts to assuming RJV capital and knowledge investments are being turned into patent applications mostly during the final stages of the projects, or in its immediate aftermath[18].

## 5- Evidence from IT and MB areas: econometric results

We employ the Poisson and Negative Binomial models proposed by Hausman et al. (1984) to estimate the relationship between patent applications and participation to R&D consortia, in the IT and MB area, controlling for size and cumulated innovative experience.

---

[18] We assessed the validity of this assumption and test the sensitivity of results to both shorter and longer lags. We do not include various lags of the RJV participation variable in the same equation in order to avoid problems related to multicollinearity. As we detect from Table 13, consortia affiliation is highly correlated over time. This high value of the correlation coefficient is partly due to the time horizon of projects, whose duration ranges from less than one year to five years.
Estimations may be also performed employing a contemporaneous RJV participation measure. Branstetter and Sakakibara (1998) use the contemporaneous RJV participation measure, as well as alternative lagged measures, and tackle the endogeneity problem employing Two Stages Least Squares. However, this specification implies the strong assumption that consortia have an immediate effect on firms' patenting activity. We prefer to assume that the full impact of RJV participation comes only after a lag and expect the patent stock variable (LAGST) to account for generic differences between firms' innovative ability, so that controlling for those differences (and

We expect these models to be more appropriate for the peculiar characteristics of our panel, which includes a relevant number of non-innovators. In fact, we do not force a continuous approximation on the phenomenon we want to analyse, patenting behaviour, whose distribution, in our dataset, is highly skewed, with zero outcome occurring with the highest frequency.

The basic specification includes, beside the lagged participation to European RJVs (RJV_2) the control variables specified in equation (6), that is the stock of cumulated knowledge (LAGST)[19], size (LEMP), and industry dummy variables.

We estimate the Random Effects, which allows to take into account also units for which the dependent variable is always equal to zero, i.e. non-innovators[20].

Tables 14-15 and 16-17 report the estimation outcomes for the two fields when employing a Poisson and a Negative Binomial model respectively. We estimate different specifications, without time dummies (Column 1), including time effects (Column 2), testing the significance of participation to both Framework and EUREKA Programmes (Column 3), substituting yearly participation with a frequent-participant dummy (Column 4), and assessing the relevance of firms' affiliation to a larger industrial group (Columns 5 and 6).

In all samples and specifications, the patent stock variable is highly significant, though the estimated coefficient is greater in the Negative Binomial model. This evidence supports the "technological cumulativeness" idea: firms with greatest cumulated knowledge stock, and recorded patenting attitude, are most likely to innovate today[21].

For the other variables, the Poisson estimates are generally more significant, as expected, than Negative Binomial ones, because of smaller standard errors.

The size variable (LEMP) is significant in the IT area, although the result is not robust to a shift from the Poisson to the Negative Binomial model. Nevertheless, this outcome appears interesting, when considering that part of the size effect is captured by the knowledge stock

---

controlling for size and industry), the lagged RJV participation variable mostly captures the effects that are due to different levels of involvement in European R&D consortia.

[19] We measure the knowledge capital stock in logarithmic terms, adding 1 to zero-value observations in order to get a definite logarithmic value, and including an adjustment dummy variable (LAGNOST), which is equal to 1 if the stock of patents at time t-1 is equal to zero. By taking the natural log of the (depreciated) patent stock, we still implicitly assume there are diminishing effects of cumulated knowledge on patent activity.
We have performed estimates employing a non-logarithmic stock variable. However, this variable captures most of the size effect (LEMP is always non significant) and the performance of the model, in terms of log-likelihood, remarkably worsen.

[20] FE specifications of these non-linear models are not suit to analyse our phenomenon, since units for which the dependent variable is equal to zero do not contribute to maximising the likelihood function.

[21] The use of a 20% depreciation rate in the calculation of the stock variable leads to higher log-likelihood and greater coefficient, as if faster depreciation would better explain the relationship between current patenting activity and cumulated innovative experience.

variable. Hence, in the IT field, we find that, even when controlling for the past innovative experience, size is still significant in explaining current patenting activity.

When stepping into technological field analysis, we find it is in the IT area that sector classification is significantly associated with different patenting levels, mostly because it is in this field that the largest share of relatively non patenting service firms join RJVs together with R&D-intensive component producers[22].

In the MB sample, we detect significance only in the Negative Binomial specification, and in relation to the dummy variable ENGMAN, which groups Engineering & Management Services, in particular Research and Testing Services, for which a relatively high propensity to innovate is to be expected[23].

Time effects seem to play a relevant role in both technological fields[24], though the effect appears stronger in the MB area. This finding is consistent with the negative trend of patenting in our sample registered over the period 1992-1996.

As far as the lagged RJV measure is concerned, when we do not include time dummies, its coefficient tends to be negative although it does not result significant. In the specification with time effects, the coefficient turns positive in all samples, though its significance varies over fields. In this case the distinction between technological areas is all the more meaningful. In fact, differences between participants in IT and MB RJVs are rather robust to variable and model specifications.

For the IT firms, RJV participation never appears to result in a significant contribution to the patenting level. In terms of the self-selection analysis presented in section 4, it appears that the average higher level of innovativeness of RJV participants is mostly explained by an intense patenting activity *prior* to entry, that is much related to size and sector, but no significant impact of RJV affiliation on this "in-built" innovative attitude can be detected.

On the contrary, in the MB area, the RJV variable coefficient is positive and significant in most of the proposed estimations. When using a Poisson model, significance of the RJV variable is not sensitive to the choice of a shorter (RJV_1) or longer (RJV_3) lag. Negative Binomial estimates are indeed sensitive to the lag specification: RJV affiliation is significant when measured with a lag of at least three years. Again, the Negative Binomial estimated coefficient is slightly smaller than the Poisson one, whereas standard error is slightly higher.

---

[22] This result is not extremely robust to the use of a Negative Binomial specification, although, even in this case, the estimates point at some significant sectoral effect. The sectoral dummy ENGMAN is always significant, whereas Computer and Communication sector dummies signal a higher propensity to patent when employing the frequent participant variable in place of yearly RJV affiliation.

[23] In the overall sample the average number of patents over the period 1992-1996 of ENGMAN firms is equal to 13.6, whereas the sample mean is equal to 10.3.

[24] For the IT field, time dummies are generally non-significant when using a Negative Binomial specification.

**Table 14**
**Balanced IT Sample (N=508)**
**Poisson Random Effects**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dep. Var. | Coef. | Coef. | Coef. | Coef. | Coef. | Coef. |
| PAT | *z-value* | *z-value* | *z-value* | *z-value* | *z-value* | *z-value* |
| LAGST | 0.640 *** | 0.769 *** | 0.716 | 0.755 *** | 0.755 *** | 0.501 *** |
| | *11.507* | *14.501* | *11.994* | *12.816* | *13.710* | *13.312* |
| LEMP | 0.086 *** | 0.063 ** | 0.490 | 0.060 * | 0.060 * | - |
| | *3.925* | *2.816* | *2.020* | *2.664* | *2.662* | |
| RJV_2 | -0.021 | 0.003 | -0.034 | - | 0.002 | -0.028 ** |
| | *-1.788* | *0.207* | *-0.523* | | *0.148* | *-2.782* |
| CHEM | (dropped) | (dropped) | (dropped) | (dropped) | (dropped) | (dropped) |
| DRUGS_P | 1.063 | 1.036 | 1.021 | 1.044 | 1.070 * | 1.351 |
| | *2.450* | *2.512* | *2.435* | *2.504* | *2.559* | *1.952* |
| COMP_P | 0.531 | 0.497 | 0.277 | 0.464 | 0.546 | 0.507 |
| | *1.829* | *1.757* | *0.907* | *1.587* | *1.889* | *1.422* |
| E_EQUIP | 0.986 * | 0.996 ** | 0.972 ** | 1.000 | 1.002 ** | 1.421 * |
| | *2.761* | *2.946* | *2.836* | *2.926* | *2.910* | *2.645* |
| COMM_P | 0.831 ** | 0.800 ** | 0.781 * | 0.783 | 0.814 ** | 1.194 ** |
| | *2.796* | *2.794* | *2.713* | *2.693* | *2.809* | *3.024* |
| P_EQUIP | 0.801 | 0.860 | 0.711 | 0.825 | 0.839 | 0.836 |
| | *2.263* | *2.538* | *2.008* | *2.362* | *2.386* | *1.607* |
| ENGMAN | 0.847 | 0.823 | 0.753 | 0.800 | 0.912 * | 1.102 * |
| | *2.502* | *2.488* | *2.246* | *2.382* | *2.656* | *2.425* |
| FOOD | (dropped) | (dropped) | (dropped) | (dropped) | (dropped) | (dropped) |
| LAGNOST | -2.962 *** | -2.806 *** | -2.841 *** | -2.808 | -2.799 *** | -1.333 *** |
| | *-12.630* | *-12.302* | *-12.275* | *-12.257* | *-12.216* | *-6.587* |
| CONS | -2.153 *** | -2.163 *** | -1.994 *** | -2.151 *** | -2.302 *** | -2.939 *** |
| | *-7.039* | *-7.186* | *-6.341* | *-7.076* | *-6.868* | *-7.480* |
| Y92 | - | 0.240 *** | 0.233 *** | 0.238 *** | 0.238 *** | 0.026 |
| | | *3.389* | *3.295* | *3.362* | *3.364* | *0.473* |
| Y94 | - | -0.037 | -0.013 | -0.029 | -0.033 | -0.070 |
| | | *-0.494* | *-0.169* | *-0.418* | *-0.440* | *-1.124* |
| Y95 | - | -0.177 | -0.156 | -0.171 *** | -0.173 | -0.199 ** |
| | | *-2.374* | -2.065 | *-2.345* | *-2.310* | *-3.079* |
| Y96 | - | -0.255 *** | -0.226 ** | -0.246 *** | -0.249 ** | -0.236 *** |
| | | *-3.233* | *-2.787* | *-3.257* | *-3.150* | *-3.454* |
| EU_EKA | - | - | 0.668 | | - | - |
| | | | *2.068* | - | | |
| R2_EUEKA | | | 0.032 | | | |
| | | | *0.510* | | | |
| FR3 | | - | - | 0.120 | - | - |
| | | | | *0.558* | | |
| G_P | | - | - | - | 0.315 | 1.073 |
| | | | | | *1.105* | *2.548* |
| G_S | - | - | - | - | 0.185 | 0.834 ** |
| | | | | | *0.089* | *3.035* |
| Log likelihood | -998.605 | -977.276 | -974.790 | -977.137 | -976.581 | -1867.380 |
| Wald chi2 (Pr>chi2) | 523.74 (0.000) | 654.56 (0.000) | 636.99 (0.000) | 647.93 (0.000) | 652.77 (0.000) | 351.65 (0.000) |
| LR test -pooled- (Pr>chi2) | 340.08 (0.000) | 362.19 (0.000) | 362.68 (0.000) | 348.61 (0.000) | 360.28 (0.000) | 559.28 (0.000) |

* significant at 10% level

** significant at 5% level

*** significant at 1% level

**Table 15**
**Balanced MB Sample (N=191)**
**Poisson Random Effects**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dep. Var. | Coef. | Coef. | Coef. | Coef. | Coef. | Coef. |
| PAT | *z-value* | *z-value* | *z-value* | *z-value* | *z-value* | *z-value* |
| LAGST | 0.354 | 0.646 *** | 0.670 *** | 0.672 *** | 0.617 *** | 0.372 *** |
| | *2.117* | *4.735* | *4.558* | *5.011* | *4.352* | *5.761* |
| LEMP | 0.282 * | 0.232 | 0.171 | 0.214 | 0.212 | - |
| | *2.677* | *2.357* | *1.713* | *2.226* | *2.092* | |
| RJV_2 | -0.042 | 0.151 *** | -0.339 | - | 0.148 *** | 0.112 *** |
| | *-1.719* | *3.948* | *-1.937* | | *3.881* | *4.068* |
| CHEM | 0.467 | 0.532 | 0.562 | 0.373 | 0.407 | 0.381 |
| | *0.572* | *0.745* | *0.794* | *0.538* | *0.551* | *0.459* |
| DRUGS_P | 1.358 | 0.791 | 0.812 | 0.570 | 0.843 | 1.667 * |
| | *2.037* | *1.397* | *1.436* | *1.025* | *1.428* | *2.701* |
| COMP_P | (dropped) | (dropped) | (dropped) | (dropped) | (dropped) | (dropped) |
| E_EQUIP | (dropped) | (dropped) | (dropped) | (dropped) | (dropped) | (dropped) |
| COMM_P | (dropped) | (dropped) | (dropped) | (dropped) | (dropped) | (dropped) |
| P_EQUIP | - | - | - | - | - | - |
| ENGMAN | 1.788 | 1.334 | 1.249 | 1.176 | 1.496 | 2.084 |
| | *2.392* | *2.048* | *1.908* | *1.883* | *2.048* | *2.414* |
| FOOD | -0.254 | -0.112 | -0.003 | -0.210 | -0.132 | -0.234 |
| | *-0.331* | *-0.164* | *-0.004* | *-0.308* | *-0.191* | *-0.352* |
| LAGNOST | -2.765 *** | -2.628 *** | -2.528 *** | -2.516 *** | -2.670 *** | -1.484 *** |
| | *-5.401* | *-5.961* | *-5.737* | *-5.834* | *-5.984* | *-4.275* |
| CONS | -3.677 *** | -3.278 *** | -3.003 *** | -3.409 *** | -3.292 *** | -2.869 *** |
| | *-4.278* | *-4.126* | *-3.766* | *-4.457* | *-3.965* | *-4.701* |
| Y92 | - | 0.393 *** | 0.380 *** | 0.311 * | 0.383 *** | 0.136 |
| | | *3.441* | *3.305* | *2.762* | *3.334* | *1.552* |
| Y94 | - | -0.137 | -0.118 | -0.059 | -0.132 | -0.215 |
| | | *-1.192* | *-1.029* | *-0.523* | *-1.151* | *-2.143* |
| Y95 | - | -0.680 *** | -0.499 ** | -0.298 | -0.667 *** | -0.633 *** |
| | | *-4.392* | *-3.052* | *-2.480* | *-4.284* | *-4.823* |
| Y96 | - | -0.944 *** | -0.755 *** | -0.447 *** | -0.929 ** | -0.864 ** |
| | | *-5.189* | *-3.974* | *-3.515* | *-5.079* | *-5.775* |
| EU_EKA | - | - | 0.323 | | - | - |
| | | | *0.607* | | | |
| R2_EUEKA | | | 0.458 ** | | | |
| | | | *2.832* | | | |
| FR3 | | - | - | 0.599 | - | - |
| | | | | *1.715* | | |
| G_P | | - | - | - | 0.491 | 1.249 |
| | | | | | *0.845* | *1.777* |
| G_S | - | - | - | - | 0.168 | 0.462 |
| | | | | | *0.372* | *0.878* |
| Log likelihood | -331.784 | -306.133 | -301.284 | -312.703 | -305.765 | -582.372 |
| Wald chi2 (Pr>chi2) | 97.70 (0.000) | 179.95(0.000 ) | 189.36 (0.000) | 185.37 (0.000) | 177.21 (0.000) | 169.98 (0.000) |
| LR test -pooled- (Pr>chi2) | 89.32 (0.000) | 57.60 (0.000) | 61.76 (0.000) | 29.52 (0.000) | 49.10 (0.000) | 146.46 (0.000) |

* significant at 10% level

** significant at 5% level

*** significant at 1% level

**Table 16**
**Balanced IT Sample (N=508)**
**Negative Binomial Random Effects**

| | (1) | | | (2) | | | (3) | | | (4) | | | (5) | | | (6) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coef. | (std err) | | Coef. | (std err) | | Coef. | (std err) | | Coef. | (std err) | | Coef. | (std err) | | Coef. | (std err) | |
| Dep. Var.<br>PAT | | | | | | | | | | | | | | | | | | |
| LAGST | 0.835 | (0.049) | *** | 0.866 | (0.050) | *** | 0.875 | (0.055) | *** | 0.911 | (0.051) | *** | 0.865 | (0.050) | *** | 0.702 | (0.048) | *** |
| LEMP | 0.076 | (0.035) | | 0.051 | (0.034) | | 0.048 | (0.037) | | 0.077 | (0.036) | | 0.048 | (0.034) | | - | | |
| RJV_2 | -0.029 | (0.017) | | -0.003 | (0.016) | | -0.195 | (0.110) | | | | | -0.004 | (0.016) | | -0.023 | (0.018) | |
| CHEM | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | |
| DRUGS_P | 0.956 | (0.353) | * | 0.865 | (0.372) | | 0.837 | (0.367) | | 0.879 | (0.360) | | 0.873 | (0.371) | | 0.370 | (0.328) | |
| COMP_P | 0.671 | (0.267) | | 0.632 | (0.277) | | 0.618 | (0.285) | | 0.728 | (0.270) | * | 0.630 | (0.279) | | 0.274 | (0.229) | |
| E_EQUIP | 0.642 | (0.303) | | 0.675 | (0.317) | | 0.657 | (0.312) | | 0.675 | (0.306) | | 0.664 | (0.317) | | 0.475 | (0.274) | |
| COMM_P | 0.568 | (0.262) | | 0.617 | (0.272) | | 0.577 | (0.269) | | 0.703 | (0.264) | * | 0.602 | (0.273) | | 0.466 | (0.226) | |
| P_EQUIP | 0.469 | (0.315) | | 0.575 | (0.328) | | 0.590 | (0.325) | | 0.691 | (0.318) | | 0.530 | (0.336) | | 0.147 | (0.287) | |
| ENGMAN | 0.945 | (0.300) | ** | 0.893 | (0.312) | ** | 0.886 | (0.310) | ** | 0.929 | (0.306) | ** | 0.900 | (0.318) | ** | 0.572 | (0.265) | |
| FOOD | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | | | | |
| LAGNOST | -2.812 | (0.248) | *** | -2.813 | (0.249) | *** | -2.809 | (0.250) | *** | -2.782 | (0.248) | *** | -2.811 | (0.249) | *** | -2.768 | (0.173) | *** |
| Y92 | - | | | 0.253 | (0.100) | | 0.255 | (0.100) | | 0.246 | (0.102) | | 0.250 | (0.100) | | -0.020 | (0.090) | |
| Y94 | - | | | -0.028 | (0.106) | | 0.011 | (0.108) | | -0.038 | (0.102) | | -0.029 | (0.106) | | -0.250 | (0.100) | |
| Y95 | - | | | -0.173 | (0.106) | | -0.148 | (0.108) | | -0.178 | (0.105) | | -0.175 | (0.106) | | -0.397 | (0.100) | *** |
| Y96 | - | | | -0.253 | (0.111) | | -0.212 | (0.112) | | -0.252 | (0.106) | | -0.257 | (0.111) | | -0.431 | (0.104) | *** |
| R2_EUEK | - | | | - | | | 0.192 | (0.109) | | - | | | - | | | - | | |
| EU_EKA | - | | | - | | | -0.057 | (0.265) | | - | | | - | | | - | | |
| FR3 | - | | | - | | | - | | | -0.452 | (0.181) | | - | | | - | | |
| G_P | - | | | - | | | - | | | - | | | 0.165 | (0.258) | | 0.366 | (0.218) | |
| G_S | - | | | - | | | - | | | - | | | 0.024 | (0.192) | | 0.121 | (0.162) | |
| CONS | -2.301 | (0.370) | *** | -1.969 | 0.386 | *** | -1.962 | (0.414) | *** | -2.206 | (0.390) | *** | -1.979 | (0.407) | *** | -1.224 | (0.275) | *** |
| Log likelihood | -921.099 | | | -909.883 | | | -908.020 | | | -906.890 | | | -909.645 | | | -1684.19 | | |
| Wald chi2 (Pr>chi2) | 878.180 | (0.000) | | 870.830 | (0.000) | | 894.840 | (0.000) | | 922.930 | (0.000) | | 875.670 | (0.000) | | 832.27 | (0.000) | |
| LR test -pooled- (Pr>chi2) | 49.640 | (0.000) | | 65.130 | (0.000) | | 63.410 | (0.000) | | 61.780 | (0.000) | | 64.500 | (0.000) | | 82.38 | (0.000) | |

* significant at 10% level

** significant at 5% level

*** significant at 1% level

**Table 17**
**Balanced MB Sample (N=191)**
**Negative Binomial Random Effects**

| | (1) | | | (2) | | | (3) | | | (4) | | | (5) | | | (6) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coef. | (std err) | | Coef. | (std err) | | Coef. | (std err) | | Coef. | (std err) | | Coef. | (std err) | | Coef. | (std err) | |
| Dep. Var. | | | | | | | | | | | | | | | | | | |
| PAT | | | | | | | | | | | | | | | | | | |
| LAGST | 0.897 | (0.109) | *** | 0.844 | (0.129) | *** | 0.874 | (0.131) | *** | 0.950 | (0.059) | *** | 0.807 | (0.123) | *** | 0.573 | (0.084) | *** |
| LEMP | 0.001 | (0.101) | | 0.072 | (0.117) | | 0.072 | (0.110) | | 0.084 | (0.078) | | -0.028 | (0.111) | | - | - | |
| RJV_2 | -0.087 | (0.045) | | 0.126 | (0.051) | | -0.390 | (0.208) | | - | | | 0.112 | (0.049) | | 0.078 | (0.040) | |
| CHEM | 1.451 | (0.641) | | 1.288 | (0.820) | | 1.178 | (0.768) | | 1.906 | (0.552) | *** | 0.978 | (0.735) | | 0.853 | (0.601) | |
| DRUGS_P | 1.167 | (0.517) | | 1.034 | (0.667) | | 0.988 | (0.628) | | 1.325 | (0.438) | *** | 1.116 | (0.573) | | 1.316 | (0.441) | ** |
| COMP_P | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | |
| E_EQUIP | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | |
| COMM_P | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | |
| P_EQUIP | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | | (dropped) | | |
| ENGMAN | 1.752 | (0.575) | ** | 1.704 | (0.654) | * | 1.633 | (0.664) | | 1.899 | (0.529) | *** | 2.184 | (0.673) | *** | 2.350 | (0.536) | *** |
| FOOD | 1.093 | (0.636) | | 0.743 | (0.836) | | 0.694 | (0.776) | | 1.130 | (0.568) | | 0.698 | (0.715) | | 0.404 | (0.543) | |
| LAGNOST | -2.727 | (0.459) | *** | -2.861 | (0.463) | *** | -2.783 | (0.469) | *** | -2.448 | (0.421) | *** | -2.978 | (0.473) | *** | -2.898 | (0.338) | *** |
| Y92 | - | | | 0.383 | (0.149) | * | 0.389 | (0.141) | * | 0.329 | (0.157) | | 0.335 | (0.148) | | 0.038 | (0.125) | |
| Y94 | - | | | -0.097 | (0.148) | | -0.090 | (0.140) | | -0.024 | (0.158) | | -0.070 | (0.147) | | -0.270 | (0.134) | |
| Y95 | - | | | -0.629 | (0.201) | ** | -0.472 | (0.195) | | -0.316 | (0.168) | | -0.546 | (0.198) | * | -0.656 | (0.174) | *** |
| Y96 | - | | | -0.909 | (0.228) | *** | -0.732 | (0.224) | *** | -0.511 | (0.176) | ** | -0.821 | (0.225) | *** | -0.913 | (0.195) | *** |
| R2_EUEK | - | | | - | | | 0.492 | (0.194) | | - | | | - | | | - | | |
| EU_EKA | - | | | - | | | 0.063 | (0.401) | | - | | | - | | | - | | |
| FR3 | - | | | - | | | - | | | 0.639 | (0.145) | *** | - | | | - | | |
| G_P | - | | | - | | | - | | | - | | | 1.217 | (0.535) | | 1.221 | (0.444) | * |
| G_S | - | | | - | | | - | | | - | | | 0.822 | (0.450) | | 0.613 | (0.344) | |
| CONS | -2.747 | (0.918) | ** | -2.128 | (0.998) | | -1.900 | (0.963) | | -3.850 | (0.713) | | -2.111 | (0.922) | | -1.690 | (0.493) | *** |
| Log likelihood | -306.901 | | | -293.119 | | | -289.343 | | | -293.611 | | | -290.017 | | | -544.603 | | |
| Wald chi2 (Pr>chi2) | 285.650 | (0.000) | | 294.640 | (0.000) | | 246.410 | (0.000) | | 1119.970 | (0.000) | | 253.120 | (0.000) | | 287.67 | (0.000) | |
| LR test -pooled- (Pr>chi2) | 14.850 | (0.000) | | 7.540 | (0.006) | | 12.090 | (0.000) | | 0.000 | (0.993) | | 7.82 | (0.005) | | 28.17 | (0.000) | |

* significant at 10% level

** significant at 5% level

*** significant at 1% level

Table 18 reports the values of incidence rate ratios (IRR) for the patent stock, size and RJV variable, when time effects are taken into account[25].

In the MB field, an increase by one project per year at time t-3 is associated with an increase in the number of patent applications at time t which ranges from 22% (Negative Binomial) to 25% (Poisson). LAGST has by far the most relevant effect for all firms, though the IRR is higher in the IT area, where also an increase in the employment level produces significant changes in patenting.

**Table 18**

**Balanced samples- Poisson and Negative Binomial Random Effects**

**Incidence Rate Ratios**

| | IT | | MB | |
|---|---|---|---|---|
| | Poisson | Neg Bin | Poisson | Neg Bin |
| LAGST[a] | 2.158*** | 2.378*** | 1.907*** | 2.326*** |
| | (14.501) | (17.430) | (4.735) | (6.568) |
| LEMP[a] | 1.065** | 1.052 | 1.261 | 1.074 |
| | (2.816) | (1.510) | (2.357) | (0.614) |
| RJV_2 | 1.002 | 0.996 | 1.162*** | 1.134 |
| | (0.207) | (-0.210) | (3.948) | (2.458) |
| RJV_3 | 0.993 | 0.990 | 1.249*** | 1.224** |
| | (-0.451) | (-0.480) | (4.352) | (3.042) |

Note:
z-values in parenthesis
*significant at 10% level
**significant at 5% level
***significant at 1% level

a: IRR referred to the specification including RJV_2

In column 3 of the Tables above presented, we test the significance of the dummy variable EU_EKA, entered as both an additional intercept term and a multiplicative term for RJV_2. No effect is detected for the IT field, whereas in the Poisson estimation of the MB sample the response effect of diversified participation (slope coefficient) is significant, although the coefficient of the RJV term becomes non-significant.

In the Poisson model, contrarily to the RJV variable, the results about the frequent-participant dummy (FR3, Column 4) are very similar across samples, and are not sensitive to the choice of

[25] The values reported in Table 18 are those of the exponentiated coefficient $e^b$, which, for the Poisson model, has the interpretation of incidence rate ratios.
The incidence rate ratio for a one unit change in variable $X_j$ is equal to:

$$\frac{e^{\beta_1 X_1 + \ldots + \beta_j (X_j + 1) + \ldots + \beta_k X_k}}{e^{\beta_1 X_1 + \ldots + \beta_j X_j + \ldots + \beta_k X_k}} = e^{\beta_j}$$

different cut-off numbers (five or ten). The quality of being a frequent (or a non-frequent) participant does not significantly contribute to explaining innovative ability.
The result changes for the MB area, when adopting a Negative Binomial specification: a relatively high frequency of participation to shared-cost research is significantly associated with patent output.

When the analysis is extended to group dummy variables (Column 5), no significant effect is detected: neither in the IT area, nor in the MB area the patenting behaviour of firms appear to be highly related to their legal status, in terms of affiliation to a holding group[26].
However, we need to take into account the strong correlation between employment level and group affiliation. If we omit the size variable (Column 6), its effect is more substantially captured by the group dummies[27]. The Subsidiary status tends to be associated with higher patenting activity in the IT field, whereas the Parent dummy is significant (10% level) in the MB area, where we detect a non-significant impact of the employment level. We may relate this result to the fact that, in the MB area, parent companies are not necessarily the largest firms, but are the catalyst for the group patenting activity, whereas in the IT area patenting activities are more widely spread among group members. The interpretation however suffers from ambiguity, in the sense that the result may reflect patent application policies within groups, rather than effective centralisation (or de-centralisation) of research activities.

In Table 19, we present an alternative (Poisson) specification aimed at evaluating the effect of cumulated RJV experience rather than of yearly RJV participation. We substitute the measure of projects per year with the count of cumulated project-years up to time t, the “stock of RJV participation” (SRJV). However, we still assume the effects of RJV experience shows up with some lag and employ a two-period lagged measure of the RJV stock, SRJV_2.
Results confirm the existence of technology field differences emphasised when referring to RJV participation flows. The variable is non significant in the IT sample, but is greatly significant in the MB area (1% level), where the estimated IRR is equal, to 1.06. Hence, the estimated effect of an increase in the stock of participation is smaller than the estimated effect of an increase in the number of projects per year.

[26] We have chosen 1992, the first year of our analysis, as the reference year for classifying firms as Parent (dummy variable g_p), Subsidiary (g_s) or Independent (g_i). Firms who are not classified by Dun & Branstreet (1992) as parent or subsidiary are labelled “independent”. However, we expect this third group to be composed of firms that are truly independent establishments and firms who may be part of a larger group but, do not enter, for some reason, Dun & Branstreet (1992) files.
[27] In addition, the inclusion of group dummies strongly reduces the coefficient of LAGST, the patent stock.

**Table 19**
**Cumulated RJV participation**
**Poisson Random Effects - Balanced Samples**

| | IT (N=508) | | MB (N=191) | |
|---|---|---|---|---|
| Dep. Var. | Coef. | | Coef. | |
| PAT | *z-value* | | *z-value* | |
| LAGST | 0.769 | *** | 0.657 | *** |
| | *14.479* | | *4.815* | |
| LEMP | 0.063 | ** | 0.233 | |
| | *2.808* | | *2.349* | |
| SRJV_2 | 0.000 | | 0.061 | *** |
| | *-0.071* | | *4.549* | |
| CHEM | (dropped) | | 0.530 | |
| | | | *0.738* | |
| DRUGS_P | 1.035 | | 0.782 | |
| | *2.509* | | *1.373* | |
| COMP_P | 0.501 | | (dropped) | |
| | *1.771* | | | |
| E_EQUIP | 0.995 | ** | (dropped) | |
| | *2.944* | | | |
| COMM_P | 0.800 | ** | (dropped) | |
| | *2.795* | | | |
| P_EQUIP | 0.860 | | (dropped) | |
| | *2.538* | | | |
| ENGMAN | 0.824 | | 1.351 | |
| | *2.491* | | *2.058* | |
| FOOD | (dropped) | | -0.121 | |
| | | | *-0.176* | |
| LAGNOST | -2.807 | *** | -2.623 | *** |
| | *-12.308* | | *-5.912* | |
| CONS | -2.164 | *** | -3.292 | |
| | *-7.181* | | *-4.109* | |
| Y92 | 0.240 | *** | 0.383 | *** |
| | *3.377* | | *3.366* | |
| Y94 | -0.031 | | -0.154 | |
| | *-0.432* | | *-1.334* | |
| Y95 | -0.172 | | -0.616 | *** |
| | *-2.269* | | *-4.432* | |
| Y96 | -0.248 | ** | -1.066 | *** |
| Log likelihood | -977.294 | | -303.371 | |
| Wald chi2 (Pr>chi2) | 653.79 (0.000) | | 181.28 (0.000) | |
| LR test -pooled- (Pr>chi2) | 354.01 (0.000) | | 61.90 (0.000) | |

* significant at 10% level
** significant at 5% level
*** significant at 1% level

The Poisson Random Effects estimation of the imputed dataset (Tables 20 and 21) confirms the balanced sample findings: patent stock and employment level are highly significant, the (aggregate) Drugs, (aggregate) Communication and Electrical Equipment sector dummies are positively significant, time effects are greatly significant and point to a steep decline in patent applications over time[28], whereas the RJV participation variable is non significant.

However, important differences with respect to the balanced estimates and between IT and MB fields appear when we introduce variables that, in different ways, are related to size and investment capabilities.

In the IT area, in contrast with the outcomes above commented, the frequent-participant dummy variable exhibits a statistically significant association with patenting, when selecting a rather high threshold: FR3 is non significant, whereas FR5 and FR10 are significant at 5% and 1% level respectively. When the sample is enlarged to include firms which are more heterogeneous in terms of size and RJV participation, and, on average, less innovative, the quality of being a frequent participant is significantly associated with a more intense patenting activity.

The significant EU_EKA dummy (intercept term) gives similar, though not identical, information, since firms which take part to both EU Framework and EUREKA programs are highly intensive RJV participants[29].

Another variable which is related to both frequency of participation and size is the group dummy variable, since most of RJV participants and biggest firms are part of large industrial groups. The G_S (Subsidiary dummy) is highly significant in the imputed estimates[30], even when including the employment variable.

The different results about significance of these variables, which are somehow related to size and investment resources, suggest that the "size component" becomes more relevant when the estimated sample comprises firms which are on average smaller and more heterogeneous.

In the MB field, the employment variable becomes highly significant. However, contrarily to the IT evidence, in the MB estimates, the "size/resource" variables, such as the frequent-participant, EU_EKA and group dummies are never significant, whatever specification we employ[31].

In the MB field, there still appears to be more convincing evidence of a positive effect of RJV participation than in the IT area, but statistical significance does not emerge when we consider our "standard" two-period lagged measure of RJV affiliation (p-value=0.025). Rather, a strong positive association with patenting (1% significance level) results when we employ a three-period lagged measure. The coefficient of RJV_3 is close to the one estimated for RJV_2 in the balanced sample: 0.129, which corresponds to an IRR equal to 1.138, i.e. one more RJV project at time t-3 is

[28] Taking 1993 as the reference year (omitted dummy) Y92, is positive, whereas Y94, Y95 and Y96 are negative.

[29] The average number of project-years for this subgroup of firms, in the imputed dataset, is equal to 18.57, against 3.33 project-years in the overall sample, and 6.29 project-years for all RJV participants.

[30] Significance is defined in relation to the omitted dummy variable G_I, which identifies independent establishments.

[31] The slope term R2_EUEKA is significant, but its inclusion produces a proportional negative effect on the participation coefficient.

associated, *ceteris paribus*, with an increase in patenting by nearly 14%. The result is robust to the use of longer lags. Indeed, the estimated coefficient increases with the lag.

The significance of long lags can be related to the structure of RJVs in the MB fields, that exhibit an average duration which is remarkably longer than in the IT field. The great majority (nearly 90%) of European R&D consortia in this technological area lasts more than 30 months, with 20% of the projects lasting more than four years.

Another explanation relates to the early development of RJVs in the MB area, for which the largest number of projects has been promoted in the second half of the 80s, within the EUREKA Program, specifically aimed at fostering close-to-market technological applications. The timing of the programmes combines with the peculiar technological evolution of the field, especially of biotechnologies.

In this sense, when the lag is very high there is a risk that the RJV variable does not really represent the intensity of RJV participation, but rather picks up the "pioneer" firms, which entered the first European R&D projects in the early years of the EUREKA Program, and already had a high innovative propensity. The issue is similar to the self-selection one, though it is more specific and concerns a clearly identified group of firms that could play the role of outliers. We have tried to assess the presence of this effect by flagging the greatly innovative firms, which entered MB RJVs in 1989, with a dummy variable, PIONEER, and re-estimating the basic specification with the three-period lagged measure of participation. RJV_3 is still significant at 1% level, whereas the pioneer dummy is non-significant[32].

As far as the IT area is concerned, the larger size of the imputed dataset allows to further break up the sample into small subgroups and investigate whether the effects of RJV participation appears less negligible when we focus on innovative sectors separately. Table 22 presents the estimates related to the Office, Computers & Accounting Equipment (US SIC code 357), Communication Equipment (365, 366, 367), Professional & Scientific Equipment (38 excl.384) subsamples. The outcome is not substantially different from the result on the overall IT sample. Indeed, the (non significant) RJV_2 coefficient tends to be negative, except for Communication Equipment. For this sector, we also estimate a positive RJV effect, significant at 5% level, when specifying a Poisson model. However, the result is sensitive to the choice of the statistical model. When adding an additional source of firm- and time- specific variance, in the Negative Binomial model, the coefficient is non-significant, whatever is the specified lag.

[32] The result is robust to the specification of a longer lag. Moreover, we obtain a similar outcome when estimating the "pioneer effect" in the balanced sample.

**Table 20**
**Imputed IT Sample (N=1043)**
**Poisson Random Effects**

| | (1) | | (2) | | (3) | | (4) | | (5) | |
|---|---|---|---|---|---|---|---|---|---|---|
| Dep. Var. | Coef. | *z-value* | Coef. | *z-value* | Coef. | *z-value* | Coef. | *z-value* | Coef. | *z-value* |
| PAT | | | | | | | | | | |
| LAGST | 0.449 | *11.356* *** | 0.646 | *13.170* *** | 0.657 | *15.821* *** | 0.617 | *12.276* *** | 0.612 | *12.477* *** |
| LEMP | 0.079 | *3.491* *** | 0.078 | *3.425* *** | 0.060 | *2.718* ** | 0.070 | *3.070* ** | 0.056 | *2.421* |
| RJV_2 | -0.015 | *-1.343* | 0.009 | *0.744* | -0.006 | *-0.127* | - | | 0.003 | *0.246* |
| CHEM | - | | - | | - | | - | | - | |
| DRUGS_P | 1.593 | *1.931* | 1.427 | *2.013* | 0.864 | *2.112* | 1.514 | *2.109* | 1.496 | *2.116* |
| COMP_P | 0.243 | *0.711* | 0.119 | *0.395* | 0.103 | *0.458* | -0.004 | *-0.014* | 0.163 | *0.511* |
| E_EQUIP | 1.892 | *3.332* *** | 1.681 | *3.439* *** | 1.057 | *3.686* *** | 1.679 | *3.388* *** | 1.763 | *3.627* *** |
| COMM_P | 1.111 | *2.909* ** | 0.922 | *2.758* * | 0.675 | *2.972* ** | 0.811 | *2.377* | 0.886 | *2.635* * |
| P_EQUIP | 0.653 | *1.254* | 0.639 | *1.409* | 0.263 | *0.857* | 0.482 | *1.029* | 0.528 | *1.130* |
| ENGMAN | 0.786 | *1.795* | 0.697 | *1.818* | 0.444 | *1.693* | 0.637 | *1.637* | 0.908 | *2.275* |
| FOOD | - | | - | | - | | - | | - | |
| LAGNOST | -0.181 | *-0.849* | -0.176 | *-0.886* | -2.327 | *-13.383* *** | -0.176 | *-0.879* | -0.122 | *-0.614* |
| CONS | -2.958 | *-8.110* *** | -3.192 | *-9.530* *** | -2.076 | *-8.273* *** | -3.206 | *-9.505* *** | 0.132 | *2.468* *** |
| Y92 | - | | 0.137 | *2.571* * | 0.155 | *2.929* ** | 0.132 | *2.470* | -0.114 | *-2.054* |
| Y94 | - | | -0.130 | *-2.350* | -0.132 | *-2.394* | -0.112 | *-2.121* | -0.133 | *-2.378* |
| Y95 | - | | -0.154 | *-2.768* ** | -0.166 | *-3.045* ** | -0.134 | *-2.466* | -0.257 | *-4.208* |
| Y96 | - | | -0.282 | *-4.626* *** | -0.304 | *-5.189* *** | -0.256 | *-4.319* *** | -3.664 | *-9.944* *** |
| EU_EKA | - | | - | | 0.672 | *3.007* ** | - | | 1.127 | *3.041* ** |
| R2_EUEKA | | | | | 0.133 | 0.262 | | | | |
| FR3 | - | | - | | - | | 0.581 | *2.349* | - | |
| G_P | - | | - | | - | | - | | 0.744 | *1.944* |
| G_S | - | | - | | - | | - | | 0.733 | *3.106* ** |
| Log likelihood | -1949.08 | | -1298.32 | | -1865.90 | | -1925.59 | | -1917.26 | |
| Wald chi2 (Pr>chi2) | 175.82 | (0.00) | 243.86 | (0.00) | 811.61 | (0.00) | 261.45 | (0.00) | 286.08 | (0.00) |
| LR test -pooled- (Pr>chi2) | 1689.42 | (0.00) | 1690.34 | (0.00) | 1244.77 | (0.00) | 1650.74 | (0.00) | 1684.99 | (0.00) |

**Table 21**
**Imputed MB Sample (N=366)**
**Poisson Random Effects**

| | (1) | | (2) | | (3) | | (4) | | (5) | |
|---|---|---|---|---|---|---|---|---|---|---|
| Dep. Var. | Coef. | *z-value* | Coef. | *z-value* | Coef. | *z-value* | Coef. | *z-value* | Coef. | *z-value* |
| PAT | | | | | | | | | | |
| LAGST | 0.194 | *1.688* | 0.698 | *5.141* *** | 0.704 | *7.580* *** | 0.659 | *4.785* *** | 0.640 | *4.516* *** |
| LEMP | 0.417 | *3.436* *** | 0.383 | *3.669* *** | 0.203 | *2.502* | 0.418 | *3.984* *** | 0.347 | *3.171* ** |
| RJV_2 | -0.048 | *-2.102* | 0.069 | *2.244* | -0.474 | *-3.059* ** | | | 0.066 | *2.155* |
| CHEM | 1.260 | *1.567* | 1.196 | *1.812* | 0.922 | *2.153* | 1.183 | *1.770* | 1.250 | *1.828* |
| DRUGS_P | 2.144 | *3.141* ** | 1.313 | *2.385* | 0.848 | *2.570* * | 1.373 | *2.450* | 1.596 | *2.554* |
| COMP_P | - | | - | | - | | - | | - | - |
| E_EQUIP | - | | - | | - | | - | | - | - |
| COMM_P | - | | - | | - | | - | | - | - |
| P_EQUIP | - | | - | | - | | - | | - | - |
| ENGMAN | 2.725 | *3.199* *** | 2.124 | *3.029* ** | 1.288 | *2.889* ** | 2.282 | *3.149* ** | 2.308 | *2.969* ** |
| FOOD | -0.378 | *-0.513* | -0.273 | *-0.432* | - | | -0.282 | *-0.438* | -0.237 | *-0.360* |
| LAGNOST | -0.824 | *-1.849* | -0.472 | *-1.236* | -2.702 | *-8.672* *** | -0.528 | *-1.357* | -0.426 | *-1.063* |
| CONS | -5.349 | *-5.314* *** | -5.470 | *-6.535* *** | -3.159 | *-4.955* *** | -5.644 | *-6.673* | -5.770 | *-6.509* *** |
| Y92 | - | | 0.183 | *1.826* | 0.203 | *2.071* | 0.153 | *1.534* | 0.168 | *1.667* |
| Y94 | - | | -0.169 | *-1.696* | -0.161 | *-1.646* | -0.135 | *-1.373* | -0.155 | *-1.544* |
| Y95 | - | | -0.482 | *-3.846* *** | -0.398 | *-3.302* *** | -0.340 | *-3.174* ** | -0.454 | *-3.584* *** |
| Y96 | - | | -0.781 | *-5.397* *** | -0.705 | *-5.155* *** | -0.598 | *-5.100* *** | -0.746 | *-5.095* *** |
| EU_EKA | - | | - | | 0.075 | *0.204* | - | | 0.956 | *1.270* |
| R2_EUEKA | | | | | 0.550 | 3.658 *** | | | | |
| FR3 | - | | - | | - | | -0.137 | *-0.361* | - | - |
| G_P | - | | - | | - | | - | | 0.220 | *0.353* |
| G_S | - | | - | | - | | - | | 0.618 | *1.366* |
| Log likelihood | -605 | | -586 | | -546 | | -588 | | -588 | |
| Wald chi2 (Pr>chi2) | 68.090 (0.000) | | 111.440 (0.000) | | 351.29 (0.000) | | 109.13 (0.000) | | 109.13 (0.000) | |
| LR test -pooled- (Pr>chi2) | 288.14 (0.000) | | 302.25 (0.000) | | 190.21 (0.000) | | 300.50 (0.000) | | 300.50 (0.000) | |

**Table 22**
**Poisson and Negative Binomial Random Effects: IT sub-sectors**
**(Balanced Sample)**

| | Professional & Scientific Equipment | | | | | | Computer & Accounting Equipment | | | | | | Communication Equipment | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dep. Var. | Poisson | | | Negative Binomial | | | Poisson | | | Negative Binomial | | | Poisson | | | Negative Binomial | | |
| PAT | Coef. | Std. Err. | | Coef. | Std. Err. | | Coef. | Std. Err. | | Coef. | Std. Err. | | Coef. | Std. Err. | | Coef. | Std. Err. | |
| LAGST | 0.848 | (0.165) | *** | 0.499 | (0.218) | | 0.772 | (0.146) | *** | 1.046 | (0.130) | *** | 0.623 | (0.069) | *** | 0.827 | (0.074) | *** |
| LEMP | 0.005 | (0.178) | | -0.067 | (0.172) | | 0.176 | (0.129) | | -0.006 | (0.129) | | 0.062 | (0.026) | | 0.064 | (0.051) | |
| RJV_2 | -0.028 | (0.189) | | -0.427 | (0.330) | | -0.009 | (0.015) | | -0.011 | (0.019) | | 0.152 | (0.053) | ** | 0.036 | (0.059) | |
| LAGNOST | -3.469 | (0.706) | *** | -3.310 | (0.718) | *** | -1.424 | (0.478) | ** | -1.757 | (0.454) | *** | -1.812 | (0.393) | *** | -2.329 | (0.413) | *** |
| Y92 | -0.059 | (0.251) | | -0.716 | (0.483) | | 0.110 | (0.124) | | 0.189 | (0.161) | | 0.114 | (0.095) | | 0.368 | (0.164) | |
| Y94 | -0.584 | (0.343) | | -0.563 | (0.442) | | 0.012 | (0.138) | | 0.024 | (0.181) | | -0.311 | (0.098) | ** | -0.144 | (0.175) | |
| Y95 | -1.413 | (0.308) | *** | -1.744 | (0.564) | ** | -0.032 | (0.138) | | -0.068 | (0.177) | | -0.213 | (0.096) | | -0.138 | (0.169) | |
| Y96 | -1.903 | (0.362) | *** | -1.168 | (0.505) | | -0.271 | (0.148) | | -0.330 | (0.187) | | -0.436 | (0.110) | *** | -0.288 | (0.178) | |
| CONS | -0.546 | (1.254) | | 0.468 | (1.346) | | -2.696 | (0.899) | *** | -1.323 | (0.869) | *** | -1.201 | (0.331) | *** | -1.831 | (0.499) | *** |

Summing up, the Poisson and Negative Binomial Random Effects estimations of the imputed dataset give results that do not contrast with the evidence from the balanced sample, though we underlined some interesting differences: in the IT area the size component, in terms of employment, group affiliation and diversification of European programs involvement, is more important in explaining patenting activity, whereas in the MB area the evidence supports the idea of a positive effect of RJV membership, but only when considering a relatively long (at least three years) time span between RJV participation and patent applications. The estimation of a positive effect results when RJV participation is specified in terms of both projects flow (number of projects per year) and projects stock (total number of project-years). We may conclude that there is evidence of a positive effect of RJV affiliation both when considering the level of investment in co-operative R&D, in a given period, and when focussing on the level of co-operative experience cumulated over time.

### 5.1 – Analysis by size class

The comparison of balanced and imputed estimates reported above emphasises some peculiar results concerning size-related variables. This evidence points at an interesting line of investigation, which concerns the role played by size in determining the ability, or interest, of firms to take advantage from RJV affiliation. Accordingly, we try to assess whether the relationship between patenting activity and our explanatory variables changes over different size classes. Therefore, we estimate the basic specification with time effects on three subsamples, grouping firms in three size classes, according to the average number of employees over the period 1992-1996. We use the balanced dataset in order to limit the effects of measurement error in the employment variable. Unfortunately, not all subgroups present a high number of units, especially for the MB area. Hence any statistical inference exercise needs to take into account the specific sample size.

Table 23 presents RJV_2 estimated coefficients (and significance level) when employing the basic specification with time effects and the RE Poisson model.

In all three samples, we find a negative coefficient for the group of small firms, although we observe statistical significance only in the ALL estimates. Overall, small firms do not appear to reap much benefit from R&D co-operation, in terms of patenting activity. In this case, the measure of output, patent applications, is all the more important for the interpretation of results. Small firms with a low propensity to patent are likely to acquire from the RJVs benefits which are not measurable with patent output, such as access to best practices, opportunities for product and process developments that do not result into "innovations", as properly defined, costs and risk sharing for those developments, sharing of technological information otherwise hardly accessible. In other words, we may expect RJVs to give "patenting" benefits to firms which already exhibit a significant propensity to patent, rather than to firms, that, for managerial abilities or attitudes, internal resources and competencies, do not generally orient their more innovative investments or experimental routines to the patenting process.

**Table 23**

**Size class analysis (RJV_2 estimates)**

**Balanced Sample- Poisson Random Effects**

| | Coeff. | Z-value |
|---|---|---|
| **ALL** | | |
| CLASS 1 (N=274) | -0.252* | -2.563 |
| CLASS 2 (N=215) | 0.154 | 2.469 |
| CLASS 3 (N=169) | 0.000 | 0.018 |
| **IT** | | |
| CLASS 1 (N=208) | -0.361 | -1.632 |
| CLASS 2 (N=176) | 0.123 | 1.952 |
| CLASS 3 (N=124) | -0.009 | -0.722 |
| **MB** | | |
| CLASS 1 (N=82) | -0.109 | -0.771 |
| CLASS 2 (N=60) | 1.127 | 1.609 |
| CLASS 3 (N=49) | 0.143*** | 3.250 |

Note: all estimated equations include time effects

MEMP= average number of employees over the period 1992-1996

CLASS 1: MEMP<250

CLASS 2: 250 < MEMP<1000

CLASS 3: MEMP>1000

At the other extreme of the dimensional space, for large firms the impact of RJVs results positive and significant in the MB area only, though caution is required given the small sample size. In the IT area, the effect on large firms is clearly negligible. A possible interpretation of this outcome relates to the small importance, in relative terms, of RJVs projects for firms with large R&D departments and R&D budgets. Moreover, strategic issues can be extremely important in affecting large firms approach to information sharing and disclosure of technological developments by way of patenting. As we noticed in the first sections of this paper, large firms entering RJVs in the IT area are highly innovative prior to entry. We may expect that, for those firms, participation does not primarily aim at increasing patent output, but rather at joining key European networks with the main sector competitors and public R&D institutions.

There is some evidence of RJVs being more effective for medium-size firms, but results are not significant when we adopt the above classification. We further explore the impact of RJV affiliation on medium size units by enlarging the sample (Table 24). First, we simply drop out of the sample small firms (with less than 250 or 500 employees), identifying a group of medium large firms. Then we define a class of medium size firms, setting lower and upper limits at 250 and 1600 employees, where the latter represents the average level of employment in the overall sample[33].

In the IT area, we find evidence of a significant participation effect on medium-size firms, though results are highly sensitive to the definition of the dimensional class[34].

**Table 24**

**Large and medium size firms (RJV_2 estimates)**

**Balanced Sample- Poisson Random Effects**

| | Coeff. | Z-value |
|---|---|---|
| **ALL** | | |
| MEMP>250 (N=384) | 0.003 | 0.273 |
| MEMP>500 (N=253) | -0.000 | -0.017 |
| 250<MEMP<1600 (N=265) | 0.093 | 1.789 |
| **IT** | | |
| MEMP>250 (N=250) | -0.007 | -1.621 |
| MEMP>500 (N=191) | -0.120 | -0.971 |
| 250<MEMP<1600 (N=215) | 0.140* | 2.560 |
| **MB** | | |
| MEMP>250 (N=109) | 1.173*** | 4.258 |
| MEMP>500 (N=77) | 0.164*** | 3.837 |
| 250<MEMP<1600 (N=73) | 1.025 | 2.161 |

Note: all estimated equations include time effects

MEMP= average number of employees over the period 1992-1996

Results about firms in the MB area are qualitatively different and more robust to size class specification. The coefficient of the RJV variable, which was non significant when focussing on medium-size firms only (class2: 250<memp<1000), becomes highly significant if we drop from the overall MB sample firms with less than 250 or 500 employees. Contrarily to what emerges in the IT area, large firms' patenting activity seems to get a significant impulse from participation to European R&D consortia.

[33] The average level of employment in ALL is equal to 1554 employees.

[34] The coefficient of RJV_2 is significant when the upper limit of the class ranges from 1550 to 1600, i.e. around the average level of employment in the overall sample. However, the variable becomes slightly non significant if we move the threshold both downward (e.g. 1400 employees) and upward (e.g. 1700 employees). It clearly turns to non-significance once we reduce the sample to firms with less than 1300 employees or extend it to firms with more than 1800 employees.

The positive result does not allow any definite conclusion but points at a very interesting line of research, inspecting the relationship between firm size and purposiveness of co-operative R&D investments.

## 6- Conclusions

Co-operative technological activities have been extensively investigated by recent theoretical and empirical literature, which has mainly referred to evidence collected through surveys of firms.

In this paper we presented an assessment exercise of European RJV effect on patenting activity, which elaborates a large panel data of European firms engaged in co-operative research activities in the areas of *Information and Communication Technology* and *Medical and Biotechnology*.

In the course of the analysis we have extensively discussed about peculiar methodological and statistical problems, which arise in this area of empirical research, such as the need to take into account the integer nature of the patent variable, the methodological questions posed by the existence, in the dataset, of a large share of non-innovative units, the issue of self-selection by more dynamic actors in the R&D consortia, the robustness of results to different variables and models specifications.

The microeconometric analysis has been performed adopting several statistical models, from standard linear fixed effects and random effects specifications to Poisson and Negative Binomial count data models. Given the peculiar nature of our dataset, which presents a very large number of non-innovative firms, hence of zero-value patent observations, we consider count data models more appropriate than standard linear regression, even when performed after adequate adjustments on the dependent variable.

The empirical analysis provides interesting results in terms of nature of patenting activity and differences between technological macro areas.

a) Cumulated knowledge is the most relevant explanatory factor of current innovative activity. Results give support to the idea that technological change is a *cumulative* process: firms which have patented most in the past are most likely to patent today, even when taking knowledge obsolescence into account. Indeed, it appears that the existing knowledge base, as proxied by the cumulated number of patents, explains current innovative activity better if we assume it depreciates faster.

The result is robust to technological differentiation, although, in general, we detect a greater effect in the IT field. This finding may be related to the different stages of the life cycle of IT and MB technologies and to the structure and dynamics of European industry.

European research in IT has mostly involved large industrial groups, having significant experience with both generic technologies and applied research. The hierarchy of innovators in this field appears to be relatively stable, as discussed by empirical research about technological regimes (e.g. Malerba and Orsenigo, 1995, 1997). European programmes have attracted firms that were already remarkably more dynamic that the average European level.

On the contrary, the development of biotechnologies has induced a greater "instability" of innovators' hierarchy, so that the effect of the cumulated number of patents, which is nevertheless remarkably important, is generally smaller.

b) RJV participants tend to be more innovative than non-members. In the IT area, the most important waves of European programs have attracted highly research intensive firms, that were already remarkably more innovative than the average European level, whereas, in the MB area, except for a peak in 1989, early RJV members did not exhibit high levels of patenting prior to entry.

This finding is related to the structure and dynamics of the European industry as underlined in a). In this respect, the target and nature of the European consortia appear to be a relevant explanatory factor. In fact, IT projects are oriented towards more generic, pre-competitive research, with the aim of providing a common technological basis for IT applications and support for the development of a European market for information services. All the major telecommunication operators and key European equipment manufacturers participate in these projects. In the MB field, a high share of consortia deals with applied projects, set under the EUREKA Programme, and attracts, together with a few major players, a large number of firms that are not endowed with a significant "innovative platform", expressed by patents.

c) Size is positively related with patenting activity in the IT field. In the MB area, there is no clear evidence of largest firms, in terms of employment level, being more innovative.

This result is supported by the finding about significance of regressors different than employment level, that are, in various ways, related to size and investment resources. For instance, the group dummy variable, which identifies subsidiaries of large industrial groups, is significant, when employing the larger imputed dataset, in the IT area, but not in the MB field. When focussing on the smaller balanced sample, which includes relatively larger MB firms, the patenting level is positively correlated with the Parent Company status. We may relate this apparently contrasting results about employment level and Parent dummy to the fact that, in the MB area, parent companies are not necessarily the largest firms, but are catalyst for the group patenting activity. The interpretation of the result should however take into account patent application policies within groups, which do not necessarily reflect an effective centralisation (or de-centralisation) of research activities.

A similar outcome is obtained when patenting activity is related to the general quality of being a frequent participant, that mostly applies to large firms, in place of the yearly number of RJV participations. When choosing a cut-off number above five project-years, the variable appears significant in the IT field, while it is always non-significant in the MB area.

d) European RJVs seem to have positively affected the patenting level of firm participating to MB projects. The favourable evidence is robust to the use of alternative statistical models and to alternative specifications of the lag between RJV affiliation and patent applications. Indeed, the magnitude of the effects appears greater the longer is the lag. This finding is consistent with the medium-long term duration of MB projects, which is, on average, significantly longer than in the IT case.

This positive result partly reflects the large number of barely innovative firms which entered the "innovation track" only after the first entry into a EU-supported RJV. Hence, the positive effect might be related to the role of RJVs in opening up "innovative networks" to new members.

However, the estimated effect may also be explained with other factors, more directly connected with the nature of R&D in these consortia and the evolution of the technological field.

As underlined, applied market-oriented projects in this field represent a relevant share within European-supported programmes, especially within EUREKA. In this sense, if we take patents as indicators of performance, it appears that the programmes' objective to promote research in emerging fields and innovative application has been reached. However, we expect the estimated impact to reflect also the dynamics of the field during

the time RJVs have been established, so that, patent applications during or in the aftermath of the project mirror the developments of fast-emerging areas of research.
Nevertheless, the inclusion in the estimated sample of a group of non-participating firms, similar to participants, in terms of countries and sectors composition, should allow to attribute this effect to participation, rather than to general trends in the field, or, at least, to recognise that European consortia have attracted firms with a high "innovative potential", which has been developed and expressed in patent output in the course or immediately after these projects.

e) There is no clear and robust evidence of a positive correlation between patenting activity and RJV affiliation in the IT area, even when focussing on sectors, such as Computers and Office Equipment, Communication Equipment, Professional and Scientific Equipment, which exhibit a relatively high propensity to innovate.
As in the MB area, the result may be related to the characteristics of research undertaken in the consortia. In the IT case, R&D projects are definitely more oriented towards pre-competitive research and development of generic technologies. When combining this typology of research with the description of participants' characteristics, it appears that consortia attract the major leaders in the area, reinforcing their role, rather than opening up innovative networks to new members with a high, but still unexpressed, innovative potential.

f) Interesting results emerge from the analysis of RJV impact by size class.
Small firms do not appear to benefit from RJV participation in terms of increase in patent applications. Rather, RJV affiliation seems to reduce their patent output.
In the IT area, RJV participation seems to positively affect medium-size firms patenting, although statistical significance of this finding is sensitive to alternative definitions of the medium-size class, but does not affect the patenting behaviour of large firms.
In the MB area, there is no evidence of medium-size firms reaping the most from RJV affiliation. Rather, the overall significant results further improve when dropping small firms out of the sample. The patent activity of medium-large firms appear to get significant impulse from RJV participation. Hence, in this field, even if size itself does not appear to explain patenting behaviour significantly, size seems to affect the ability to benefit from co-operative R&D.

In this sense, the most interesting outcome concerns the non-significant or negative effect of participation on the innovative performance of small entities, as if a minimum amount of resources and extension of productive activities was required to be able to exploit joint investments and the externalities generated in the course of the interaction. However, this is also likely to be the case in which the patent measure mostly under-estimates innovative ability, given the low propensity to patent of small firms.

The results of the empirical investigation point at interesting areas for further studies. In particular, the differences between technological macro sectors and the role of the cumulated innovative experience and size in determining ability to reap benefits from co-operation could be further investigated.

In this respect, an important place in the research agenda should be assigned to the identification of appropriate measures of knowledge-related characteristics. In this perspective, the patent measure exhibits great limitations, since it captures only one aspect of the structure and extent of the knowledge base of economic agents. Of peculiar relevance for our analysis, is, for instance, the general low propensity to patent of European firms, which may adopt different devices for protecting their knowledge output. Moreover, patent output does not effectively capture “innovative potential”, which may represent a critical asset for taking advantage from co-operative activities.

In this sense, empirical research is to benefit from the development of tools for investigation that more closely reflect the rich conceptualisation of knowledge and innovative processes proposed by the theoretical literature about technological regimes and knowledge creation in co-operative settings.

As far as policy implications are concerned, empirical results are to be interpreted taking into account that IT and MB industries are at different stages of their life cycle. It appears that co-operative policies reinforce existing leaders and networks in "mature industries", where a "network of excellence" has already emerged and hierarchy of innovators is rather stable, whereas it favours the exploitation of innovative potential by new actors in the case of emerging technologies.

This finding suggests that a peculiar attention is to be given to sector dynamics at the stage of policy design and in the evaluation process of policy targets and results.

# References

Arrow K.J.(1962), 'Economic Welfare and the Allocation of Resources for Invention', in *The Rate and Direction of Inventive Activity: Economic And Social Factors*, NBER, Princeton

Aw B.Y., Chung S., Roberts M.J. (2000), 'Productivity and turnover in the export market: micro-level evidence from the Republic of Korea and Taiwan', *The World Bank Economic Review*, vol 14, no 1, pp. 65-90

Benfratello L., Sambenelli A. (2000), *Research Joint Ventures and Firm Level Performance*, W.P. 13-2000, Fondazione Eni "Enrico Mattei", Milan

Bernard A.B., Jensen J. B. (1995), 'Exceptional exporter performance: cause, effect or both?', *Journal of International Economics* 47 , pp. 1-25

Branstetter L., Sakakibara M. (1998), 'Japanese Research Consortia: a microeconometric analysis of industrial policy', *Journal of Industrial Economics*, vol XLVI, pp 207-233

Bussoli P. (2000*), An empirical analysis of technological convergence process and RJVs in Europe at the firm level*, WP 12-2000, Fondazione Eni "Enrico Mattei", Milan

Clerides K.S., Lach S., Tybout J.R. (1998), 'Is learning by exporting important? Micro-dynamic evidence from Colombia, Mexico and Morocco', *Quarterly Journal of Economics*, 113, pp. 903-947

Contractor F.J., Lorange P. (eds) (1988), *Cooperative Strategies in International Business*, Lexington Books, Lexington

Cuneo P., Mairesse J. (1984), 'Productivity and R&D at the Firm Level in French Manufacturing', in Griliches Z. (ed), *R&D, Patents and Productivity*, University of Chicago Press, Chicago

Geroski P. (1995), 'Markets for Technology', in P. Stoneman (ed), *Handbook of the Economics of Innovation and Technological Change*, Basil Blackwell, Oxford

Griliches Z. (1984), *R&D, Patents and Productivity*, University of Chicago Press, Chicago

Griliches Z. (1995), 'R&D and Productivity: Econometric Results and Measurement Issues', in P. Stoneman (ed), *Handbook of the Economics of Innovation and Technological Change*, Basil Blackwell, Oxford

Griliches Z. (1998), *R&D and Productivity. The econometric evidence*, University of Chicago Press, Chicago

Griliches Z., Mairesse J. (1984), 'Productivity and R&D at the Firm Level', in Z. Griliches (ed), *R&D, Patents and Productivity*, University of Chicago Press, Chicago

Hagedoorn J. (1993), 'Understanding the rationale of strategic technology partnering: interorganizational modes of cooperation and sectoral differences', *Strategic Management Journal*, vol 14, pp. 371-385

Hagedoorn J., Narula R. (1996), 'Choosing modes of governance for strategic technology partnering: international and sectoral differences', *Journal of International Business Studies*, vol 27, pp. 265-284

Hall B., Mairesse G. (1995), 'Exploring the Relationship between R&D and Productivity Growth in French Manufacturing Firms', *Journal of Econometrics*, vol 65, pp. 263-294

Hausman J., Hall B. Griliches Z. (1984), 'Econometric Models for Count Data with Application to Patents- R&D Relationship', *Econometrica*, 52, pp. 909-938

Henderson R., Cockburn I. (1996), 'Scale, scope, and spillovers: the determinants of research productivity in drug discovery', *Rand Journal of Economics*, vol 27, No 1, pp. 32-59

Hsiao C. (1986), *Analysis of panel data*, Cambridge University Press, Cambridge

Jaffe A. (1986), 'Technological Opportunity and Spillovers of R&D: evidence from firms' patents, profits and market value', *American Economic Review*, vol 76, No 5, pp. 984-1001

Kamien M.I., Muller E., Zang I. (1992), 'Research Joint Ventures and R&D Cartels', *American Economic Review*, vol 82, No 5, pp. 1293-1306

Katz M.L. (1986), 'An Analysis of Cooperative R&D', *Rand Journal of Economics*, 17 (4), pp. 527-543

Little R.J.A., Rubin D.B. (1987), *Statistical Analysis with Missing Data*, John Wiley & Sons, New York

Llerena P., Matt M. (1999), *Inter-organizational collaborations: the theories and their policy implications*, WP 9703, BETA, Université "L. Pasteur", Strasbourg

Malerba F., Orsenigo L. (1993), 'Technological Regimes and Firm Behavior', *Industrial and Corporate Change*, vol 2, No 1, pp. 45-74

Malerba F., Orsenigo L. (1997), 'Technological Regimes and Sectoral Patterns of Innovative Activities', *Industrial and Corporate Change*, 6, pp. 83-118

Narula R., Dunning J. (1998), 'Explaining International R&D Alliances and the Role of Governments', *International Business Review*

Nelson R.R, Winter S.G. (1982), *An Evolutionary Theory of Economic Change*, Harvard University Press, Cambridge, MA

Sakakibara M. (1997), 'Evaluating government-sponsored R&D consortia in Japan: who benefits and how?', *Research Policy*, 26, pp. 447-473

Sinha D.K., Cusumano M.A. (1991), 'Complementary resources and co-operative research: a model of research joint ventures among competitors', *Management Science*, 37, 9, pp. 1091-1105

Spence M. (1984), 'Cost Reduction, Competition and Industry Performance', *Econometrica*, 52, pp. 101-121

Stoneman P., Vickers J. (1988), 'The assessment: the economics of technology policy', *Oxford Review of Economic Policy* 4 (4), pp.1-16

Vonortas N.S. (1997), 'Research joint ventures in the US', *Research Policy*, 26, pp. 577-595

Winter S.G. (1984), 'Schumpeterian Competition in Alternative Technological Regimes' *Journal of Economic Behaviour and Organisations*, 5, pp. 287-320