UNIVERSITÀ DEGLI STUDI DI FERRARA
DIPARTIMENTO DI ECONOMIA ISTITUZIONI TERRITORIO
Via Voltapaletto, 11 – 44100 Ferrara

Quaderno n. 9/2008

March 2008

Probability estimates from the Poisson model using repeated observations: an application to partnership between firms

Rosa Bernardini Papalia Silvia Bertarelli

Probability estimates from the Poisson model using repeated observations: an application to partnership between firms

Rosa Bernardini Papalia[1]

Silvia Bertarelli[2]

Abstract:
The purpose of this paper is to test Poisson models using repeated cross-section data, in the evaluation of the opportunities of partnership between small and medium sized firms - favoured by the European Union through Europartenariat - when the subject of the analysis is the typical European firm. Since contacts between firms are relatively few in numbers, Poisson regression models related to different projects and firms' profiles are estimated by implementing a GEE procedure. The idea of analysing a representative firm in a cohort-based framework allows us to study it in a dynamic perspective. In addition, normative evaluation of economic policy impact will be possible.

Keywords: Poisson model, repeated observation, Generalised Estimating Equation Approach, partnership between firms

JEL Classification: C33, C35, L50.

---

[1] Dipartimento di Scienze Statistiche, Università di Bologna,Via Belle Arti, 41 - 40126 Bologna, e-mail: rossella.bernardini@unibo.it.

[2] Corresponding author: Dipartimento di Economia Istituzioni Territorio,Università di Ferrara, Via Voltapaletto, 11 – 44100 Ferrara, e-mail: silvia.bertarelli@unife.it.

I. Introduction

In this paper we study the probability of finding contacts between firms interested in participating in the Europartenariat initiative, created to encourage international business and co-operation links between European small and medium sized enterprises. The identification of a units' problem appears in the evaluation of the opportunities of partnership between firms when the subject of the analysis is the typical European firm. This is due to the fact that the available data set contains information about firms that change over time. Firms observed in subsequent periods come from very different regions, but share some common characteristics such as the number of employees and their location in a less-favoured European region. Therefore, Europartenariat meetings cannot directly give evidence about the representative firm operating in the European economy.

Only by analysing common information drawn from repeated cross-section data, can this objective be reached. Traditionally, the creation of homogenous groups is a necessary condition to follow the same units over time. In addition, we argue that a pseudo-panel approach solves the units' identification problem appearing in the evaluation of partnership opportunities between small and medium sized firms. We try to evaluate it in terms of differentiated probabilities to find potential partners. Since contacts between firms are relatively few in numbers, Poisson regression models related to different projects and firms' profiles are estimated by implementing a GEE procedure.

The interest of this study comes from the idea that there exists a close relation between the contact and the success of a partnership. From this point of view, the high *ex-post* frequency of successful co-operation initiatives seems to give empirical support to our statement. *Ex-post* monitoring of contacts signals a high percentage of success in co-operation initiatives and small firms would potentially (like to) establish such economic relationships, which implies the presence of some form of impediment caused by general difficulties of creating opportunities of contact between such firms in the absence of EU intervention.

An overview of the paper is as follows. In section 2, we introduce the basic estimation problem of the probability of contacting potential partners. In section 3 we describe the database building procedure and the sample data structure. Section 4 presents the Poisson regression model. The estimation procedure is reported in section 5 and empirical results with reference to parameter

estimates, predicted probabilities, and marginal effects on event probabilities are discussed in section 6. Finally, section 7 summarises some concluding remarks.

II. Basic problem

The aim of this paper is to study the probability to find contacts between firms interested in participating in the Europartenariat initiative, created to encourage international business and co-operation partnership between European small and medium sized enterprises[3]. We refer to a representative European firm with two main characteristics: i) the firm is small or medium sized; ii) the firm operates in a less-favoured European region.

The Europartenariat program was launched by the European Commission in 1987 in order to stimulate the development of less-favoured regions by encouraging small and medium-sized business relationships with their counterparts all over the Community. Since then, the European authorities have organised 18 meetings. The participation rate of firms coming from more than 50 countries has grown, reaching more than two thousands firms for each Europartenariat event.

The main objectives in participating in the Europartenariat event are: i) create a joint venture; ii) find a distributor; iii) license technologies; iv) find components to complete a product; v) discuss joint research projects; vi) find a partner to subcontract research to. In our analysis these objectives are grouped in four categories - productive, technological, commercial, and financial - and the related probability of finding contacts are modelled separately.

We analyse the behaviour of the European firm by basing our analysis on repeated cross-section observations collected by sequentially studying Europartenariat initiatives. This information is characterised by several independent samples drawn sequentially over time, where observations refer to the same population (EU firms) but each unit is observed only once. A longitudinal analysis which follows the same subjects is possible using a pseudo-panel approach. The basic idea is that the population observed in the survey can be partitioned in groups called cohorts. Each cohort collects individuals sharing some common time-invariant characteristics. The pseudo panel aggregates all the observations at cohort level and therefore - subject to a few assumptions and some conditions - traditional longitudinal analysis can be performed on these cohorts (Deaton 1985, Moffit 1993).

[3] In this framework, Bernardini and Bertarelli (1998b) have analysed the theoretical problem of partnership with reference to search theory, while a simulation study related to the probability of contact between firms has been presented in Bernardini and Bertarelli (1998a).

In order to define the cohort we consider as a common time-invariant characteristic the number of employees (see table A2 in Appendix A for a description of cohorts). Contacts between firms in a given instant of time - measured by a nonnegative integer - are relatively few in numbers and are generated by a Poisson process. For modelling such an event, a Poisson regression model will be presented in the following sections.

III. Description of data

We have considered data from five Europartenariat meetings from 1992 to 1994, organised twice a year; specifically, we refer to Greece 1992, Italy 1992, France 1993, Scotland 1993, and Poland 1994. The data comes from catalogues issued by Europartenariat organisers and distributed to European firms interested in visiting local firms. The main feature of a local (host) firm is its location in a less-favoured European region. The catalogue contains information about economic activity, import-export position and collaboration proposals of the host firm. In addition, information about turnover, imports and exports, year of foundation, etc., is available.

A Europartenariat catalogue contains profiles of the participating host companies with details of their co-operation proposals. Copies of the catalogue are distributed throughout several countries including EFTA, Central and Eastern European and Mediterranean countries. EC and overseas companies have the opportunity to identify from the catalogue those local companies that they would like to meet. For their participation in a meeting, visiting companies receive support for practical arrangements without being charged any fee[4].

Each catalogue provides cross-section observations. We have already remarked in the previous section that the selected firms described in the catalogue change every time, so we cannot analyse the same group of local companies across time. Common characteristics are the location in a less-favoured European region, and the small-medium size. Therefore, the available data *i*) does not imply the observation of the same firm for subsequent periods; *ii*) does not directly give information about the representative firm operating in the European economy. More specifically, because of the second limitation, we cannot analyse the decision of partnership between European firms. To overcome these problems we have decided to build a pseudo-panel to be able to observe collaboration decisions with reference to homogenous groups from a dynamic perspective, and to concentrate the analysis on the representative European firm.

[4] For Europartenariat events organised in Member States, two thirds of the cost is born by the European Commission. The host region pays the other third. No fees are charged to visiting companies.

Time-invariant characteristics are assumed within each group, in order to guarantee homogeneity. So the new data set contains observations over time concerning 13 groups or cohorts, constructed with reference to the number of workers[5]. We can estimate the probability of contacting firms with common projects of collaboration using repeated observations, and study the incidence of time/regional differences.

Table 1 shows all relevant variables, and some related statistics. The database contains information on 1839 firms. We observe the presence of different projects as to commercial, productive, technological, and financial co-operation. The highest relative frequency of firms is connected with commercial proposals (73%) and the lowest one refers to financial projects (17%). Looking at the distribution by sector (table 2), we note it is not uniform and specifically firms are more concentrated in *mechanical*, *textile*, and *food* sectors (14,3%, 12,8%, 12%, respectively), but the number of firms in other sectors is always significant. Import-export oriented firms are approximately uniformly distributed across cohorts, with an average frequency of 68,2% for export-oriented firms and 52,6% for import-oriented ones. Looking at the distribution by cohort (table 3) approximately 31% of total firms is concentrated in the first and second cohorts related to firms with less than 25 employees; however, in the other cohorts the number of firms is always significant.

For the estimation, the following set of explanatory variables is available: average turnover per employee, exports, imports, the proportions of firms operating in different sectors and dummies related to regional and cohort effects. Definitions and some descriptive statistics for the whole sample are shown in table 3. Concerning the choice of the relevant variables, the selection has been implemented on the basis of a stepwise analysis[6].

IV. The Poisson model

Contacts between firms are relatively rare events, so we consider the stochastic process generating the number of contacts by a Poisson distribution. The model is based on the Poisson probability function:

[5] Appendix A shows the partition of firms by size.

[6] Most of the stepwise methods have been used as they indicate promptly how many explanatory variables may be needed in the model. Therefore the “best” single explained (dependent) variable and the “best” set of explanatory (independent) variables can be selected. More specifically, using the Efroymson’s procedure we start with an empty subset, and at each step we add the independent variable which gives the largest reduction of the residual sum of squares. When each new variable is added to the subset, partial correlations are considered to see if any of the variables in the subset has to be dropped.

$$f(y_{it},\theta_{it}) = P(Y_{it} = y_{it}) = \frac{e^{-\theta_{it}}\theta_{it}^{y_{it}}}{y_{it}!}, \quad for \quad y_{it} = 0,1,\dots,\infty \quad \theta_{it} > 0, \tag{1}$$

where y_{it} is the response variable indicating the number of contacts and x_{it} is the vector of explanatory variables observed at time t for cohort i. So i refers to the i-th group (or cohort) of homogenous firms with the same size and t is the time dimension. θ_{it} is the expected value of y_{it}, $E(y_{it})$, i.e. group i expected probability to have y_{it} contacts at time t.

If we express the model in terms of the observed y_i or its expectations $\theta_{it} = E(y_{it})$ we derive the formulation of the Poisson regression:

$$\ln(\theta_{it}) = \ln[E(y_{it})] = x'_{it}\beta \tag{2}$$

for $i = 1, \dots, m$ and $t = 1, \dots, n_i$, where there are n_i observations for each cohort identifier i. This formulation ensures that θ_{it} is always greater than zero because $\theta_{it} = \exp(x'_{it}\beta)$.

With reference to the within-group correlation structure, the correlation matrix is defined for any cohort i, $i = 1, \dots, m$, as:

$$R_{t,s} = \begin{cases} 1 & if \quad t = s \\ \rho & otherwise \end{cases} \tag{3}$$

where the i subscript has been suppressed to simplify the notation. In our analysis, we estimate two different models, assuming the following time correlation structures[7]:

i) an *independent* structure of within-group correlation (*IP model*), defined as

$$R_{t,s} = \begin{cases} 1 & if \quad t = s \\ 0 & otherwise \end{cases} \tag{4}$$

ii) a *multiplicative* correlation structure (*ARP model*), as an autoregressive AR(1) of the form

$$R_{t,s} = \begin{cases} 1 & if \quad t = s \\ \rho^{|t-s|} & otherwise \end{cases} \tag{5}$$

With reference to the latter assumption, an auto-regressive formulation implies decreasing correlation as the distance between observations increases. This hypothesis seems to be appropriate in our context. However, this reasoning also involves an alternative interpretation. Each time observation of the representative firm described by the cohort i can also be interpreted as an observation referred to a specific location in a European region. If we consider the sequence of Europartenariat initiatives from 1992 to 1994 (Greece, Italy, France, Scotland, and Poland), we can assume that the correlation of the group variable between two consequent events is greater than the correlation between two meetings far away from one other, because in the former case the two

[7] Unstructured and exchangeable within-group correlation structures are also plausible assumptions to be tested, but convergence problems have arisen in the estimation.

regions share more economic similarities than the latter one. For example, Italy and France are characterised by more comparable economic structures and institutions, than Italy and Poland.

V. Estimation procedure

A Poisson regression model describes contacts between firms sharing common projects. Estimates of the relevant parameters are implemented using the Generalised Estimating Equation (GEE) approach described by Liang and Zeger (1986). This represents a multivariate extension of the quasi-likelihood approach used to estimate the regression coefficients without completely specifying the joint distribution of the multivariate response (Hall and Severini, 1998). The term GEE has been introduced because nuisance parameters appear in these equations but not in quasi likelihood equations and because it is often impossible to obtain a unique value for the integral of GEEs to be used as an objective function (Liang-Qaqish, 1992).

The GEE approach assumes that the marginal density of y_{it} is

$$f(y_{it}) = exp\{[y_{it}\ \gamma_{it} - a(\gamma_{it}) + b(y_{it})]\phi\}.$$

The first two moments are given by

$$E(y_{it}) = \alpha'(\gamma_{it})$$

$$var(y_{it}) = v(\theta_{it})\phi = \alpha''(\gamma_{it})/\phi$$

The GEE approach proposed by Liang and Zeger (1986) is able to directly model the distribution of the response variable as a function of the covariates. This method proceeds by choosing a matrix (called “working correlation matrix”, WCM) to describe the relationships between the repeated observations on each subject. Specifically, the working correlation matrix for the *ARP model* - described in section 4 - is calculated as a function of Toeplitz matrices. Estimates of parameters are obtained by solving the generalised estimating equation with reference to the WCM matrix.

In the GEE approach, the covariance structure is essentially an incidental structure necessary to make the model adequate and it is important to select the working correlation matrix as close as possible to the empirical correlation. Under the hypothesis of no correlation between observations, it is assumed that the WCM is the identity matrix, while under the first order dependence hypothesis the observations are assumed to be correlated only with those immediately before or after them and hence the WCM is a triangle matrix.

VI Empirical results

In this section results concerning the empirical application are presented and discussed.

1 *Parameter estimates*

As we have already remarked in section 4, we estimate two models assuming independent and multiplicative correlation structures (*IP model* and *ARP model*) with reference to commercial, productive, technological, and financial projects.

The *ARP model*, which assumes an auto-regressive time correlation structure, seem to fit better by looking at coefficient significativities than the *IP model* for all four objectives, as is shown in table 4a and table 4b that describe coefficient and related standard errors for both models. As to the correlation structure of the *ARP model,* we find a correlation ranging in the following intervals: [-0.51, 0.26], [-0.81, 0.65], [-0.68, 0.46] and [-0.34, 0.12] for commercial, productive, technological, and financial projects, respectively.

It is noteworthy that the inclusion of cohort dummies provides a much better fit of data - both in the *IP* and *ARP model* - than the correspondent models without them. However, estimated coefficients are significant only in the *ARP* specification relating to some explanatory variables. A positive influence of the average turnover on the number of contacts emerges in commercial, productive, and technological objectives, even if the magnitude of the effect is moderate.

2 *Predicted probabilities*

Using equation (1) - presented in section 4 - we can calculate the predicted probability that the event count is equal to 0, 1, 2, 3, ... given some fixed values for the independent variables. This equation expresses event probability as a function of parameters β and explanatory variables x, so it is possible to have a predicted probability of some "selected" sub-groups of interest or types of firms. These predicted probabilities tell us proportionally how many members of each group may have success in 1, 2, 3, ... contacts. Predicted probabilities are intuitively appealing because they give an idea of how likely success is in a contact for different firms' profiles.

The probabilities calculated for different levels of turnover *per* employed are presented in Figure 1 with reference to all four objectives. The probability of having 1, 3, 5, and 10 contacts associated to different projects, increases as the turnover increases, except for financial objectives. Commercial objectives are associated with the highest probabilities, while financial projects present the smallest values.

We have also calculated predicted values restricting the analysis on one hand to regional aspects and on the other hand on dimensional ones. Regional and cohort effects on the probability of contacting potential partners provide other interesting results reported in figure 2 and in tables 5a-5b. With reference to regional effects, predicted probabilities are not uniform across regions for each objective. All plots depicted in figure 2 consider the exact timing of the meetings organised by the regions. Therefore the estimated probabilities across regions are equal to the estimated probabilities across time. In this view, the probability partly increases over time for commercial projects. The opposite arises for the financial and the technological cases. In productive projects, the probability neither constantly increases nor decreases. It is difficult to accept the idea of no learning in the organisation of meetings even though data apparently shows that the probability of finding contacts in Europartenariat events does not improve across time. We suspect regional differences strongly condition the success of the event.

With reference to firms' dimensions, several differences between cohorts emerge by the analysis of the predicted probabilities reported in tables 5a-5b. The shaded areas of tables 5a-5b include the estimated probabilities, which are approximately equal to zero. By comparing these zones in terms of dimension, we observe the highest number of null result in financial projects. With reference to the technological objective, characterised by the smallest shaded area, null values emerge in some cases where firms are very small and the number of contacts is low. This result does not appear in the other cases where null values are often referred to the highest number of contacts. In addition, we observe a high probability of having one contact (more than a 10% chance) for several cohorts in all projects: for the commercial objective in cohorts 5 and 12; for the productive objective all cohorts; for the technological objective cohorts 7, 9-12; for the financial objective cohorts 1-4 and 13.

The highest probabilities of finding one contact is always observed with reference to the larger cohorts for all objectives except the productive one (maximum value: 35% localised in cohort 5). The maximum values for the common project are in cohorts 9, 11, and 12; for the technological one the highest probability refers to cohort 12 and with reference to the financial one, 26% of firms in cohort 13 can have a successful contact.

In addition, with reference to each cohort we have studied the evolution of predicted probability when the number of contacts increases. We note they decrease almost every time, except for the technological case where a positive correlation for cohorts 1 and 2 emerges for a number of contacts less than (or equal to) 25.

To summarise, the probabilities for productive and financial projects are at least equal to 10% (but less than 30%) for the “1 contact” event, but they slump to less than 1% for more than 1 contact. It seems there are difficulties in arranging more than one meeting. As to the commercial and technological objectives, small firms have more difficulties than medium sized firms.

3 *Marginal effects on event probabilities*

We can look at the marginal effect of some variables on the probability of the event. We determine the marginal effect on event probability in Poisson models, by calculating the partial derivative of the probability with respect to an independent variable x_k. In general, the effect is given by the following equation:

$$\frac{\partial Prob(Y=y)}{\partial x_k} = \frac{\beta_k \theta_{it} e^{-\theta}\left(y\theta_{it}^{\,y-1} - \theta_{it}^{\,y}\right)}{y!} \qquad (6)$$

where $y = 0, 1, 2, 3 \ldots$, is the number of contacts between firms. The marginal effect on the probability changes as the value of independent variables varies (see Greene, 1990). Specifically, the marginal effect on the probability in the Poisson regression model measures the change of the expected probability to have y contacts given a unit change in x_k.

Marginal effects have been calculated with reference to the turnover per employed and the proportion of firms in the mechanical, textile, and food sectors. Results are not reported in the paper.

Confirming the previous analysis concerning the predicted probability, the marginal effects connected with turnover changes is smaller the larger ~~is~~ the number of contacts y for all types of project, other than for financial ones. In addition, a negative variation of the expected probability comes from a 1% change of the participation rate of firms operating in *mechanical*, *textile*, and *food* sectors. These results emerge in the general model, as well as in the formulations which consider the conditioning of different regions. As to cohort effects, we do not observe a common behaviour.

VII. Remarks and conclusions

The main significance of the paper consists in testing Poisson models as adequate tools to describe the probability of creating contacts between firms interested in participation in Europartenariat meetings. The purpose of this initiative is to encourage international business and

co-operation partnership between small and medium sized enterprises by organising forms of contact where potential partners can meet at limited cost.

Without the EU intervention, small firms would potentially start economic relationship with some form of impediment in implementation as a consequence of general difficulties in creating opportunities of contact between firms. The idea comes from the observation that there is a high e*x-post* percentage of success in collaboration.

We have argued that Europartenariat intervention can be evaluated in terms of differentiated probabilities of finding potential partners compared to the alternative opportunities that EU authorities offer a typical European firm. The objective can be reached with a strategy of exploiting cross-sectional data in order to extrapolate relevant information about its dynamic structure in the estimation procedure.

Contacts between firms are relatively few in numbers; this feature has justified the introduction of a Poisson regression model to describe the phenomena: specifically, a pseudo-panel approach has been used. Estimates of the Poisson model parameters have been implemented using the Generalised Estimating Equation approach.

We have analysed the effects of different types of objectives in estimating the probability. Productive, technological, commercial, and financial projects have been considered, and the related probabilities of finding contacts have been specified separately. For each of them, we have estimated two models assuming independent and multiplicative correlation structures. The latter model seems to better fit the data than the former one in terms of coefficient significance. In our context, the result confirms the idea that hosting countries with more similar economic structures share higher within group correlations. Probabilities for different levels of turnover *per* worker have been calculated. We find that the probability of having contacts associated with differentiated projects increases as the turnover increases, except for the financial one. In other words, firms with a high average turnover, i.e. low labour-intensive firms, have more chance of finding contacts. Commercial objectives are associated with the highest probabilities, while financial projects present the smallest values.

Regional and cohort effects on the probability of contacting potential partners provide other interesting results. With reference to regional effects, it is not evident that the estimated probability of finding contacts in Europartenariat initiatives improves across time. This evidence can be interpreted by the existence of regional differences between meetings. As to cohort effects, two main results emerge. First, predicted probabilities decrease almost every time when the number of contacts increases except for the technological case. Second, small firms have more difficulties than medium sized firms concerning technological objectives. These results need a further investigation.

Marginal effects have been calculated with reference to the turnover per employed and to the proportion of firms operating in the *mechanical*, *textile*, and *food* sectors. Marginal effects connected with turnover changes are positive for all types of project, other than the financial one. In addition, a negative variation of the expected probability comes from an increase of the participation rate of firms included in these sectors. The presence at the meeting of the widest number of represented sectors, instead of a high number of firms for each sector, seems to guarantee a high probability of finding successful contacts.

We can conclude that our first goal of testing the Poisson models has been achieved. Estimation results give a good fit of data for all estimated models. In addition, the idea of analysing a representative European firm in a cohort-based framework has been effective for different reasons. First, we have been able to study the dynamic behaviour of a typical European firm. Second, normative evaluation of economic policy impact is possible: specifically, we find the existence of local influences in the successful organisation of a Europartenariat event. Another interesting result of the cohort analysis is related to the emergence of quite important size differences between firms referred to alternative projects. Looking at the evidence, the *a priori* strategy of building differentiated models, one for each project, has been profitable.

References

Aitkin M. (1995). Probability model choice in single samples from exponential families using Poisson log-linear modelling, and model comparison using Bayes and posterior Bayes factors, *Statistics and Computing* **5**, 113-120.

Bernardini Papalia, R., Bertarelli S. (1998a). An empirical dynamic programming model of partnership among firms, *La Statistica per le Imprese*, *SIS* 1997. Torino: Giappicchelli Editore.

Bernardini Papalia, R., Bertarelli S. (1998b). An optimal partnership search model: theoretical implications for the Europartenariat event, Collana di Teoria Economica, Dipartimento di Scienze Economiche Università di Bologna, Working Paper n. 310.

Geweke J. F. Keane P.M., Runkle D.E. (1997). Statistical inference in the multinomial multiperiod probit model, *Journal of Econometrics* **80**, 125-165.

Greene, W. H. (1990). *Econometric analysis*. Englewood Cliffs, NJ: Prentice Hall.

Hall B., Severini T.A. (1998). Extended Generalized Estimating Equations for Clustered Data, *Jounal of the American Statistical Association* **93**, 1365-1375.

Kelejian H. (1992). The logit model and panel data via repeated observations, *Economic Letters* **40**, 135-140.

Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes, *Biometrcs* **42**, 121-130.

Liang, K.Y., Zeger S.L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika* **73**, 13-22.

Liang K.Y. Zeger S. L., Qaqish B. (1992). Multivariate Regression Analyses for Categorica Data, *J. R. Statistic Society B* **54**, 3-40.

McCullagh P. and J. A. Nelder (1989) *Generalized Linear Models*. London: Chapman & Hall.

Moffit R. (1993). Identification and estimation of dynamics models with a time series of cross-sections, *Journal of Econometrics* **59**, 99-123.

Wedderburm R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton methods, *Biometrika* **61**, 439-447.

Table 1: Regional distribution of firms by sectors and import-export activities

	N firms	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5	Sector 6	Sector 7	Sector 8	Sector 9	Sector 10	Sector 11	N exp-firms	N imp-firms
Greece	303	11	40	0	30	10	38	67	56	20	31	0	224	217
%	100	3,63	13,20	0,00	9,90	3,30	12,54	22,11	18,48	6,60	10,23	0,00	73,93	71,62
Italy	386	0	55	46	65	28	49	46	65	5	17	10	187	144
%	100	0,00	14,25	11,92	16,84	7,25	12,69	11,92	16,84	1,30	4,40	2,59	48,45	37,31
France	414	20	40	15	144	25	53	28	41	48	0	0	283	155
%	100	4,83	9,66	3,62	34,78	6,04	12,80	6,76	9,90	11,59	0,00	0,00	68,36	37,44
Scotland	335	31	17	34	34	25	20	45	37	21	41	16	190	105
%	100	9,25	5,07	10,15	10,15	7,46	5,97	13,43	11,04	6,27	12,24	4,78	56,72	31,34
Poland	401	49	46	33	50	46	36	50	29	32	30	0	304	272
%	100	12,22	11,47	8,23	12,47	11,47	8,98	12,47	7,23	7,98	7,48	0,00	75,81	67,83
Total	1839	111	198	128	323	134	196	236	228	126	119	26	1188	893
	100	6,04	10,77	6,96	17,56	7,29	10,66	12,83	12,40	6,85	6,47	1,41	64,60	48,56

Table 2: Regional distribution of firms by cohort

COHORT	Greece	Italy	France	Scotland	Poland	Total
1	21	70	81	91	56	319
	6,58%	21,94%	25,39%	28,53%	17,55%	100%
2	39	118	88	80	50	375
	10,40%	31,47%	23,47%	21,33%	13,33%	100%
3	26	56	62	35	34	213
	12,21%	26,29%	29,11%	16,43%	15,96%	100%
4	48	38	50	30	40	206
	23,30%	18,45%	24,27%	14,56%	19,42%	100%
5	27	17	18	14	11	87
	31,03%	19,54%	20,69%	16,09%	12,64%	100%
6	18	19	12	9	18	76
	23,68%	25,00%	15,79%	11,84%	23,68%	100,00%
7	9	10	20	12	11	62
	14,52%	16,13%	32,26%	19,35%	17,74%	100%
8	11	9	23	10	22	75
	14,67%	12,00%	30,67%	13,33%	29,33%	100%
9	15	2	4	6	6	33
	45,45%	6,06%	12,12%	18,18%	18,18%	100%
10	10	7	9	10	11	47
	21,28%	14,89%	19,15%	21,28%	23,40%	100%
11	7	8	3	4	6	28
	25,00%	28,57%	10,71%	14,29%	21,43%	100%
12	8	6	6	4	9	33
	24,24%	18,18%	18,18%	12,12%	27,27%	100%
13	64	26	38	30	127	285
	22,46%	9,12%	13,33%	10,53%	44,56%	100%
Firms *per* region	303	386	414	335	401	1839

Table 3: Description of variables and related statistics

Variable	*Description*	MIN	I PERC.	MEDIAN	III PERC.	MAX	MEAN	ST. DEV.	VAR.	SKEWN.	KURT.
ATURN	Turnover per employed*	8,79	49,37	88,64	116,96	313,78	91,64	58,32	3401,4	1,36	5,71
EXP	Exports	7,50	306,41	586,55	1237,25	7923,25	1158,05	1455,97	2119852	2,43	9,74
IMP	Imports	12,83	186,26	361,25	697,17	5696,94	685,45	983,45	967179,9	3,09	13,79
NS1	Relative number of firms in sector 1	0	0	0,037	0,110	0,25	0,0590	0,0669	0,0045	0,89	2,80
NS2	Relative number of firms in sector 2	0	0,038	0,097	0,179	0,333	0,1098	0,0932	0,0087	0,55	2,43
NS3	Relative number of firms in sector 3	0	0	0,037	0,111	0,333	0,0660	0,0812	0,0066	1,25	4,00
NS4	Relative number of firms in sector 4	0	0,067	0,143	0,250	0,611	0,1744	0,1426	0,0203	0,94	3,47
NS5	Relative number of firms in sector 5	0	0	0,077	0,111	0,333	0,0826	0,0890	0,0079	1,32	4,43
NS6	Relative number of firms in sector 6	0	0,025	0,105	0,147	0,5	0,1082	0,1026	0,0105	1,30	5,17
NS7	Relative number of firms in sector 7	0	0,050	0,128	0,211	0,5	0,1401	0,1166	0,0136	0,97	4,12
NS8	Relative number of firms in sector 8	0	0,043	0,120	0,182	0,5	0,1327	0,1150	0,0132	0,91	3,58
NS9	Relative number of firms in sector 9	0	0	0,021	0,091	0,247	0,0511	0,0632	0,0040	1,14	3,53
NS10	Relative number of firms in sector 10	0	0	0,026	0,102	0,286	0,0569	0,0721	0,0052	1,16	3,64
NS11	Relative number of firms in sector 11	0	0	0	0	0,25	0,0162	0,0498	0,0025	3,64	16,25
NEXP	Relative number of export-oriented firms	0,333	0,564	0,692	0,8	1	0,6820	0,1649	0,0272	-0,06	2,40
NIMP	Relative number of import-oriented firms	0,167	0,353	0,5	0,701	1	0,5260	0,2105	0,0443	0,25	2,42
NOBC	Relative number of firms with a commercial project	0,333	0,606	0,722	0,857	1	0,7282	0,1712	0,0293	-0,02	2,45
NOBP	Relative number of firms with a productive project	0	0,3	0,436	0,531	1	0,4179	0,1959	0,0384	0,17	3,55
NOBT	Relative number of firms with a technological project	0	0,252	0,333	0,462	0,857	0,3524	0,1725	0,2975	0,18	3,29
NOBF	Relative number of firms with a financial project	0	0,013	0,125	0,286	0,545	0,1677	0,1648	0,0272	0,80	2,47
OBC	Number of firms with a commercial project	1	6	12	29	81	20,3846	20,7061	428,7404	1,62	4,80
OBP	Number of firms with a productive project	0	3	7	18	68	12,0308	13,0444	170,1553	2,03	8,07
OBT	Number of firms with a technological project	0	3	6	13	48	9,9385	10,1810	103,6525	1,52	5,02
OBF	Number of firms with a financial project	0	1	3	6	33	4,8000	6,6572	44,3187	2,46	9,54

Table 4a: Parameter estimates from the Poisson model

	Commercial obj.				Productive obj.			
	IP Model		*ARP Model*		*IP Model*		*ARP Model*	
Expl. var.	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
aturn	0,0041	0,00117	0,0039	0,00110	0,0038	0,00159	0,0025	0,00096
exp	0,0000	0,00005	0,00001	0,00004	-0,00001	0,00007	0,0001	0,00004
imp	0,0000	0,00005	-0,0001	0,00006	0,00003	0,00006	-0,0001	0,00005
ns1	-7,0997	2,63824	-8,6451	2,50273	-10,8641	4,31241	-2,602	2,851
ns2	-12,252	2,574	-13,518	2,420	-14,384	4,179	-6,077	2,748
ns3	-12,565	2,415	-14,134	2,191	-14,706	3,984	-8,617	2,497
ns4	-12,487	2,437	-13,626	2,298	-13,890	4,063	-4,224	2,699
ns5	-13,819	2,634	-14,526	2,411	-16,376	4,252	-6,552	2,732
ns6	-10,568	2,372	-11,344	2,170	-12,470	4,005	-3,794	2,530
ns7	-12,923	2,516	-14,244	2,361	-15,126	4,147	-6,211	2,705
ns8	-9,971	2,528	-9,965	2,368	-13,795	4,144	-3,356	2,688
ns9	-17,921	3,038	-18,077	2,855	-20,199	4,620	-8,249	3,035
ns10	-12,333	2,504	-14,652	2,223	-14,249	4,093	-7,267	2,471
ns11	-12,341	2,938	-13,686	2,461	-18,095	4,888	-12,000	2,615
reg1	-0,4041	0,17669	-0,3991	0,16717	-0,4844	0,22986	-0,5451	0,15164
reg2	-0,4616	0,22916	-0,5644	0,22464	-0,5863	0,33039	-0,3172	0,24220
reg3	0,0899	0,20022	-0,0539	0,19434	-0,4697	0,27096	-0,9117	0,18141
reg4	-0,6959	0,22694	-0,7108	0,23660	-1,7018	0,33054	-1,0141	0,27093
coo1					0,0259	0,35089		
coo2	0,1320	0,14887	0,1951	0,13001	0,2283	0,28448	0,0338	0,11399
coo3	-0,4469	0,18577	-0,3387	0,17136	-0,2344	0,26203	-0,5040	0,15401
coo4	-0,4026	0,19995	-0,3581	0,18370	-0,5077	0,27123	-0,9037	0,16535
coo5	-1,1970	0,24936	-1,1150	0,22389	-1,3121	0,30098	-1,7624	0,20055
coo6	-1,5728	0,23522	-1,6365	0,20509	-1,9221	0,31497	-2,4815	0,18885
coo7	-1,8190	0,24270	-1,8289	0,20544	-1,8191	0,29057	-2,3468	0,18581
coo8	-1,4455	0,23270	-1,5177	0,21347	-1,4604	0,28452	-2,0260	0,19653
coo9	-2,6622	0,32771	-2,6274	0,29393	-2,7496	0,38011	-3,3397	0,27115
coo10	-2,1364	0,28930	-2,3037	0,26152	-2,0656	0,36040	-2,6922	0,24849
coo11	-2,7585	0,32784	-2,5614	0,25316	-2,0963	0,35503	-1,9148	0,21091
coo12	-2,5774	0,30724	-2,4202	0,25689	-2,2687	0,34107	-2,2584	0,19600
coo13	-0,3438	0,25248	-0,2905	0,23837			-0,4438	0,21194
constant	15,8014	2,41732	16,9235	2,23332	17,9040	4,07900	9,3075	2,57360

Table 4b: Parameter estimates from the Poisson model

	Technological obj.				Financial obj.			
	IP Model		*ARP Model*		*IP Model*		*ARP Model*	
Expl. var.	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
aturn	0,0050	0,00176	0,0019	0,00137	-0,0030	0,00375	-0,0024	0,0036
exp	0,0001	0,00007	0,0001	0,00005	0,00005	0,00014	0,00003	0,0001
imp	-0,000005	0,00007	-0,0001	0,00007	0,00001	0,00011	-0,00001	0,0001
ns1	-10,987	4,472	-4,957	3,389	-16,731	10,819	-15,379	10,694
ns2	-14,903	4,261	-9,010	3,219	-21,491	10,501	-20,375	10,190
ns3	-16,275	3,997	-12,152	2,917	-19,486	10,124	-20,175	9,720
ns4	-15,860	4,064	-9,653	3,050	-20,023	10,220	-19,466	9,993
ns5	-15,009	4,302	-8,898	3,145	-19,540	10,527	-18,269	10,226
ns6	-13,440	3,992	-8,065	2,905	-16,824	10,273	-16,120	10,039
ns7	-17,690	4,197	-11,707	3,128	-21,144	10,516	-20,629	10,196
ns8	-12,934	4,184	-5,537	3,137	-17,498	10,395	-16,060	10,164
ns9	-21,590	4,821	-12,680	3,710	-23,497	11,073	-21,940	10,676
ns10	-15,885	4,095	-12,370	2,904	-19,416	10,321	-19,322	9,810
ns11	-15,676	4,971	-12,986	3,387	-24,245	12,514	-22,887	12,004
reg1	0,4323	0,26513	0,5115	0,22014	0,5299	0,37973	0,4796	0,3706
reg2	-0,2707	0,35652	0,1549	0,31258	-0,8096	0,70630	-0,7650	0,6966
reg3	0,1061	0,30037	0,0134	0,25101	-0,5744	0,55826	-0,6208	0,5716
reg4	-1,1432	0,35898	-0,4796	0,33049	-2,9658	0,79721	-2,8904	0,7949
coo1					0,1432	0,66343		
coo2	0,0308	0,21887	-0,1875	0,15577	0,0840	0,58797	-0,1026	0,3055
coo3	-0,3864	0,26566	-0,6910	0,20535	-0,5491	0,54151	-0,5820	0,3493
coo4	-0,4876	0,29588	-0,9445	0,22179	-0,7014	0,53134	-0,7518	0,3934
coo5	-1,0994	0,34885	-1,5537	0,26496	-1,0122	0,54443	-1,1973	0,4283
coo6	-1,5207	0,35467	-2,1496	0,25181	-1,8409	0,55696	-2,1134	0,5074
coo7	-2,1978	0,38265	-2,7364	0,26283	-1,4721	0,49029	-1,6963	0,4920
coo8	-1,8281	0,35327	-2,3822	0,27343	-2,0986	0,59993	-2,4752	0,4860
coo9	-2,6347	0,44573	-3,0641	0,35263	-2,9028	0,60401	-3,1303	0,5954
coo10	-2,7110	0,46706	-3,3836	0,34434	-2,8826	0,71846	-3,2169	0,6751
coo11	-2,7762	0,41918	-2,7404	0,30205	-3,3535	0,75880	-3,2385	0,5916
coo12	-3,0676	0,47932	-3,0740	0,33058	-3,1170	0,69705	-3,3442	0,6634
coo13	-0,3846	0,37632	-0,7321	0,29838			-0,1698	0,6545
constant	18,1703	4,04040	12,6808	2,98221	22,6180	10,40011	21,9560	9,9287

Tab. 5a: Cohort-effects on the probability of contact between firms

	Commercial objective							
N. of contacts	1	3	5	7	10	15	20	25
Cohort								
1	0,0003	0,0057	0,0307	0,0791	0,1241	0,0421	0,0028	0,0001
2	0,0000	0,0011	0,0088	0,0335	0,0934	0,0828	0,0142	0,0007
3	0,0047	0,0422	0,1144	0,1477	0,0819	0,0049	0,0001	
4	0,0024	0,0266	0,0876	0,1375	0,1021	0,0100	0,0002	
5	0,1487	0,2240	0,1013	0,0218	0,0008			
6	0,1275	0,2221	0,1161	0,0289	0,0014			
7	0,2346	0,2014	0,0519	0,0064	0,0001			
8	0,1730	0,2221	0,0855	0,0157	0,0005			
9	0,3652	0,0476	0,0019					
10	0,3189	0,1419	0,0189					
11	0,3669	0,0526	0,0023					
12	0,3665	0,0513	0,0022					
13	0,0020	0,0229	0,0798	0,1327	0,1074	0,0121	0,0003	0,0000

	Productive objective							
N. of contacts	1	3	5	7	10	15	20	25
Cohort								
1	0,1968	0,2165	0,0715	0,0112	0,0003			
2	0,1214	0,2208	0,1205	0,0313	0,0016			
3	0,2403	0,1985	0,0492	0,0058	0,0001			
4	0,2979	0,1604	0,0259	0,0020				
5	0,3498	0,0299	0,0008					
6	0,2930	0,0107	0,0001					
7	0,2848	0,0094	0,0001					
8	0,3299	0,0200	0,0004					
9	0,1628	0,0011						
10	0,2285	0,0037						
11	0,2368	0,0043						
12	0,2125	0,0028						
13	0,2194	0,2083	0,0593	0,0081	0,0002			

Note: The shaded areas indicate predicted probabilities approximately equal to zero.

Tab. 5b: Cohort-effects on the probability of contact between firms

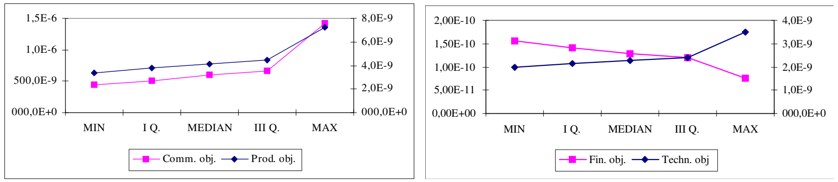
	Technological							
N. of contacts	1	3	5	7	10	15	20	25
Cohort								
1					0,0001	0,0027	0,0247	0,0671
2					0,0001	0,0027	0,0248	0,0672
3			0,0002	0,0019	0,0150	0,0786	0,0798	0,0236
4			0,0010	0,0059	0,0337	0,0990	0,0563	0,0093
5	0,0056	0,0480	0,1224	0,1488	0,0753	0,0039	0,0000	0,0000
6	0,0154	0,0910	0,1616	0,1368	0,0402	0,0008	0,0000	0,0000
7	0,1098	0,2174	0,1291	0,0365	0,0021			
8	0,0183	0,1008	0,1667	0,1313	0,0347			
9	0,2742	0,1781	0,0347	0,0032	0,0000			
10	0,1885	0,2188	0,0762	0,0126	0,0003			
11	0,1955	0,2169	0,0722	0,0114	0,0003			
12	0,3216	0,1392	0,0181	0,0011				
13								

	Financial objective							
N. of contacts	1	3	5	7	10	15	20	25
Cohort								
1	0,2356	0,0042						
2	0,2496	0,0053						
3	0,1701	0,0012						
4	0,1882	0,0018						
5	0,0948	0,0002						
6	0,0739	0,0001						
7	0,0935	0,0002						
8	0,0507							
9	0,0227							
10	0,0188							
11	0,0167							
12	0,0164							
13	0,2615	0,0064						

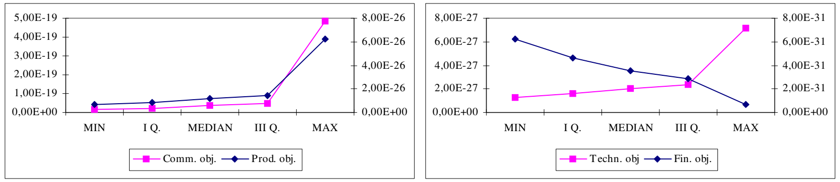
Note: The shaded areas indicate predicted probabilities approximately equal to zero.

Fig. 1 Per-capita turnover effects on the probability of contacting firms

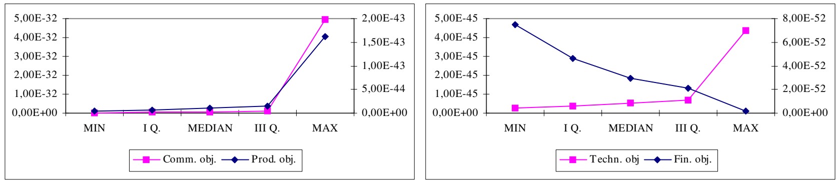
a) *Probability of having 1 contact*



b) *Probability of having 3 contacts*



b) *Probability of having 5 contacts*



b) *Probability of having 10 contacts*

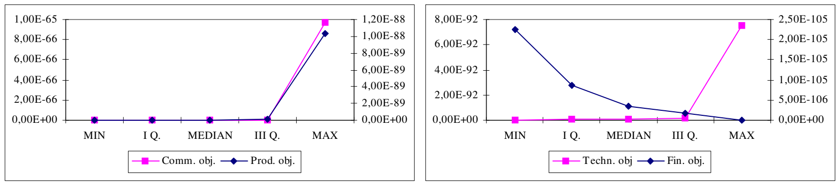
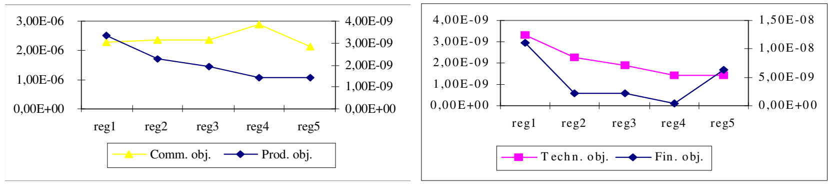
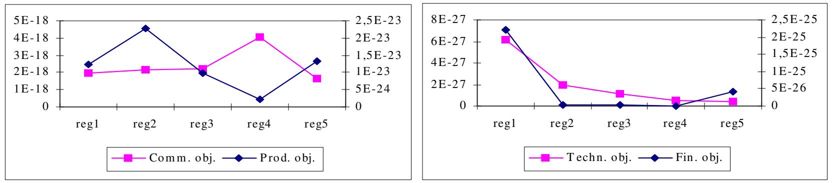


Fig. 2 Regional effects on the probability of contacting firms

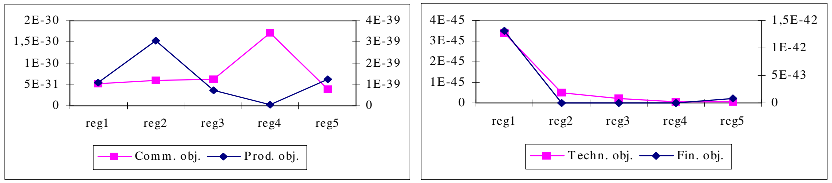
a) *Probability of having 1 contact*



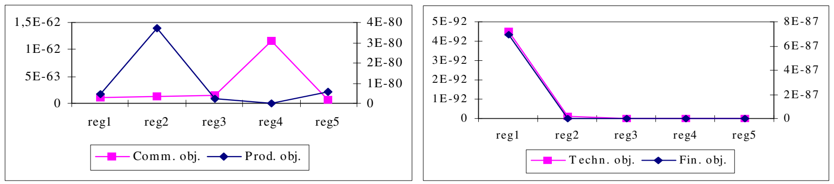
b) *Probability of having 3 contacts*



c) *Probability of having 5 contacts*



d) *Probability of having 10 contacts*



Appendix A

Table A1 Sectors

Sector	Description
S1	Building (materials, products and services)
S2	Chemicals and pharmaceuticals (plastic materials, rubber, glass, and ceramics)
S3	Electronics, electrical products
S4	Mechanics
S5	Transport services and manufacturing
S6	Paper, wood, and furniture
S7	Textiles, leather, and footwear
S8	Food
S9	Trade services
S10	R&D
S11	Computing and software

Table A2 Cohorts

Cohort	Number of employees (n)
1	$n < 13$
2	$13 \leq n < 26$
3	$26 \leq n < 39$
4	$39 \leq n < 52$
5	$52 \leq n < 65$
6	$65 \leq n < 78$
7	$78 \leq n < 91$
8	$91 \leq n < 104$
9	$104 \leq n < 117$
10	$117 \leq n < 130$
11	$130 \leq n < 143$
12	$143 \leq n < 156$
13	$n \geq 156$