Prediction of firm's decision to innovate by means of rough sets

Javier Saiz Briones y María Jesús Segovia Vargas

Department of Financial Economics and Accounting I, Universidad Complutense de Madrid. Campus de Somosaguas. 28223 Madrid, Spain.

Abstract

This paper tries to further investigate the factors behind the decision to innovate. By using a large sample of manufacture firms in the period 2000 to 2005, it is intended to apply a method coming from the Artificial Intelligence (Rough Sets theory) and analyze the role of a set of internal, external and innovative variables in explaining firm's decision.

This technique presents some advantages: it is a useful tool to analyse information systems representing knowledge gained by experience; through this analysis the elimination of the redundant variables is got, then the cost of the decision making process and time employed by the decision maker are reduced; the analysis process results in a model consisting of a set of easily understandable decision rules supported by a set of real examples allowing to argue the decisions made using these rules. Moreover, this method does not need that variables satisfy any assumptions. Statistical methods need that the explicative variables satisfy statistical assumptions what is quite difficult to happen. This fact complicates the analysis. We have obtained quite good results showing the suitability of this methodology.

Keywords: artificial intelligence, rough sets, innovation

JEL: O3

1.- INTRODUCTION

Innovation literature has focus on analyze the relationship between technological activity and different firm characteristics. Usually, those studies use only one characteristic each time for developing their analysis.

In this article we develop the model including the most important factors which influence in innovation process. Therefore, we can contrast the accuracy effect of each one on R&D process. It can be obtained which ones have influence without excluding other ones that are relevant too.

This article has 6 sections. The next one is focus on theoretical literature which explains firm innovation process. Section 3 explains the variables and the data use for contrasting the hypothesis. The next one develops the statistical model for estimating. In section 5 we present the main results and the last one shows the conclusions.

2. LITERATURE REVIEW.

From Schumpeter (1934) to National System of Innovation and the evolution theory (Nelson and Winter, 1982), Innovation has been presented in the economic thought as playing a crucial role especially in macroeconomic and growing theories.

Schumpeter.

Schumpeter conceives innovation like a brake up with the past organization. It means a change in productive framework that pushs up economic growth using a new combination of productive factors. It's a "destructive creation" process where the new is substituting the old. In that scene, businessmen are substituted by new ones, and from an economic point of view it means a resources reallocation (Schumpeter, 1942).

The introduction of innovation means a reduction of firm's costs or an increasing of goods price produced by the company that introduced it. However, the difference between incomes and expenditures of an innovative enterprise is bigger than in the non innovative one. Thus, that implies an improvement of profits. Later on, innovation expands to the economy leads to a continuous reduction of profits until making them close to cero. That eradicates the additional component of them. In case that innovation does not expand to the whole system, benefits become a monopoly income.

Evolution Theory.

Evolution theory sets the start up idea that technology is knowledge. It resides in firms that develop it and use it. The assimilation of this knowledge is only possible through study and experience.

This theory was developed by Nelson and Winter (1982) and their idea is that technological innovation is the result of searching solutions to production problems developing a learning process. However, Dosi (1982) develops implications of this idea[1].

Evolution authors emphasize sources of learning can be internal and external and say that there are very important differences between firms. Coriat and Dosi (1994) suggest internal capacities of a firm are affected by its ability for innovating and developing new technologies. However, Dosi develops a model where innovation firm behavior is related with its structure, environment and technological sector.

Sectoral patterns of technical change.

Taking into account the diversity of technological and learning processes, all the ways of developing technology and the different technological opportunities, Pavitt[2] (1984) developed taxonomy of sectoral patterns of technical change. It is based on the idea that there is not only one model of innovation development for all firms. In fact, there is more than one pattern that can be used for very similar enterprises.

There are four types of sectors as we can see in table 1:

- The first one includes those sectors whose innovation is based on suppliers, especially those who provide capital assets, materials or components and technological advice services. Those firms are located in agriculture, construction and market services sectors. The main purpose of these firms is cost reduction for competing on markets where customers are highly price sensitive.

- The second one is referred to those intensive production sectors that have economies of scale and strong division of labor (automobile, durable goods…). Their technology is provided from external sources like machinery suppliers. It is also internal developed in R&D department. Therefore, enterprise contributes to knowledge generation. Market demand is very elastic so technology is oriented to production process improving, reducing cost, and also it focuses on product design differentiation.

---

1 For a review of the evolution theory see Arena and Lazaric (2003) and Dosi and Winter (2003).

2 Very different teorical formulation are developed for helping generate patterns like technological paradigm and technological path of Dosi (1982), (1984), (1988) and Pavitt (1984), techno-economical paradigm of Pérez (1983), Freeman and Pérez (1988), technological regime of Nelson and Winter (1982), Winter (1984), Malerba and Orsenigo (1990) and Technological guidepost and innovation avenues of Sahal (1985).

- The third one includes that firms that provide to other sectors. They are highly specialized and manufacture instruments and capital assets. Demand looks for assets oriented to a very specific necessity who incorporate a very particular technology use for a very specific production process. Technological path is focus on obtaining differentiated products with a high level of design.

- The last one is referred to those firms who produce scientific knowledge, goods such as electric and electronic machines, chemical industry. Demand is price sensitive, moreover new developed products create new markets for them. Technological change is due to internal factors. Firms use it to make new internal laboratories where they develop knowledge needed to acquire and adjust external scientific advances.

Table 1. Pavitt Taxonomy of Innovation Patters.

sector type/ variables	Science Based	Scale Intensive	Supplier dominated	Specialised suppliers
Firm size	Large firms	Large firms	Large firms	Large firms
Type of innovation	Mixed products and processes	Processes	Processes	Products
Locus of innovation	R&D departments	Production	External	Decentralized
Sources of innovation	Universities and research centers	Production and specialized suppliers	Specialized suppliers	Science based firms/customers
Means of appropriability	Patents and entry barriers	Tacit knowledge and entry barriers	Tacit knowledge	Tacit knowledge/reputation
Competitive parameter	Performance/ quality/ price	Price/quality	Price	Quality/ performance
Sectors	Pharmaceuticals/ microelectronics	Automobiles	Stone/ glass	Mechanical engineering
Trajectories	Performance and physical properties and improvements in process yields	Improvements in process yields and increases in the scale of production processes and automation.	Improvements in process yields	Improvements in performance and reliability of products
Learning regime	Learning by searching/ learning by doing	Learning by doing/ learning by using	Learning by using	Learning by interacting/ learning by doing

Schumpeterian patterns of innovation.

Malerba and Orsenigo base on Nelson and Winter ideas developed a theory for explaining technological diversity. It's called *schumpeterian patterns of innovation*[3]. They are defined like industrial organization models, based on technological activity. Schumpeterian patterns are made taking into account technological opportunity, apropiability, cumulative conditions and technological knowledge sources[4]. They can be defined like:

- Technological opportunity is referred to how easily a firm obtains innovations given a certain level of resources.

- Apropiability is related with how innovators capture results and benefits of their innovation, moreover , how protect their output against imitators.

- Technological knowledge sources define learning processes and how they develop innovative activity (basic and applied). It refers to underlying knowledge of technological activity.

- Cumulative conditions is related to the higher probability that an actual innovator has of going on with being an innovator on the future related with non innovative enterprises.

Using patent data for 49 technology classes in 6 countries, Malerba and Orsenigo developed a sectoral pattern and they made up two regimes:

- The first one called *Schumpeter Mark I* is made by sectors with low innovative activity. Innovator size is small. There is a high firm rotation with lots of firms coming in and out. The main reason is due to a great number of technological opportunities that do not need a high research effort. Furthermore, results appropiability is very difficult to get and there is a low knowledge cumulativity too.

- In *Schumpeter Mark II* there is a relative high concentration in innovative activities. Innovative enterprises are big and they have a high level of organization in this task too. Thus, there is a low rotation of innovative firms

---

[3] Malerba and Orsenigo, (1990), (1995) and (1996).

[4] Many authors show how technological opportunity, apropiability and technological knowledge sources influence the way innovation is developed. See Nelson and Winter (1982), Kamien and Schuwartz (1982), Winter (1984), Dosi (1984) and Cohen and Levin (1989a).

in the market. This is due to technological condition in opportunity levels, apropiability and cumulativity of knowledge. They all are very high and make advantages for the first one who came into the sector. It develops barriers to entry in the market.

Patterns of innovation.

The third theory summarizes the technology behavior of firms basing on Knowledge bases, resources, product and process orientation, relations with other agents and appropiability conditions. It is the *patterns of innovation* theory. There is not an agreement on the empirical results but they are very close[5].

Only a few studies deal with innovation using the sectoral patterns of technological change and it can be seen on table 2. As we can see, there are different types of indicators and, therefore, methodology used to identify them is different too.

Table 2. Empirical studies of patterns.

Author and year	Country	Data	Type of indicator for identify patterns	statistical technique	Type of taxonomy and number of patterns
Pavitt (1984)	England	SPRU databank of British innovations.	Technological and Structural	not explained	Sectoral 3 patterns
Napolitano (1991)	Italy	ISTAT survey	Technological	Correlation matrix	Sectoral 3 patterns
Archigugi et al. (1991)	Italy	CNR-ISTAT suvery	Technological and Structural	not explained	Sectoral 5 patterns
Molero and Buesa (1996)	Spain	Authors' designed survey	Technological and Structural	Cluster analisys	Business 7 regimes
Baldwin and Johnson (1997)	Canada	Survey of Growing Small and Medium-Size Enterprises and balance data.	Technological	factoral analisys	Business 3 types
Coombs and Tomlinson (1998)	England	CBI Innovation Survey	Technological	factoral analisys	Business 3 types
Buesa and Zubiarre (1998)	Spain	ESTE-Eusko Ikaskuntza survey	Technological and Structural	Cluster analisys	Business 6 patterns
Fonfria (1998)	Spain	IAIF-CDTI suvery	Technological	Factoral and cluster	Business 6 patterns

5 See Buesa and Molero (1992) (1996), Buesa and Zubiaurre (1999), Zubiaurre (2000) and Galende del Canto (2003).

				analisys	
Arvanitis and Hollestein (1998)	Switzerland	KOF/ETH Survey	Technological	Cluster analisys	Business 5 types
Urraca (2000)	Spain	ESEE Survey	Technological and Structural	not explained	Sectoral 4 patterns
Hollestein (2003)	Switzerland	Swiss innovation survey	Technological	Factoral and cluster analisys	Business 5 types
Raymond et al. (2004)	Holand	CIS 2, 2.5 and 3. Production survey and Finance suvery.	Technological and Structural	Tobit Model	Sectoral 3 patterns
Peeters et al. (2004)	Belgium	CIS 3.	Technological	Factoral and cluster analisys	Business 8 patterns

National/Regional Systems of Innovation.

A National System of Innovation can be defined like a group of institutional and business organizations inside a territory that interact between them with the main purpose of allocate resources to activities orientated to generation and diffusion of technological knowledge. Those organizations are, in a Schumpeterian sense, the foundation of economic development (Buesa, M., 2005).

This concept was used on the first time by Freman (1988) and has been developed in the last decade. The main empirical studies that work on this topics are Lundvall (1992), Nelson (1993), Edquist (1997) and OECD (1999) and from a global point of view Braczyck, Cooke and Heidenreich (1998) and Koschatzky, Kulicke and Zenker (2000). Studies of the Spanish case are Heijs (2001) Navarro (2002) and Martínez Pellitero (2002).

Lundavll (1992) shows five different components of a National System of Innovation:

1. *The internal organization of firms.* It is a very important part of innovations generation.

2. *Relations between companies.* They are crucial for allowing knowledge transfer.

3. *Public Sector.* It drives quantity, quality and direction of investigation, and, innovation development too.

4. *Institutional framework.* It has a prominent role inside NSI.

5. *The organization and intensity of R+D.* Innovation processes are linked with it.

Nelson (1993) starts his empirical study with some assumptions:

1. NSI are wide structures composed of all factors that can affect national technological capacities.

2. Technological change is very significant and it is totally dependent of sciences, technologies and connections between them in a very narrow way. These factors drive a change on institutions and firms that set the system.

3. The main agents of the system are firms, industries, research laboratories, universities and public units of research. They set the core of the system. Technological advances are the results of interaction between them.

It is very important to introduce all determinants of innovation: the economics ones, the social ones, the politics ones and the institutional ones when you are specifying a NSI (Edquist, 1997).

Buesa (2005) recommends structure NSI in four blocks. It can be seen in table 3:

- **Economic and productive framework:** Here is where all organization of the system is described. The variables are referred to market size, economic development… They determine the allocation of resources used for creation of knowledge and production activities too.

- **Scientific Research:** It is made by Universities and Public research Institutions. They are responsible of abstract knowledge, evolution and improvement of labor force qualification.

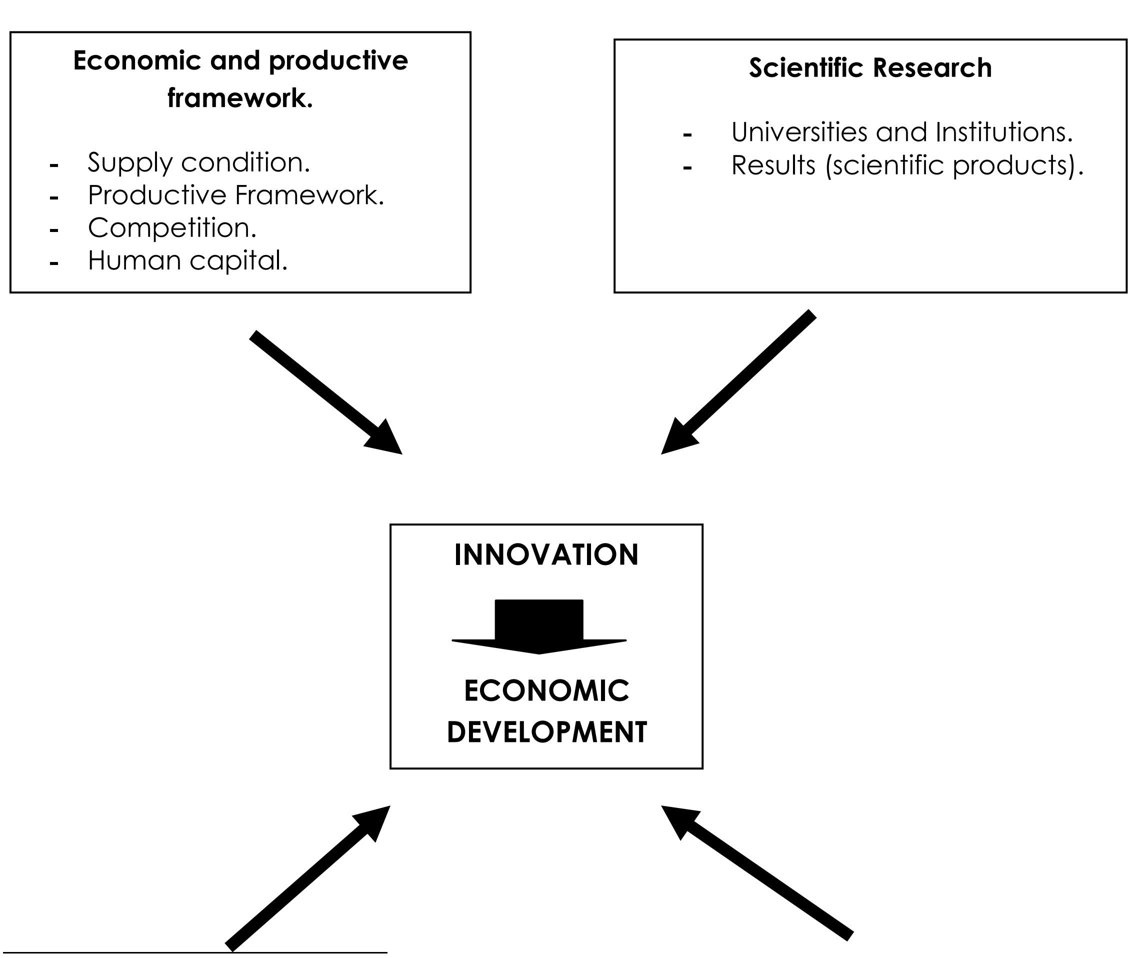
- **Innovative Firms:** they develop a special role in technology development from creation of new technologies to their adoption and diffusion.

- **Support politics and institutions:** They correct market failures due to special characteristics of technological good, particular in reallocation of resources and knowledge creation activities. There are institutions helping apropiability of innovation results. Also, promote the cooperation between organizations and agents of the innovation system.

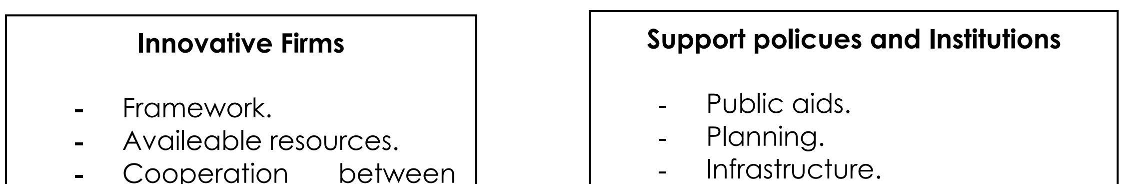
As we can see, SNI get away neoclassical equilibrium that used to look for a static optimum. SNI propose that there is a dynamic equilibrium as a result of resources' reallocation[6].

In the present research we have made a literature review concerning innovation processes inside the firm. We found two types of research. Ones focus on the internal factors of the enterprise that make innovation possible. Others, contrary, focus on the external characteristics of the firm that influence on innovation activity.

Table 3. Framework of National (Regional) Systems of Innovation.



[6] Buesa, M. (2005). For a detail description see Edquist (1997).



Taking into account all the literature we can make three clusters. The first one is related to internal characteristics of the firm. As we can see in **table 4** are: size, industrial property, company team, organizational resources and debt. The second one resumes market characteristics like internationalization, market concentration, commercial resources and economic opportunity. The last one shows innovation characteristics as human resources, membership of a Regional System of Innovation, appropiability, diversification and technological opportunity.

Table 4. Internal and external determinants of innovation.

	DETERMINANT	TEORICAL FRAMEWORK	EMPIRICAL STUDIES
Firm features	SIZE	LARGE: It can be obtained economies of scale. Risk became lower. There is a bigger market and better appropiability. SMALL: There is more flexibility, better communication and high specialization.	LARGE: Horowitz (1962), Lunn and Martin (1986), Braga and Willmore (1991) and Gumbau (1997). SMALL: Acs and Audrestch (1990), Grabowski (1968), Adams (1970) and Graves y Langowitz (1993). U FORM: Bound et al. (1984), Rothwell (1986), Rotwell and Dogson (1994). INVERTED U FORM: Scherer (1965b), Smith (1974).
	INDUSTRIAL PROPERTY	POSITIVE: bigger result appropiability.	IT DEPENDS ON SECTOR: Mansfield et al. (1981), Levin et al. (1987), Harabi (1995), Pavitt (1984), Malerba and Orsenigo (1990).
	INDUSTRY GROUP	POSITIVE: there is more resources and better technological information. NEGATIVE: There is innovation centralization and lower flexibility.	POSITIVE: Busom (1991), Braga and Willmore (1991). NEGATIVE: Merino y Salas (1995).
	ORGANIZATION RESOURCES	POSITIVE: There is better coordination and communications. Firm's departments are better integrated.	POSITIVE: Freeman (1973), Rothwell et al. (1974), Rothwell (1986), Bughin and Jacques (1994), Kumar and Saquib (1996), Gumbau (1997) and Kuemmerle (1998), Huergo and Jaumandreu (2004).
	DEBT	NEGATIVE: It is because of high risk of technology and asymmetric information.	NEGATIVE: Long and Ravenscraft (1993), Mansfield (1965), Minasian (1969), Bailey (1972), Branch (1974), Grabowski and Mueller (1978), Ben-Zion (1978), Sougiannis (1994). POSITIVO: Scherer (1965a).
Market features	INTERNACIONALIZATIÓN	POSITIVE: There is more competition and bigger market	POSITIVE: Vernon (1966), Bijaoui (1985), Brouwer and Kleinknecht (1993), Sterlacchini (1999), Wakelin (1998), Roper and Love (2002), Basile (2001), Kumar and Siddhartan (1994), Braunerhjelm (1996), Bernard and Jensen (1999), Anderton (1999a,b), Salomón and Shaver (2002), Lunn and Martin (1986), Braga and Willmore (1991), Busom (1991), Kumar and Saqib (1996), Galende and Suárez (1998, 1999), Grossman and Helpman (1991, 1993), Kogut (1991), Nelson (1993) and Almeida and Kogut (1999).
	MARKET CONCENTRATION	POSITIVE: Better result's apropiability. Bigger resources and lower risk. NEGATIVE: Lower competition	POSITIVE: Horowitz (1962), Dasgupta and Stiglitz (1980) NEGATIVE: Link (1982), Acs and Audretsch (1988).
	COMERCIAL RESOURCES	POSITIVE: There is better reputation and brand, also more information.	POSITIVE: Freeman (1973), Rothwell et al. (1974), Doi (1985), Lunn and Martin (1986), Gumbau (1997).
	ECONOMIC OPORTUNITY	POSITIVE: There is bigger market and high possibility of growth.	POSITIVE: Schmookler (1962, 1966), Lunn and Martin (1986) and Gumbau (1997).
Innovative features	HUMAN RESOURCES	POSITIVE: Better qualification, higher experiences and better innovation development.	POSITIVE: Galende and Suarez (1998, 1999), Martínez-Ros and Salas (1995).
	REGIONAL SYSTEM OF INNOVATION	POSITIVE: spillover effects. Regional policy effectiveness.	VARY WITH SYSTEM: Pavitt (1984), Pavitt el tal. (1989), Malerba and Orsenigo (1990), Archibugi et al. (1991), De Marchi et al (1996), Molero and Buesa (1996), Pomares (1998), Urraca (2000).
	RESULTS' APPROPIABILITY	POSITIVE: There is no spillovers NEGATIVE: on imitative innovations	POSITIVE: Levin et al. (1985), Stadler (1992), Cohen and Levinthal (1989a, 1989b) and Gumbau (1997).
	DIVERSIFICATION	POSITIVE: Risk becoming lower, there is economies of scale, more resources and a bigger market. NEGATIVE: Formal and financial controls that difficult innovation appears.	POSITIVE: McEacern and Romeo (1978), Link (1982) and Chen (1996). NEGATIVE: Hoskisson and Hitt (1988), Baysinger and Hoskisson (1989) and Hoskisson and Johnson (1992).
	TECHNOLOGICAL OPPORTUNITY	POSITIVE: Time, cost and uncertainty reduction.	POSITIVE: Scherer (1965b, 1967), Lunn and Martin (1986), Meisel and Lin (1983), Levin et al. (1985, 1987), Cohen et a. (1987), Cohen and Levinthal (1989), Doi (1985) and Gumbau (1997).

We can conclude that the great majority of them are directly related with internal and external factors, showing contradictory results. Also, there are different ways of measure innovation processes and factors that we are studying. All of them can influence on the significance of the results.

The rest of the paper is structured as follows: Section 2 we describe the data and input variables; section 3 introduces some concepts of the RS theory; in section 4 the results of the RS approach are presented and finally, section 5 describes our conclusions.

3.- SELECTION OF VARIABLES AND DATA

In this stage of our research, we have proceeded to the selection of the data and variables that will be used to develop our model.

As for the *data*, we have used a sample of 4050 Spanish firms. For each firm we have five year data (from 2001-2005). The design of the Survey is made by Foundation SEPI. It controls the quality process from collection to consistency of information. The number of firms by year is shown in **table 5**. This fact guarantee sample representative.

Table 5. Summary of the sample's evolution in the years 1990-2006 All the firms.

	2000	2001	2002	2003	2004	2005	2006
1. Current sample	1754	1870	1724	1708	1380	1374	1911
1.1 Firms which answer	1634	1693	1635	1380	1374	1277	1716
	(93,1)	(90,5)	(94,8)	(80,8)	(100,0)	(92,9)	(89,8)
1.2 Firms which disappear and not accessible	96	177	77	240	6	85	181
1.3 Firms which do not collaborate	24	0	12	88	0	12	14
2. Firms recovered			73			46	0
3.Entries in the current year	236	31	0	0	0	588	307

As for the *variables*, we have to mention that choosing the variables for prediction the decisions firm to innovate is a critical issue. These variables could influence the quality of the results generated. In this research, each firm is described by 25 variables (qualitative and quantitative) that have come from previous researches about innovation processes as shown in **table 6**. The variables are:

Table 6. Variables definition.

VARIABLE	DEFINITION
number of employees	Number of total employees the firm.
industry	Manufacturing firms are made of 20 sector provided by ESEE: Meat-processing industry, Foodstuffs and tobacco, Drinks, Textiles, Leather and footware, Wood industry, Paper, Editing and printing, Chemical industry, Rubber and plastics, Non-metallic minerals products, Iron and steel, Metallic products, Machinery and mechanical goods, Office machinery, computers, processing, optical and similar, Electrical and electronic machinery and material, Motor vehicles, Other transport material, Furniture and Other manufacturing.
foreign capital participated	Firms have at least 1% of foreign capital.
foreing direct investment	Firms has at least 1 subsidiary abroad
Age	Age of the firm
long term debt	Total amount of long term debt of the firm
short term debt	Total amount of short term deb of the firm
cashflow	Cash flow of the firm
volume of exports in euros	Total amount of exports made by the firm.
market share	Market share.
advertisement expenditures	Total amount of advertisement expenditures made by the firm
market stage	Market stage where the firm develop its activity. It has three values, stable, increasing or decreasing stage.
number of graduate employees	It is referred to the total amount of graduate employees working on the firm
Region	It is made by 17 regions Spain is made of.
mean of patents in industry	Mean of total patents per industry excluded firm ones.
diversification index	Diversification index provide by ESEE. It's made by three values, the firm doesn't diversificate, the

	firm makes related diversification and the firm makes unrelated diversification.
technological opportunity	Mean of industry R&D expenditures divided by sales excluding firms ones.

4. ROUGH SET (RS) THEORY: MAIN CONCEPTS

RS Theory was firstly developed by Pawlak (1991) in the 1980s as a mathematical tool to deal with the uncertainty or vagueness inherent in a decision making process. Though nowadays this theory has been extended (Greco *et al*, 1998 and 2001), we refer to classical approach that does not order attribute domains as it assumes that different values of the same attribute are equally preferable and that only the predictive value of the attribute, as revealed by the data, will be factored into the model. The extended approach handle dominance relations, in addition to indescernibility relations, incorporating data about the ordering properties of the attributes analyzed, if these exit and are known. The resulted model is potentially more compact since some rules conflicts for certain cases are eliminated. Therefore, it uses additional information to generate a simpler final model, but the classical approach makes a less restrictive data assumption than does the extended approach (McKee, 2000, p.162).

RS theory is related in some aspects to other tools that deal with uncertainty. However, RS approach is somewhat different to statistical probability, which deals with random events in nature or fuzzy set theory, which deals with objects that may not belong only to one category but may belong to more than one category by differing degrees. On the contrary, RS theory deals with the uncertainty produced when some objects described by the same data or knowledge (so, they are indiscernible) can be classified into different classes, that is, there is not a unique inclusion of these indiscernible objects. This fact prevents their precise assignment to a set. These differences show one of the main advantages of RS theory: an agent is not required to assign precise numerical values to express imprecision of his knowledge, such as probability distributions in statistics or grade of membership in fuzzy set theory (Nurmi *et al.*, 1996).

This section presents some concepts of RS Theory following Pawlak´s reference and some remarks by Slowinski (1993) and Dimitras *et al.* (1999).

The philosophy of this approach is based on the assumption that with every object of the universe we are considering we can associate knowledge, data. Knowledge is regarded as ability to classify objects. Therefore knowledge consists of a family of various classification patterns of a domain of interest. Objects described by the same data or knowledge are indiscernible in view of such knowledge. The indiscernibility relation leads to mathematical basis for the RS Theory. Vague information causes indiscernibility of objects by means of data available and, as a result, this prevents their precise assignment to a set. Intuitively, a rough set is a set or a subset of objects that cannot be expressed exactly by employing available knowledge. If this information or knowledge consists of a set of objects described by another set of attributes, we consider a rough set as a collection of objects that, in general, cannot be precisely characterized in terms of the values of the set of attributes.

RS Theory represents knowledge about the objects as a data table, that is, an *information table*. Rows of which are labelled by objects (states, processes, firms, patients, candidates…) and columns are labelled by attributes. Entries of the table are attribute values. Therefore, for each pair object-attribute, *x-q*, there is known a value called *descriptor*, *f(x, q)*. The *indiscernibility relation* would occur if for two objects, *x* and *y*, all their descriptors in the table have the same values, that is if, and only if, f(x,q)=f(y, q).

In general, all properties of rough sets are not absolute, but are related to what we know about them. Indiscernible objects by means of attributes prevent their precise assignment to a class. Therefore, some categories (subsets of objects) can not be expressed exactly by employing available knowledge and, consequently, the idea of approximation of a set by other sets is reached. A rough set is a pair of a *lower and an upper approximation* of a set in terms of the classes of indiscernible objects. That is, it is a collection of objects that, in general, cannot be precisely characterized in terms of the values of the set of attributes, while a lower and an upper approximation of the collection can be. Therefore, each rough set has boundary-line cases, that is, objects that cannot be classified certainly as members of the set or of its complement and can be represented by a pair of crisp sets, called the lower and the upper approximation. The lower approximation consists of all objects that certainly belong to the set and can be certainly classified as elements of that set, employing the set of attributes in the table (the knowledge we are considering). The upper approximation contains objects that possibly belong to the set and can be possibly classified as elements of that set using the set of attributes in the table. The *boundary* or *doubtful region* is the difference between the lower and the upper approximation and is the set of elements that cannot be certainly classified to a set using the set of attributes. Therefore, the borderline region is the undecidable area of the universe, that is, none of the objects belonging to the boundary can be classified with certainty into a set or its complement as far as knowledge is concerned.

Inexactness of a set is due to the existence of the boundary. The greater the doubtful region of a set is, the lower the accuracy of that set. The *accuracy of approximation* is defined as the quotient between the cardinality of the lower approximation and the cardinality of the upper one. This ratio expresses the percentage of possible correct decisions when classifying objects employing knowledge available. Therefore, using the lower and the upper approximation we can define precisely those subsets that cannot be expressed exactly using the available attributes.

Because we are interested in classifications, the *quality of classification* is defined as the quotient between the addition of the cardinalities of all the lower approximations of the classes in which the objects set is classified, and the cardinality of the objects set. It expresses the percentage of objects which can be correctly classified to classes employing the knowledge available.

A fundamental problem in the rough set approach is discovering dependencies between attributes in an information table because it enables to reduce the set of attributes removing those that are not essential (unnecessary) to characterize knowledge. This problem will be referred to as knowledge reduction or, in more general terms, as a f*eature selection problem*. The main concepts related to this question are the *core* and the *reduct.* A reduct is the minimal subset of attributes which provides the same quality of classification as the set of all attributes. If the information table has more than one reduct, the intersection of all of them is called the core and is the collection of the most relevant attributes in the table.

An information table which contains condition and decision attributes is referred as a *decision table*. A decision table specifies what decisions (actions) should be undertaken when some conditions are satisfied. So a reduced information table may provide decision rules of the form "*if* conditions *then* decisions".

These rules can be *deterministic* when the rules describe the decisions to be made when some conditions are satisfied and *non-deterministic* when the decisions are not uniquely determined by the conditions so they can lead to several possible decisions if their conditions are satisfied. The number of objects that satisfy the condition part of the rule is called the *strength* of the rule and is a useful concept to assign objects to the strongest decision class when rules are non-deterministic.

The rules derived from a decision table do not usually need to be interpreted by an expert as they are easily understandable by the user or decision maker. The most important result in this approach is the generation of decision rules because they can be used to assign new objects to a decision class by matching the condition part of one of the decision rule to the description of the object. Therefore rules can be used for decision support.

RS Theory can analyse several multiattribute decision problems. It is especially well suited to classification problems. We want to mention that rough set analysis has been performed using RSES2 developed by

Institute of Mathematics, Warsaw, Poland (http://logic.mimuw.edu.pl/~rses/)

5. EMPIRICAL MODEL AND RESULTS

The complete information table contains data from 4050 firms in the period 2000-2005 described by the 25 variables explained in section 3, and assigned to a decision class (innovate –1 or not -0-). If we developed a model and we test it with the same sample, the results obtained could be conditioned. So in order to avoid it, we have formed a training set, and a holdout sample to validate the obtained model (decision rules), i.e., the test set. Both sets have been randomly selected. The training information table makes up 70% of total firms and the test information table is made up of the rest of the firms.

The training information table was entered into an input file in RSES2. We have recoded the continuous variables into qualitative terms (very low, low, medium, high and very high) with corresponding numeric values such us 1, 2, 3, 4 y 5. This recoding has been made dividing the original domain into subintervals. This recoding is not imposed by the RS theory but it is very useful in order to draw general conclusions from the variables in terms of dependencies, reducts and decision rules (Dimitras *et al.,* 1999).

The definition of the boundary values can influence results of the RS analysis, in particular the quality of classification. There is not a general way to define the optimal boundary values. It is usually done by experts according to their experience, knowledge, habits or conventions, as in financial problems (Slowinski and Zopounidis, 1995; Dimitras *et al.*, 1999). If there is not an expert to recode the variables that could follow their experience or standards of financial analysis, it is deemed desirable to avoid subjective inputs to the extent possible. Accordingly, the subintervals are based on the percentiles (20-40-60-80) for the whole sample because they are frequently used in scientific researches to divide a domain into subintervals (Laitinen, 1992; McKee, 2000). **Table 7** shows the list of subintervals

Table 7. List of Subintervals (quartiles)

Variable	1st	2nd	3rd	4th

CRECIM.	(-∞, 2.9]	(2.9, 3.5]	(3.5, 4.6]	(4.6, +∞)
CBANKINDEP	(-∞, 0.37]	(0.37, 0.59]	(0.59, 0.82]	(0.82, +∞)
REAL INTEREST	(-∞, 0.33]	(0.33, 3.64]	(3.64, 6.15]	(6.15, +∞)
DOM.CREDIT.GROWTH	(-∞, 7.90]	(7.90, 15.88]	(15.88, 28.47]	(28.47, +∞)
BANK CASH REV.	(-∞, 0.02]	(0.02, 0.06]	(0.06, 0.15]	(0.15, +∞)
FOR_LIAB_REV	(-∞, 0.36]	(0.36, 0.52]	(0.52, 0.65]	(0.65, +∞)
GDP_GROWTH	(-∞, 2]	(2, 4]	(4, 6]	(6, +∞)
INFLATION	(-∞, 2.81]	(2.81, 7.25]	(7.25, 16.04]	(16.04, +∞)
%NETKFLOWS	(-∞, -0.003]	(-0.003, 0.896]	(0.896, 4,724]	(4,724, +∞)
GDPPERHEAD	(-∞, 3144]	(3144, 8180]	(8180, 17392]	(17392, +∞)

The first results obtained from RS analysis of the coded information table were the core and the reducts. The core is consisted of 16 attributes (number of employees, industry, foreign direct investment, age, long term debt, short term debt, cashflow, volume of exports in euros, market share, advertisement expenditures, market stage, number of graduate employees, region, mean of patents in industry, technological opportunity, volume of export lag1), which represent the most relevant attributes in the table. We have obtained 6 reducts from the table which contain 18-19 attributes. Next table shows the six reducts.

Table 8 Reducts.

Reduct	Nº variables	
R1	18	core, diversification index, market stage lag1,
R2	19	core, market stage lag1, no diversification, related diversification
R3	18	core, diversification index, stable stage
R4	19	core, diversification index, increasing stage, decreasing stage
R5	19	core, stable stage, no diversification, related diversification
R6	20	core, increasing stage, decreasing stage, no diversification, related diversification,

This result means that at least 5 attributes are redundant and, therefore, they could be eliminated. Consequently, this result shows the strong support of this approach in feature selection. There are two variables (foreign capital participated and unrelated diversification) that do not appear in any reduct. Therefore they are not necessary to the problem we are considering.

We are going to analyze the six models we have obtained but in this paper we have focus on R1 and R3 because they are the shortest ones. So we have obtained two reduced tables to obtain the decision rules. We have employed two algorithms for each model to obtain the decision rules. This way we can check the results in order to interpret the rules.

The two decision models (R1 and R3) have been tested on data from the testing tests, i.e., on the 30% firms that have not been used to estimate the two algorithms. The classification accuracy in percent of correctly classified firms by R1 model is 77% and by R3 model is 76,7%. These results are quite satisfactory and validate both decision rules sets.

To interpret the rules, we have only focus on the strongest ones (**Table 9** and **10** in the Appendix shows the strongest rules for both models).

We can draw some general conclusions in a preliminary study of the two models:

- The number of rules for class 1 (innovate) is higher is higher than the number of rules for class 0. The number of attributes in the rules varies from 5 to 13.

- The variables that appear at least in the 50% of the rules are: number of employees, foreign direct investment, cashflow, volume of exports in euros, market share, advertisement expenditures, market stage, number of graduate employees, diversification index, market stage lag1 and volume of export lag1.

- The variables *number of employees, long term debt, short term debt, cashflow, volume of exports in euros and number of graduate employees* are situated above percentile 80 (very high values) for firms that belong to class 1. On the contrary, these variables are situated below the 20 percentile (very low values).

- Firms that belong to class 0 have a very small market share.

- The firms that decide to innovate belong to the following industrial sectors: 17 (Motor vehicles), 9 (Chemical industry) and 16 (Electrical and electronic machinery and Material) whereas firms that do not decide to innovate belong to 6 (Wood industry) industrial sector.

6. CONCLUSIONS

In this paper we have applied a data analysis methodology of the field of Machine Learning, Rough Set theory, on a sample of firms in the period 2000 to 2005 with the purpose of analyzing the role of a set of variables in explaining the decision of innovating.

Through the exposition we have mentioned some advantages of this approach so we can conclude that this method is an effective tool for supporting managerial decision making in general. In the light of the experiments carried out, this theory is a competitive alternative to existing prediction models for this problem and has great potential capacities that undoubtedly make it attractive for application to the field of business classification.

The general results we have obtained in terms of classification are quite good for both models with a 77% of correct classifications using the test.

Besides that our empirical results show that this technique offers a great predictive accuracy, it is non-parametric, or distribution free, methods, so it does not require the pre-specification of a functional form, or the adoption of restrictive assumptions about the characteristics of statistical distributions of the variables and errors of the model. The decision models are easily understandable and interpretable. This representation of the results makes easier the economical interpretation than other non-parametric techniques like Neural Networks or Support Vector Machines. Moreover, the flexibility of the decision rules with changes of the models over the time allows us to adapt them gradually to the appearance of new cases representing changes in the situation.

Though we have drawn some general conclusions, we have to study in depth in order to look for a more exhaustive explanation.

REFERENCES

ACS, Z. J. & AUDRETSCH, D. B. (1990) *The economics of small firms : a European challenge,* Dordrecht, Kluwer.

ARCHIBUGI, D., CESARATTO, S. & SIRILLI, G. (1991) Sources of innovative activities and industrial organization in Italy. *Research Policy,* 20.

ARENA, R. & LAZARIC, N. (2003) La théorie évolutionniste du changement économique de Nelson et Winter. Une analyse économique rétrospective. *Revue Économique,* 54.

BAYSINGER, R. D. & HOSKISSON, R. (1989) Diversification strategy and R&D intensity in large multiproduct firms. *Academy of Management Journal,* 32.

BRAGA, H. & WILLMORE, L. (1991) Technological imports and technological effort: an analysis of their determinants in Brazilian firms. *Journal of Industrial Economics,* 39.

BROUWER, E. & KLEINKNECHT, A. (1993) Employment growth and innovation at the firm level An empirical study *Journal of evolutionary economics,* 3.

BUESA BLANCO, M. (2005) Innovación. *El espacio común de conocimiento en la Unión Europea. Un enfoque al problema desde España.* Madrid, Academia Europea de Ciencias y Artes.

BUESA BLANCO, M. & MOLERO ZAYAS, J. (1996) Patterns of technological change among Spanish innovative firmas: the case of Madrid Region. *Research Policy,* 25.

BUESA BLANCO, M. & ZUBIAURRE, A. (1999) Patrones tecnológicos y competitividad: un análisis de las empresas innovadoras en el País Vasco. *Ekonomiaz. Revista Vasca de Economía*.

BUSOM, I. (1991) Impacto de las ayudas públicas a las actividades de I + D de las empresas: un análisis empírico. *D: Herri-Ekonomiaz = Economía Pública,* 11.

CHEN, R. (1996) Technological expansion: the interaction between diversification strategy and organizational capability. *Journal of Management Studies,* 33.

COHEN, W., LEVIN, R. & MOWERY, D. (1987) Firm size and R&D intensity: a re-examination. *Journal of Industrial Economics,* 35.

COHEN, W. & LEVINTHAL, D. (1989) The implications of spillovers for R&D investment and welfare: A new perspective. IN LINK, A. & SMITH, K. (Eds.) *Advances in applied microeconomics: The factors affecting technological change.* Greenwich, Conn., Jai Press.

COHEN, W. & LEVINTHAL, D. (1989) Innovation and learning: the two faces of R&D implications for the analysis of R&D investment. *Economic Journal,* 99.

CORIAT, B. & DOSI, G. (1994) Learning how to Govern and Learning how to Solve Problems: on the Co-Evolution of Competences, Conflicts and Organizational Routines. *MIMEO*.

DAGUPTA, P. & STIGLITZ, J. (1980) Uncertainty, industrial structure, and the speed of R&D. *Bell Journal of Economics,* 11.

DIMITRAS, A., SLOWINSKI, R., SUSMAGA, R. and ZOPOUNIDIS, C. (1999): "Business failure prediction using Rough Sets", *European Journal of Operational Research*, nº 114, pp. 263-280.

DOI, N. (1985) Diversification and R&D activity in Japanese manufacturing firms. *Managerial and Decision Economics,* 6.

DOSI, G. (1982) Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy***,** 147-162.

DOSI, G. (1984) *Technical change and industrial transformation : the theory and an application to the semiconductor industry,* London, Macmillan.

EDQUIST, C. (1997) *Systems of innovation : technologies, institutions, and organizations,* London, Pinter.

FREEMAN, C. (1973) A study of successs and failure in industrial innovation. IN WILLIAMS, B. R. & INTERNATIONAL ECONOMIC ASSOCIATION. (Eds.) *Science and technology in economic growth : proceedings of a conference held by the International Economic Association at St. Anton, Austria.* London, Macmillan.

FREEMAN, C. & PÉREZ, C. (1988) Structural crises of adjustment: business cycles. IN DOSI, G. & ALI, E. (Eds.) *Technical Change and Economic Theory.* Londres, Pinter Publishers.

GALENDE DEL CANTO, J. (2003) Determinantes, patrones y resultados del proceso de innovación en la empresa española. IN NAVAS, J. E. & NIETO, M. (Eds.) *Estrategias de innovación y creación de conocimiento tecnológico en las empresas industriales españolas.* Madrid, Civitas Ediciones.

GALENDE DEL CANTO, J. & SUÁREZ, I. (1998) Los factores determinantes de las inversiones empresariales en I + D. *Economía Industrial*.

GALENDE DEL CANTO, J. & SUÁREZ, I. (1999) A resource-based análisis of the factors determining a firm´s R&D activities. *Research Policy,* 28.

GRABOWSKI, H. (1968) The determinants of industrial research and development: a study of the chemical, drug, and petroleum industries. *Journal of Political Economy,* 76.

GRABOWSKI, H. & MUELLER, D. (1974) Industrial research and development, intangible capital stocs, and firma profit rates. *Bell journal of economics,* 9.

GRECO, S., MATARAZZO, B., and SLOWINSKI, R. (1998): "A new rough set approach to evaluation of bankruptcy risk", In Zopounidis, C. (ed.), *New Operational Tools in the Management of Financial Risks*, Kluwer Academic Publishers, Dordrecht, pp.121-136.

GRECO, S., MATARAZZO, B., andSLOWINSKI, R. (2001): "Rough Sets Theory for Multicriteria Decision Analysis", *European Journal of Operational Research*, vol. 129, nº 1, pp. 1-47.

GUMBAU, M. (1997) Análisis microeconómico de los determinantes de la innovación: aplicación a las empresas industriales españolas. *Revista Española de Economía,* 14.

HARABI, N. (1995) Appropriability of technical innovations: an empirical análisis. *Research Policy,* 24.

HEIJS, J. (2001) Justificación de la política tecnológica: un enfoque teórico. *Madri+d*.

HOROWITZ, I. (1962) Firm size and research activity. *Southern Economic Journal,* 29.

HOSKISSON, R. & HITT, M. (1988) Strategic control systems and relative R&D investment in large multiproduct firms. *Strategic Management Journal,* 9.

HUERGO, E. & JAUMANDREU, J. (2004) Firms' age, process innovation and productivity growth. *International Journal of Industrial Organization,* 22.

KAMIEN, M. I. & SCHWARTZ, N. L. (1982) *Market structure and innovation,* Cambridge ; New York, Cambridge University Press.

KOGUT, B. (1991) Country Capabilities and the Permeability of Borders. *Strategic Management Journal,* 12.

LAITINEN, E. K. (1992): “Prediction of failure of a newly founded firm”, Journal of Business Venturing 1992; july; 323-340.

LEVIN, R., COHEN, W. & MOWERY, D. (1985) R&D appropiability, opportunity, and market structure: new evidence on some Shumpeterian hypotheses. *American Economic Review,* 75.

LUNDVALL, B.-A. (1992) *National systems of innovation : towards a theory of innovation and interactive learning,* London, Pinter.

LUNN, J. & MARTIN, S. (1986) Market structure, firm structure and research and development. *Quarterly Review of Economics and Business,* 26.

MCKEE, T. and LENSBERG, T**.** (1999): “Using a genetic algorithm to obtain a causally ordered model from a rough sets derived bankruptcy prediction model”, paper presented at T*he International Symposium on Audit Research,* The University of Southern California.

MCKEE, T. (2000): “Developing a Bankruptcy Prediction Model via Rough Sets Theory”, *International Journal of Intelligent Systems in Accounting, Finance and Management* nº 9, pp. 159-173.

MALERBA, F. & ORSENIGO, L. (1990) Technological Regimes and Patterns of Innovation: A Theoretical and Empirical Investigation of the Italian Case. IN HEERTJE, A. & PERLMAN, M. (Eds.) *Evolving technology and market structure : studies in Schumpeterian economics.* Ann Arbor, University of Michigan Press.

MALERBA, F. & ORSENIGO, L. (1995) Schumpeterian Patterns of Innovation. *Cambridge Jounal of Economics***,** 47-65.

MALERBA, F. & ORSENIGO, L. (1996) Schumpeterian Patterns of Innovation are technology-specific. *Research policy,* 25.

NELSON, R. R. (1993) *National innovation systems : a comparative analysis,* New York ; Oxford, Oxford University Press.

NELSON, R. R. & WINTER, S. G. (1982) *An evolutionary theory of economic change,* Cambridge,Massachusetts, The Belknap Press.

NURMI, H., KACPRZYK, J. and FEDRIZZI, M. (1996): "Probabilistic, fuzzy and rough concepts in social choice"; *European Journal of Operational Research*, nº 95, pp.264-277.

PAWLAK, Z. (1991): *Rough Sets. Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht/ Boston/ London.

PAVITT, K. (1984) Sectorial patterns of technical change: Towards a taxonomy and theory. *Research policy,* 13.

PAVITT, K., ROBSON, M. & TOWNSEND, J. (1988) Sectoral patterns of production and use of innovations in the UK: 1945-1983. *Research Policy,* 17.

ROTHWELL, R. (1986) The role of small firms in technological innovation. IN CURRAN, J., STANWORTH, J. & WATKINS, D. S. (Eds.) *The survival of the small firm,* Aldershot, Gower.

ROTHWELL, R. & DODGSON, M. (1994) Innovation and size of firm. IN DODGSON, M. & ROTHWELL, R. (Eds.) *The handbook of industrial innovation.* Aldershot, Hants, England ; Vermont, USA, E. Elgar Pub. Co.

SALOMON, R. & SHAVER, J. (2002) Export and domestic sales: their interrelationship and determinants. *Strategic Management Journal,* 26.

SCHERER, F. M. (1965) Firm size, market structure, opportunity, and the output of patented inventions. IN SCHERER, F. M. (Ed.) *Innovation and growth : Schumpeterian perspectives* Cambridge, Mass. ; London, MIT Press.

SCHUMPETER, J. A. (1936) *The theory of economic development : an inquiry into profits, capital, credit, interest, and the business cycle,* Cambridge, Harvard University Press.

SCHUMPETER, J. A. (1942) *Capitalism, socialism, and democracy,* New York, Harper.

SLOWINSKI, R. (1983): "Rough set learning of preferential attitude in multicriteria decision making", in: J. Komorowski and Z. W. Ras (eds.), *Methodologies for Intelligent Systems. Lecture Notes in Artificial Intelligence* vol. 689, Springer-Verlag, Berlin, pp. 642-651.

SLOWINSKI, R. and ZOPOUNIDIS, C. (1995): "Application of the rough set approach to evaluation of bankruptcy risk", *International Journal of Intelligent Systems in Accounting, Finance and Management* vol. 4; nº1; pp.27-41.

VERNON, R. (1966) International Investment and International Trade in the Product Cycle. *Quarterly Journal of Economics,* 80.

WINTER, S. G. (1984) Schumpeterian Competition in Alternative Technological Regimes. *Journal of economic behavior & organization,* 5.

Table 9: Decisión rules R1	employ	industr	fdi	age	debtlt	debtst	cash	exports	share	advert	stage	graduate	region	aprop.	divers.	techno	stlag1	explag1	innova
1	1		0		1	1	1	0	0	0					0		1	0	0
2	1		0		1	1	1	0	0	0	0	0			0		1	0	0
3			1			4				4					0		1		1
4	5		1				5	4				4			0	0	0	0	1
5	5		1				5	4		4		4		1	0		0		1
6			0		1			0	0	0	0		9		0		1	0	0
7			1					3		0	1				0		0		1
8			0		5	5		4			2	4			0		0		1
9	2		0		2	1		0	0	0		0			0		0	0	0
10	5		1		5					4	0	4			0		0		1
11			0					4	0		2	4			0		0	4	1
12	4		1						0			4			0		0	4	1
13	5		1	5								3			0		1	4	1
14	5		0				5	4		4	1	4			0	0	0	0	1
15	5		0		5	5			0	0	1				0		0		1
16			0	2	3				0	0	0			3	0	3	1		0
17	5		0						0	4	1	4			0	0	0	0	1
18	5		1		5	5	5	4							2		0		1
19			0				1	0	0		0			4	0	2	1	0	0
20			0			1		0	0		0				2		0	0	0
21	5		0		5	5	5	4		0	0	4			0	4	1	4	1
22	5		0		5	5			1	0	0	4			0		1		1
23	1		0		3		2	0	0	0	0	0			0		1	0	0
24			0		1	1	1		0	0	0	0			0	2	1		0
25	1		0				1				0			3	0		0		0
26			0			5	5		0	0		4		5	0	4	0	4	1
27	5		0			5	5	4		4	0	4			0		1	4	1
28	5		1		5		5			0	1	4			0		0	4	1
29			0	4		4	5	4			1				0		0		1
30	4		1					3						3	0		0		1
31	5		0			5				0	1	4		5	0	4	0	5	1
32			0				5	4	1	4	1					0	0	0	1
33	5		0		5	5	5		1	0	0	4					1		1
34	1		0			2	1	0	0	0		0			0		0	0	0
35	1		0				1	0	0	0	2				0	4	0	0	0
36	5	9	0			5	5	4		0	0	4		5	0	4	1	4	1
37	5		1	5						4	0	4			0		1	4	1
38	2		0		1		2	0	0	0	0	0			0		1	0	0
39			0			2		2	0	0	0	0			0		1	2	0

Table 10: Decisión Rules R3

	employ	industry	fdi	age	debtlt	debtst	cash	exports	share	advert	stage	graduate	region	aprop.	diversif.	techno	explag1	stg1lag1	innova
1	5		1				5	4		4		4			0			0	1
2	5		1	5	5	5	5	4				4						0	1
3	1		0		1	1	1	0	0	0	0	0			0		0	1	0
4			1				5			4		4			0		1	1	1
5	5		1		5	5	5	4		4		4			0			0	1
6	5		1			5	5			0	0					4	4	1	1
7			0		1			0	0	0	0		9		0		0	1	0
8	5		1					4			0				0	0	0	0	1
9	1		0			1			0		2				0			1	0
10	1		0			1	1	0	0	0	2	0			0		0	0	0
11	5		0						0	4	1				0	0	0	0	1
12	1		0	3					0	0	0	0			0	3		1	0
13	5		1		5	5	5	4		1	1	4			0		1	0	1
14	5		1		5	5	5	4		4		4			0		4	0	1
15		4	0				1	0	0	0	4	0		4	0		0	0	0
16	2		0		1		2	0	0	0					0		0	0	0
17			0			1	1	0	0		2				0		0	1	0
18		9	0	5					0		1		9		0			0	1
19			0						1	4	2	4			0			0	1
20	5		1				5	4		4	0	4			0		4	1	1
21	1		0	3	1	1	1	0			0				0		0	1	0
22	5		0			5			1	0	1	4			0	4		0	1
23			0					4	1	0	1			5	0	4		0	1
24	4		1				5	4			1				0		4	0	1
25	5		1			5	5	4		4	0	4			0	4	4	1	1
26	5		0		5			4			2	4			0		4	0	1
27	5		1	5	5	5		4		0		4			0		4	0	1
28	5		1	5						4	0				0		4	1	1
29	2		0			2		0	0		0	0		5	0		0	1	0
30	5		0						1	4	1	4			0	0	0	0	1
31	4	9	0			5	5	4		0	0	4		5	0	4	4	1	1
32	5		1	5	5	5	5	4		0	0				0		4	1	1
33	5		1		5	5			1	0				5	0			0	1
34			0		5	4				0	0				1			1	1

	employ	industry	fdi	age	debtlt	debtst	cash	exports	share	advert	stage	graduate	region	aprop.	diversif.	techno	explag1	stg1lag1	innova
35	2		0					0	0		0			2	0	2	0	1	0
36	1		0	1	1	1	1	0	0		0				0			1	0
37			0				4			4	0	3		1	0	0		0	1
38	5		1	4		5	5	4				4			0		4	1	1
39			1	5				3		4	0				0			0	1
40			0			2		2	0	0	0	0			0		2	1	0
41	1		0	3	1			0	0		2				0		0	0	0
42		2	0				1				0			1	0	0	0	0	0