NEMO Working Paper #10

Determinants of collaboration in European R&D networks: Empirical evidence from a binary choice model perspective

Manfred Paier and Thomas Scherngell
(Austrian Research Centers GmbH - ARC)

Supported by the EU FP6-NEST-Adventure Programme

Contract n° 028875 (NEST)

This version April 2008

Determinants of collaboration in European R&D networks: Empirical evidence from a binary choice model perspective

Manfred Paier[*] and Thomas Scherngell

[*]Corresponding author
Department of Technology Policy, Austrian Research Centers GmbH – ARC
Donau-City-Strasse 1, A-1220 Vienna, Austria
mailto:manfred.paier@arcs.ac.at

April 2008

Abstract. R&D collaboration networks stimulated by the European Union in its Framework Programmes (FP) on Research and Technological Development exhibit – like many other large social networks – characteristic features of complex networks. Existing analyses of these networks, however, are mostly built on strongly simplifying assumptions for network construction, like fully connected project graphs or star-graphs. Still, apart from qualitative evidence, little is known about the determinants that affect partner choice and network dynamics.

It is the aim of this paper to provide empirical evidence for a more differentiated picture of partner selection – and thus network formation – in the European FPs. We adopt an econometric perspective to identify determinants of link formation, including various actor characteristics, relational and network effects as well as geographical effects. We employ a binary choice model estimated by means of logistic regressions, with a dependent variable that represents the establishment of a formal cooperation between two organisations in FP5. We use data on EU FP projects from the *sysres EUPRO* database and from a representative survey of FP 5 participants.

The study produces statistically significant evidence that R&D collaboration choices of organisations participating in European FPs are affected by geography, FP experience and relational factors including network characteristics. Thematic proximity matters more than geographic proximity, while most influential for collaboration appears to be prior acquaintance of the actors. Also, network effects significantly determine the collaboration choice, but to a slightly smaller extent than geographical effects.

JEL Classification: L14, R15, O38, D85
Keywords: R&D collaboration, network formation; EU Framework Programmes, binary choice model, logistic regression

1 Introduction

An organisation that decides to engage in collaborative R&D has to choose with whom to collaborate. Understanding factors that affect partner selection is crucial for understanding the dynamics of the resulting network, as they link micro-level behaviour and meso- or macro-level networks. In economics, three different theoretical strands have been discussing collaboration motives of organisations, mainly in the context of complementarity (for an overview see Paier, Ahrweiler et al. 2007): First, transaction costs theory considers inter-firm partnership as a ‘hybrid’ form of organisation that arises when transaction costs are large enough for the market to function efficiently, but not high enough to favour the integration of production in a single firm. Second, resource based theory takes into account that resources are scarce, inimitable, and only imperfectly substitutable. To gain access to resources of other organisations, a firm can establish a long-run relationship. And third, according to organisational learning theory the primary motive for collaboration is learning and joint knowledge production – a view that is in line with the recent shift of research from incentive-based to knowledge-based approaches to network formation (Pyka 2002).

In social network theory, we find two arguments that may influence an organisation´s collaborative strategy: On the one hand, high structural closure may be favourable since repetitive and redundant ties within a group of organisations create trust and help to deal with the problems of co-ordination and opportunism, and hence improve the efficiency of transactions within the group (Coleman 1988). At the same time, closure may lead to encapsulation of information and other resources within the group, and may cause greater separation between different groups, reducing the overall efficiency of the network. On the other hand, the existence of non-redundant ties generates arbitrage opportunities for the holder; this is called the structural holes argument (Burt 2001): Organisations that connect otherwise disconnected parts of the network, are able to control information flows between the groups. This argument suggests that it is attractive to take up distant partnerships, but also risky because of potentially opportunistic behaviour of the prospective partner. Jointly, the closure and the structural holes arguments are likely to generate networks with small world properties (high clustering and short average path length) (Watts and Strogatz 1998).

Another stream in the literature stresses the importance of geography for knowledge diffusion and R&D collaborations (see, for example, Almeida and Kogut 1999, Fischer, Scherngell and Jansenberger 2006). This is based on the assumption in theoretical models of New Growth Theory that different kinds of geographical space matter in the knowledge diffusion process since important parts of new knowledge have some degree of tacitness and are, thus, embedded in the routines of individuals and organisations, and difficult to transfer across space (see Romer 1990). Autant-Bernard et al. (2007a) note that the geographical dimension of knowledge diffusion deserves further attention by analysing such phenomena as R&D collaborations.

This paper aims to contribute to the existing empirical literature by focusing on interorganisational R&D collaborations as captured by data on research projects of the EU FP5 (1998-2003). The objective is to estimate how different types of external effects, including geographical effects, effects of prior FP experience and network effects, determine interorganisational collaboration choice. The study at hand departs from existing literature by widening geographic and thematic coverage as compared to other empirical studies in this field (see, for instance, Autant-Bernard et al. 2007b, Ponds, van Oort and Frenken 2007), and by using unique data drawn from a representative survey of FP5 participants, from where we select a sample of 191 organisations and create variables accounting for network and other individual effects in our empirical setting. We adopt a binary choice modelling perspective that provides a suitable analytical framework to address these questions.

The remainder of the paper is organised as follows. The next section sheds some light on R&D Networks within the EU FPs and discusses the state of the art of FP network analysis. *Section 3* introduces the empirical model that is used to identify geographical, FP experience and network effects on the collaboration choice of organisations. We specify a binary choice model that is estimated by means of a logistic regression. *Section 4* describes the construction of the dependent and the independent variables and introduces the data in some more detail, while *Section 5* discusses the estimation results derived from Maximum Likelihood estimation. *Section 6* concludes with a summary of the main results.

2 R&D networks in the European Framework Programmes

It is widely believed that interaction between firms, universities and research organisations is crucial for successful innovation in the knowledge based economy, in particular in knowledge intensive industries (see, for example, OECD 1992). Over the last few decades, technology and innovation policies of the EU have adopted this view, with the Framework Programmes (FPs) on Research and Technological Development as its central instrument for supporting collaborative R&D, with a special focus on transnational cooperation and mobility for training purposes. Since the start of FP1 (1984-1988), more than 50,000 collaborative R&D projects have been funded in the subsequent FPs. Currently, FP7 (2007-2013) is under way[1].

The EU FPs were implemented to follow two main strategic objectives: *First*, strengthening the scientific and technological bases of European industry to foster international competitiveness and, *second*, the promotion of research activities in support of other EU policies (CORDIS 2006). In spite of their different scope, the fundamental rationale of the FPs has remained unchanged. All FPs share a few common structural key elements: *First*, only projects of limited duration that mobilise private and public funds at the national level are funded. *Second*, the focus of funding is on multinational and multi-actor collaborations that add value by operating at the European level. *Third*, project proposals are to be submitted by self-organised consortia and the selection for funding is based on specific scientific excellence and socio-economic relevance criteria (see Roediger-Schluga and Barber 2006).

Mainly for reasons of data availability, attempts to evaluate quantitatively the structure and function of the large social networks generated in the EU FPs have just begun in the last few years, using social network analysis and complex networks methodologies. Breschi and Cusmano (2004), for example, reveal the emergence of a dense and hierarchical network using data from FP3 and FP4. A highly connected core of frequent participants, taking leading roles within consortia, is linked to a large number of peripheral actors, forming a giant component that exhibits the characteristics of a small world. From a policy perspective, analysis of R&D networks has to go beyond

[1] See Roediger-Schluga and Barber (2006) for a detailed discussion on the history and different scopes of the EU FPs since 1984.

analysing the characteristics of complex networks (for an overview of characteristics see da F. Costa 2005), and one has to ask for the determinants of network formation in R&D collaboration (Miotti and Sachwald 2003). Especially in the light of exploring their dynamics and governance, the determinants for partner choice are crucial and need thorough empirical investigation, as has been analysed recently for the case of Spanish firms (Arranz and Fernández de Arroyabe 2008). It is this line of research that this paper is devoted to, focusing, however, on the European perspective and on geographical and network effects.

3 The empirical model

By nature of our research questions, we seek to model organisational collaboration choices with the objective to examine how specific individual characteristics, spatial effects and network effects determine the choice of collaboration. In our analytical framework, the constitution of a collaboration Y_{ij} between two organisations i and j depends on an unobserved continuous variable Y_{ij}^* that corresponds to the profit that two organisations i and j receive when they collaborate. Since we cannot observe Y_{ij}^* but only its dichotomous realisations Y_{ij}, we assume Y_{ij} = 1 if $Y_{ij}^* > 0$ and Y_{ij} = 0 if $Y_{ij}^* \leq 0$. Y_{ij} is assumed to follow a *Bernoulli* distribution so that Y_{ij} can take the values one and zero with probabilities π_{ij} and 1- π_{ij} , respectively. The probability function can be written as

$$\Pr\left(Y_{ij}\right) = \pi_{ij}^{Y_{ij}}\left(1-\pi_{ij}\right)^{Y_{ij}} \qquad i, j = 1,\ldots, n \qquad (1)$$

with $E\left[Y_{ij}\right] = \mu_{ij} = \pi_{ij}$ and $Var\left[Y_{ij}\right] = \sigma_{ij}^2 = \pi_{ij}\left(1-\pi_{ij}\right)$ where μ_{ij} denotes some mean.

The next step in designing a model for our research purposes concerns the systematic structure, i.e. we aim to have the probabilities π_{ij} depend on a matrix of observed covariates. Thus we let the probabilities π_{ij} be a linear function of the covariates as given by

$$\pi_{ij} = \boldsymbol{X}_{ij}^{(k)}\boldsymbol{\beta} \tag{2}$$

where $\boldsymbol{X}_{ij}^{(k)}$ is the matrix containing a constant and k-1 explanatory variables, including geographical effects, relational effects and FP experience characteristics of the organisations i and j. $\boldsymbol{\beta}_k = (\beta_0, \boldsymbol{\beta}_{(k-1)})^T$ is the k-by-1 parameter vector, where β_0 is a scalar constant term and $\boldsymbol{\beta}_{(k-1)}$ is the vector of parameters associated with the k-1 explanatory variables.

However, estimating this model using OLS procedures is not convenient since the probability π_{ij} has to be between zero and one, while the linear predictor can take any real value. Thus, there is no guarantee that the predicted values will be in the correct range without imposing any complex restrictions (see Johnston and Dinardo 2007). A very promising solution to this problem is to use the logit transform of π_{ij} in the model, i.e. replacing (2) by the following ansatz:

$$Logit\left(\pi_{ij}\right) = \log\frac{\pi_{ij}}{1-\pi_{ij}} = \beta_0 + \beta_1 X_{ij}^{(1)} + \beta_2 X_{ij}^{(2)} + \ldots + \beta_K X_{ij}^{(K)} \tag{3}$$

This leads to the binary logistic regression model to be estimated given by

$$\Pr(Y_{ij} = 1 \mid \boldsymbol{X}_{ij}^{(k)}) = \pi_{ij} = \frac{e^{(\beta_0 + \beta_1 X_{ij}^{(1)} + \beta_2 X_{ij}^{(2)} + \ \ldots \ + \beta_K X_{ij}^{(K)})}}{(1 + e^{(\beta_0 + \beta_1 X_{ij}^{(1)} + \beta_2 X_{ij}^{(2)} + \ \ldots \ + \beta_K X_{ij}^{(K)})})} \tag{4}$$

The focus of interest is on estimating the parameters $\boldsymbol{\beta}_k = (\beta_0, \boldsymbol{\beta}_{(k-1)})^T$. The standard estimator for the logistic model is the maximum likelihood estimator. The reduced log-likelihood function is given by (see Johnston and Dinardo 2007)

$$\log L(\boldsymbol{\beta} \mid Y_{ij}) = -\sum_{i=1}^{n}\sum_{j=1}^{n}\log(1 + \exp[(1-2Y_{ij})\sum_{k=1}^{K}\beta_K X_{ij}^{(K)}]) \tag{5}$$

assuming independence over the observations Y_{ij}. The resulting variance matrix $\boldsymbol{V}(\hat{\boldsymbol{\beta}})$ of the parameters is used to calculate standard errors. $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically efficient when the observations of Y_{ij} are stochastic and in absence of multicollinearity among the covariates.

4 Data, variable construction and some descriptive statistics

This section sheds some light on the empirical setting of the current study and discusses in some detail the construction of the dependent and the independent variables accompanied by some descriptive statistics. We draw on different data sources for constructing the dependent and the independent variables. The core data source combines the *sysres EUPRO* database[2] and a representative survey[3] among FP5 participants[4].

The *sysres EUPRO* database presently comprises data on funded research projects of the EU FPs (complete for FP1-FP5, and about 70% for FP6) and all participating organisations. It contains systematic information on project objectives and achievements, project costs, project funding and contract type as well as on the participating organisations including the full name, the full address and the type of the organisation. In addition we use geographical information systems data to construct the geographical variables.

The survey focuses on the issues of partner selection, intra-project collaboration and output performance of EU projects on the level of bilateral partnerships (individuals as well as organisations). Restricting itself to small collaborative projects (i.e. projects with a minimum of two and a maximum of 20 partners), the survey addresses a subset of 9,107 relevant (59% of all FP5) projects. It yielded 1,686 valid responses,

---

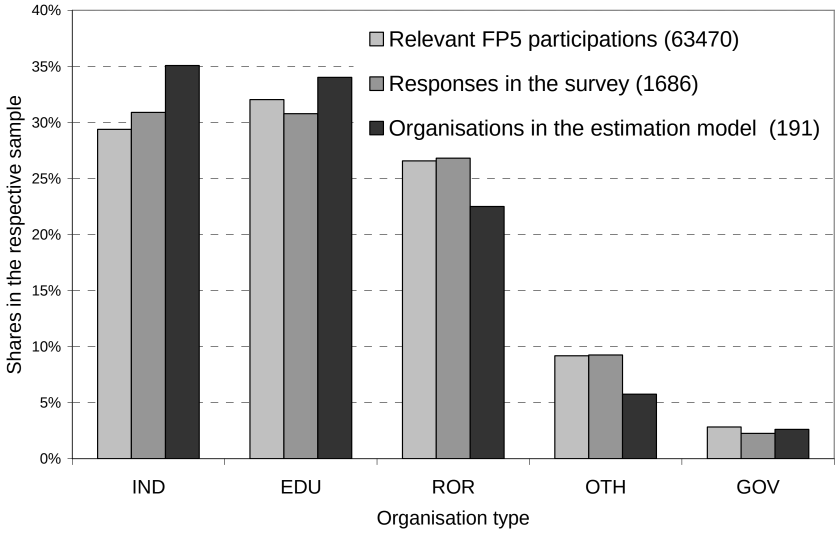
[2] The *sysres EUPRO* database is constructed and maintained by the Austrian Research Centers GmbH by substantially standardising raw data on EU FP research collaborations obtained from the CORDIS database (see Roediger-Schluga and Barber 2008, Barber et al. 2008).

[3] This survey was conducted in 2007 by the Austrian Research Centers GmbH, Vienna, Austria and operated by b-wise GmbH, Karlsruhe, Germany.

[4] We chose FP5 (1998-2002) for the survey, in order to cover some of the developments over time, including prior as well as subsequent bilateral collaborations, and effects of the collaboration both with respect to scientific and commercial outcome. Thus, the survey is able to complement the *sysres EUPRO* database.

representing 3% of all (relevant) participants, and covering 1,089 (12% of all relevant) projects.

Figure 1: Distribution of analysed entities by organisation type



Note: IND…industry, EDU…university, ROR…public research, OTH…other organisations, GOV…government.

The present study uses cross-section data on n =191 organisations that are selected from the survey data in the following way: We employ the collaboration network of the respondents on the organisation level (this network comprises 1,173 organisations collaborating in 1,089 projects) and extract the 2-*core* (deNooy, Mrvar et al. 2004) of its largest component (203 organisations representing 17% of all vertices)[5]. Finally, another 12 organisations are excluded due to non-availability of geographical distance data, so that we end up with a sample of n=191 organisations that are used to construct the binary dependent variable as described below.

The dependent variable

To construct the dependent variable Y_{ij} that corresponds to observed collaborations between two organisations i and j we construct the n-by-n collaboration matrix $\boldsymbol{Y}$ that contains the collaborative links between the (i, j)-organisations. One element Y_{ij} denotes the existence of collaboration between two organisations i and j as measured in terms of

[5] This technical trick ensures optimal utilisation of observed collaborations in the estimation model, while keeping the size of the model small. It is important to note that it does not make use of the network properties on this – somewhat arbitrary – sub-network.

the existence of a common project. ***Y*** is symmetric by construction so that $Y_{ij} = Y_{ji}$. Note that the matrix is very sparse. The number of observed collaborations is 702 so that proportion of zeros is about 98%. The mean collaboration intensity between all (i, j)-organisations is 0.02.

Variables accounting for geographical effects

We use two variables $x_{ij}^{(1)}$ and $x_{ij}^{(2)}$ to account for geographical effects on the collaboration choice. The first step is to a assign specific NUTS-2 regions to each of the $i, j = 1, \ldots, n = 191$ organisations that are given in the *sysres EUPRO* database[6]. Then we take the great circle distance between the economic centres of the regions where the organisations i and j are located to measure the geographcial distance variable $x_{ij}^{(1)}$. The second variable, $x_{ij}^{(2)}$, controls for country border effects and is measured in terms of a dummy variable that takes a value of zero if two organisations i and j are located in the same country, and zero otherwise, in order to get empirical insight on the role of country borders for collaboration choice of organisations.

Variables accounting for FP experience of organisations

This set of variables controls for the experience of the organisations with respect to participation in the European FPs. First, thematic specialisation within FP5 is expected to influence the potential to collaborate. We define a measure of thematic distance $x_{ij}^{(3)}$ between any two organisations that is constructed in the following way: Each organisation is associated with a unit vector of specialisation s_i that relates to the number of project participations $N_{i,1}, \ldots, N_{i,7}$ of organisation i in the seven sub-programmes of FP5[7].

$$s_i = \left(N_{i,1}, \ldots, N_{i,7}\right)\left(N_{i,1}^2 + \cdots + N_{i,7}^2\right)^{-1/2} \qquad i = 1, \ldots, n = 191 \qquad (6)$$

6 NUTS is an acronym of the French for the “nomenclature of territorial units for statistics", which is a hierarchical system of regions used by the statistical office of the European Community for the production of regional statistics. At the top of the hierarchy are NUTS-0 regions (countries) below which are NUTS-1 regions and then NUTS-2 regions.

7 EESD, GROWTH, HUMAN POTENTIAL, INCO 2, INNOVATION-SME, IST, and LIFE QUALITY.

The thematic distance of organisations *i* and *j* is then defined as the Euclidean distance of their respective specialisation vectors s_i and s_j,

$$x_{ij}^{(3)} = \left|s_i - s_j\right| = \left[\left(s_{i,1} - s_{j,1}\right)^2 + \cdots + \left(s_{i,7} - s_{j,7}\right)^2\right]^{1/2} \qquad i, j = 1, \ldots, n = 191 \qquad (7)$$

with $x_{ij}^{(3)} = x_{ji}^{(3)}$ and a value ranging from $0 \le x_{ij}^{(3)} \le \sqrt{2}$. The second variable accounting for FP experience focuses on the individual (or research group) level, and takes into account the respondent´s inclination or openness to FP research. As a proxy for openness of an organisation to FP research, we choose the total number P_i of FP5 projects in the respondents' own organisation, that they are aware of[8]. Then we define

$$x_{ij}^{(4)} = P_i + P_j \qquad i, j = 1, \ldots, n = 191 \qquad (8)$$

as a measure for the aggregated openness of the respective pair of organisations to FP research. The third variable related with FP experience is the overall number of FP5 project participations an organisation is engaged in. Denoting, like above, $N_i = N_{i1} + ... + N_{i7}$ as the total number of project participations of organisation *i* in FP5, we define

$$x_{ij}^{(5)} = \left|N_i - N_j\right| \qquad i, j = 1, \ldots, n = 191 \qquad (9)$$

as the difference in the number of participations of organisation *i* and *j* in FP5. It is taken from the *sysres EUPRO* database and is an integer ranging from $0 \le x_{ij}^{(5)} \le 1{,}228$, resulting from the minimal value of one participation and the maximum of 1,229 participations among the sample of 191 organisations.

Variables accounting for relational effects

We consider a set of three variables accounting for potential relational effects on the decision to collaborate. Hereby, we distinguish between 'joint history' and network

[8] The exact wording of the question was, 'How many FP5 projects of your organisation are you aware of?' For multiple responses from an organisation, the numbers of known projects are summarised. In cases of missing data, this number is set to zero.

effects. The first factor to be taken into account is prior acquaintance of two organisations, and is measured by a binary variable denoting acquaintance on the individual (research group) level before the FP5 collaboration started. It is taken from the survey[9]. By convention, $x_{ij}^{(6)} = 1$, if at least one respondent from organisation $i(j)$ nominated organisation $j(i)$ as prior acquainted, $x_{ij}^{(6)} = 0$ otherwise. All other relational factors we take into account in the model are network effects. For conceptual reasons we have to look at the global FP5 network, where we make use of the structural embeddedness of our 191 sample organisations. One of the most important centrality measures is betweenness centrality. Betweenness is a centrality concept based on the question to what extent a vertex in a network is able to control the information flow through the whole network (Wasserman and Faust 1994). Organisations that are high in betweenness, may thus be especially attractive as collaboration partners. More formally, the betweenness centrality of a vertex can be defined as the probability that a shortest path between a pair of vertices of the network passes through this vertex. Thus, if $B(k,l;i)$ is the number of shortest paths between vertices k and l passing through vertex i, and $B(k,l)$ is the total number of shortest paths between vertices k and l, then

$$b(i) = \sum_{k \neq l} \frac{B(k,l;i)}{B(k,l)} \tag{10}$$

is called the betweenness centrality of vertex i (Dorogovtsev and Mendes 2004). We calculate the betweenness centralities in the global FP5 network and include

$$x_{ij}^{(7)} = b(i)\, b(j) \qquad i, j = 1, \ldots, n = 191 \tag{11}$$

as a combined betweenness measure. The third variable accounting for relational effects is local clustering. Due to social closure, we may assume that within densely connected clusters organisations are mutually quite similar, so that it might be strategically advantageous to search for complementary partners from outside. Hereby, communities with lower clustering may be easier to access. We use the clustering coefficient CC_1,

9 The exact wording of the question was, ‘Which of your [*project acronym*] partners (i.e. persons from which organisation) did you know before the project began?’

which is the share of existing links in the number of all possible links in the direct neighbourhood (at distance $d = 1$ of a vertex). Thus, let k_i be the number of direct neighbours and T_i the number of existing links among these direct neighbours, then

$$CC_1(i) = T_i \frac{2}{k_i(k_i - 1)} \tag{12}$$

is the relevant clustering coefficient (Watts and Strogatz 1998). We employ the difference in the local clustering coefficient CC_1 within the global FP5 network for inclusion in the statistical model, by setting

$$x_{ij}^{(8)} = \left| CC_1(i) - CC_1(j) \right| \qquad i, j = 1, \ldots, n = 191 \tag{13}$$

in order to obtain a symmetric variable in *i* and *j*.

Summing up, the three groups of independent variables are geographical effects, FP experience and relational effects. A maximum of eight variables are included in the estimation model. Table 1 reviews the defined variables, and provides some descriptive statistics.

Table 1: List of independent variables and some descriptive statistics

Variable name	Variable	Mean	Standard Deviation	Min	Max
Geographical effects					
Geographical distance	$x_{ij}^{(1)}$	1,010.01	585.13	0	3,409
Country border effects	$x_{ij}^{(2)}$	0.12	0.34	0	1
FP experience					
Thematic distance	$x_{ij}^{(3)}$	0.89	0.39	0	1.41
Openness to FP research	$x_{ij}^{(4)}$	28.86	42.43	0	508
Difference in number of participations	$x_{ij}^{(5)}$	92.53	156.78	0	1,228
Relational effects					
Prior acquaintance	$x_{ij}^{(6)}$	0.04	0.20	0	1
Betweeness centrality	$x_{ij}^{(7)}$	0.31	0.72	0	6.41
Difference in local clustering	$x_{ij}^{(8)}$	0.35	0.31	0	0.97

*Note that country border effects and prior acquaintance are dummy variables.

5 Estimation results

This section discusses the estimation results of the binary choice model of R&D collaborations as given by specification (4). The binary dependent variable corresponds to observed collaborations between two organisations i and j, taking a value of one if they collaborate and zero otherwise. The independent variables are geographical separation variables, variables capturing FP experience of the organisations and relational effects (joint history and network effects). We estimate three model versions: The standard model includes one variable for geographical effects and FP experience, respectively, and two variables accounting for relational effects. In the extended model version we add country border effects as additional geographical variable in order to isolate country border effects from geographical distance effects, and openess to FP research as additional FP experience variable. The full model additionally includes balance variables accounting for FP experience and network effects, respectively.

Table 2 presents the sample estimates derived from Maximum likelihood estimation for the model versions. The number of observations is equal to 36,481, asymptotic standard errors are given in parentheses. The statistics given at the bottom of Table 2 indicate that the selected covariates show a quite high predictive ability. The Goodman-Kruskal-Gamma statistic ranges from 0.769 for the basic and 0.782 for the extended model to 0.786 for the full model, indicating that more than 75% fewer errors are made in predicting interorganisational collaboration choices by using the estimated probabilities than by the probability distribution of the dependent variable alone. The Somer´s D statistic and the c index confirm these findings. The Nagelkerke´s R-Squared is 0.391 for the basic model, 0.395 for the extended model and 0.397 for the full model version, respectively[10]. A likelihood ratio test for the null hypothesis of $\boldsymbol{\beta}_k = 0$ yields a χ_4^2 test statistic of 2,565.165 for the basic model, a χ_6^2 test statistic of 2,582.421 for the extended model and a χ_8^2 test statistic of 2,597.911 for the full model. These are

[10] Nagelkerke´s R-squared is an attempt to imitate the interpretation of multiple R-Squared measures from linear regressions based on the log likelihood of the final model versus log likelihood of the null model. It is defined as $R^2_{Nag} = [1-(L_0/L_1)^{2/n}]/1-L_0^{2/n}$ where L_0 is the log likelihood of the null model, L_1 is the log likelihood of the model to be evaluated and n is the number of observations.

statistically significant and we reject the null hypothesis that the model parameters are zero for all model versions.

Table 2: Estimation results of the choice of collaboration model (n^2 = 36,481 observations; asymptotic standard errors in brackets)

Coefficient	ML estimates		
	basic	extended	full
Intercept	-1.882*** (0.313)	-1.951*** (0.342)	-1.816*** (0.385)
Geographical effects			
Geographical distance [β_1]	-0.145*** (0.038)	-0.116*** (0.039)	-0.128*** (0.040)
Country border effects [β_2]	–	-0.103*** (0.034)	-0.094** (0.034)
FP experience			
Thematic distance [β_3]	-1.477*** (0.110)	-1.465*** (0.114)	-1.589*** (0.117)
Openness to FP research [β_4]	–	0.004*** (0.001)	0.003*** (0.001)
Difference in number of participations [β_5]	–	–	0.001 (0.000)
Relational effects			
Prior acquaintance [β_6]	4.224*** (0.089)	4.189*** (0.089)	4.194*** (0.089)
Betweeness centrality [β_7]	0.161*** (0.023)	0.135*** (0.025)	0.119*** (0.027)
Difference in local clustering [β_8]	–	–	0.070** (0.025)
Model Performance			
Somer´s D	0.733	0.746	0.753
Goodman Kruskal Gamma	0.769	0.782	0.786
C index	0.876	0.873	0.875
Nagelkerke R-squared index	0.391	0.395	0.397
Log-Likelihood	-2190.151	-2176.768	-2169.578
Likelihood Ratio Test	2,565.165***	2,582.421***	2,597.911***

Notes: The dependent variable is binary corresponding to observed collaborations between the (i, j)-organisation pairs, the independent variables are defined as given in the text. Parameter estimates are derived from Maximum Likelihood estimation. ***significant at the 0.001 significance level, **significant at the 0.01 significance level, *significant at the 0.05 significance level

The model reveals some promising empirical insight in the context of the relevant literature on innovation as well as on social networks. The results provide a fairly remarkable confirmation of the role of geographical effects, FP experience effects and network effects for interorganisational collaboration choice in EU FP R&D networks. In general, the parameter estimates are statistically significant and quite robust over different model versions.

The results of the basic model show that geographical distance between two organisations significantly determines the probability to collaborate. The parameter estimate of β_1 = -0.145 indicates that for any additional 100 km between two organisations the mean collaboration frequency decreases by about 15.6 percent. Geographical effects matter but effects of FP experience of organisations are more important. As evidenced by the estimate β_3 = -1.477 it is most likely that organisations choose partners that are located closely in thematic space. A one percent increase in thematic distance reduces the probability of collaboration by more than 3.25 percent. Most important determinants of collaboration choice are network effects. The estimate of β_6 = 4.224 tells us that the probability of collaboration between to organisations increases by 68.89 percent when they are prior acquaintances. Also network embeddedness matters as given by the estimate for β_7 = 0.161 indicating that choice of collaboration is more likely between organisations that are central players in the network with respect to betweeness centrality.

Turning to the results of the extended model version it can be seen that taking into account country border effects decreases geographical distance effects by about 24 percent (β_1 = -0.116). The existence of a country border between two organisations has a significant negative effect on their collaboration probability, the effect is slightly smaller than geographical distance effects (β_2 = -0.103). Adding openness to FPs as an additional variable to capture FP experience does not influence the other model parameters much. Openness to FPs shows – though significant – only a very small impact on collaboration choice.

In the full model version we add one balance variable accounting for FP experience and network effects, respectively. The difference in the number of submitted FP projects has virtually no effect on the choice of collaboration as given by the estimate of β_5. An interesting result from a social network analysis perspective provides the integration of the difference between two organisations with respect to the clustering coefficient (see Section 4). The estimate of β_8 = 0.070 tells us that it is more likely that two organisations collaborate when the difference of their cluster coefficients is higher. This result points to the existence of strategic collaboration choices for organisations that are highly cross-linked searching for organisations to collaborate with lower clustering coefficients, and the other way round. The effect is statistically significant but smaller than other network effects and geographical effects.

6 Concluding remarks

The focus of this study is on the collaboration choice of organisations participating in EU FP5 research projects. We shift attention to the role of geography, FP experience and relational factors including network characteristics for the choice of an organisation who to collaborate with. The objective was to measure the impact of these groups of determinants on interorganisational collaboration choice. We used a binary choice model with a latent dependent variable estimated by means of logistic regressions as analytical framework to address these questions.

The study produces some interesting results in the context of the empirical literature on R&D networks in Europe. In contrast to the study of Autant-Bernard (2007b) that investigates collaboration choice in a very specific thematic field (micro- and nanotechnologies) our results show that interorganisational R&D collaborations are significantly determined by geographical effects. Collaborative activities are more likely to occur between organisations that are not too far from each other in geographical space. Also the existence of a country border between two organisations significantly decreases the probability that they collaborate.

Geographical effects matter, but more important determinants of collaboration choice are FP experience and relational characteristics including network effects. The results indicate that it is most likely that organisations choose partners that are similar with respect to their thematic profile in the FP. The other measures of FP experience, however, namely individual openness to FP research and difference with respect to the number of project participations, do not influence partner choice.

The results reveal that the most important factors of collaboration choice are relational effects, predominantly joint history: The probability of collaboration between two organisations highly increases when they know each other from prior collaborative activities. Furthermore, network effects significantly affect the collaboration choice of organisations, as for instance, network embeddedness and difference between organisations in local clustering, but such network effects are slightly smaller than geographical effects.

The study raises some points for a future research agenda. *First*, it would be of interest to estimate the impact of these determinants on the collaboration choice for different types of organisations. For instance, it can be assumed that industry participants show different behaviour concerning collaboration choice than organisations from public research. *Second*, it would be promising to specify a model that takes into account the project substructure, such as intensive R&D collaboration, in addition to the use of the formal relationships represented by the cooperation contract, which is current state of the art.

Acknowledgements. The work presented here is partly funded by the EU-FP6-NEST project NEMO ('Network Models, Governance and R&D collaboration networks'), contract number 028875

References

Almeida, P. and Kogut, B. (1999): The Localization of knowledge and the mobility of engineers in regional networks, *Management Science* 45(7), 905-917

Arranz, N. and J. C. Fernández de Arroyabe (2008). The choice of partners in R&D cooperation: An empirical analysis of Spanish firms. *Technovation* 28: 88–100.

Autant-Bernard, C., Mairesse, J. and Massard, N. (2007a): Spatial knowledge diffusion through collaborative networks, *Papers in Regional Science* 86, 341-350

Autant-Bernard, C., Billand, P., Frachisse, D. and Massard, N. (2007b): Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies, *Papers in Regional Science* 86, 495-519

Barber, M., Heller-Schuh, B., Roediger-Schluga, T., and Scherngell, T. (2008): The sysres EUPRO database manual. NEMO Deliverable D4.1. Vienna.

Barker, K. and Cameron, H. (2004): European Union science and technology policy, RJV collaboration and competition policy. In Caloghirou, Y., Vonortas, N.S. and Ioannides, S. (eds.): *European Collaboration in Research and Development*, Cheltenham, Edward Elgar, 154-186

Batagelj, V. and A. Mrvar (2005): *Pajek - Program for Analysis and Visualization of Large Networks. Reference Manual.* Ljubljana, Slovenia.

Breschi, S. and Cusmano, L. (2004): Unveiling the texture of a European Research Area: Emergence of oligarchic networks under EU Framework Programmes, *International Journal of Technology Management* 27(8): 747-772.

Burt, R. S. (2001): Structural Holes versus Network Closure as Social Capital. In N. Lin, K. S. Cook and R. S. Burt (eds.): *Social Capital: Theory and Research.* Aldine de Gruyter.

Coleman, J. S. (1988): Social Capital in the Creation of Human Capital. *The American Journal of Sociology* 94 (Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure): pp. S95-S120.

Cowan, R. (2004): Network models of innovation and knowledge diffusion, Unu-MERIT working paper series, 2004-016

CORDIS (2006): FP5 - EC Programme: Maximum Amounts and Breakdown (1998-2002), available from http://cordis.europa.eu/fp5/src/budget.htm

da F. Costa, L., F. A. Rodrigues, G. Travieso, and Villas Boas, P. R. (2005): Characterization of Complex Networks: A Survey of measurements. *arXiv:cond-mat/0505185 v3* 30 Jun 2005.

deNooy, W., Mrvar, A. and Batagelj, V. (2004): *Exploratory Social Network Analysis with Pajek,* Cambridge University Press.

Dorogovtsev, S. N. and Mendes, J. F. F. (2004): The shortest path to complex networks. *arXiv:cond-mat/0404593v4 [cond-mat.stat-mech]* 24 Jul 2004.

Fischer, M.M., Scherngell, T. and Jansenberger, E. (2006): The geography of knowledge spillovers between high-technology firms in Europe. Evidence from a spatial interaction modelling perspective, *Geographical Analysis* 38, 288-309

Johnston, J. and Dinardo, J. (2007): *Econometric Methods*. Fourth Edition, New York, McGraw-Hill

Miotti, L. and Sachwald, F. (2003). Co-operative R&D: why and with whom? An integrated framework of analysis. *Research Policy* 32(8): 1481-1499.

Paier, M., Ahrweiler, P., Barber, M., Brandes, A., Heller-Schuh, B., Jonard, N., Kutz, M., Pyka, A., Sanditov, B., Wagner-Luptacik, P. and Weber, M. (2007): Conceptual and empirical foundations of R&D network dynamics. NEMO Deliverable Report D1.1. Austrian Research Centers GmbH, Vienna.

Ponds, R., van Oort, F. and Frenken, K. (2007): The geographical and institutional proximity of research collaboration, *Papers in Regional Science* 86, 423-443

Pyka, A. (2002): Innovation networks in economics: from the incentive-based to the knowledge-based approaches. *European Journal of Innovation Management* 5(3): 152-163.

Roediger-Schluga, T. and Barber, M. (2006): The structure of R&D collaboration networks in the European Framework Programmes, Unu-MERIT working paper series, 2006-36

Roediger-Schluga, T. and Barber, M. (2008): R&D collaboration networks in the European Framework Programmes: Data processing, network construction and selected results, *International Journal of Foresight and Innovation Policy* 4, 321–347

Romer, P.M. (1990): Endogenous technological change, Journal of Political Economy 98, 71-102

Wasserman, S. and Faust, K. (1994): *Social Network Analysis -- Methods and Applications*. Cambridge, Cambridge University Press

Watts, D. J. and Strogatz, S. H. (1998): Collective dynamics of 'small-world' networks. *Nature* 393(6684): 440-442