

**Etude économétrique des liens entre R&D  
interne et coopération en R&D**  
projet dirigé par Matthieu MANANT

Stanley DURRLEMAN, Nadia FERAOUN, Thomas FUHR,  
Alexandre GRAMFORT, Aurélien LOUIS et Julie ROBILLARD

22 juin 2005

## Table des matières

<b>1</b>	<b>Une revue de la littérature : les bases de la théorie de la coopération en R&amp;D</b>	<b>3</b>
1.1	Concentration du marché et incitation à innover . . . . .	3
1.2	Coopération et spillovers . . . . .	3
1.3	La notion de capacité d'absorption . . . . .	4
1.4	Réseaux de collaboration en R&D . . . . .	5
1.5	Les limites de la théorie . . . . .	5
<b>2</b>	<b>Description des variables</b>	<b>7</b>
<b>3</b>	<b>Le modèle</b>	<b>11</b>
<b>4</b>	<b>L'analyse des résultats</b>	<b>13</b>
4.1	Les résultats . . . . .	13
4.2	Leur pertinence . . . . .	14
4.3	Leur analyse . . . . .	14
4.4	La coopération se substitue-t-elle aux dépenses de R&D ? Les variables COOP et IRD . . . . .	16
4.5	Comparaison avec une autre étude. . . . .	17
<b>5</b>	<b>Annexe : Quelques mots d'économétrie...</b>	<b>20</b>
5.1	Moindres carrés ordinaires et modèle linéaire multiple . . . . .	20
5.2	Du modèle exponentiel au modèle linéaire . . . . .	21
5.3	Modèle Probit . . . . .	22
<b>6</b>	<b>Annexe : Origine des données (Intitulés des questions posées)</b>	<b>24</b>

## Introduction

Le département de Recherche et Développement (R&D) est un moteur crucial de l'activité d'une entreprise, fournissant les innovations nécessaires à la croissance. Toutefois, les entreprises peuvent choisir entre deux stratégies d'innovation. La première consiste à développer intégralement les technologies en interne, leur département R&D gérant de manière autonome l'ensemble des recherches. La seconde stratégie repose sur l'existence d'accords avec d'autres entreprises et d'échanges de savoirs entre départements de R&D, il s'agit d'une recherche menée sous forme de collaboration technologique.

L'enquête sur l'innovation CIS3-France 2001 [2] fournit des données pertinentes sur les stratégies de recherche employées par près de 7000 entreprises françaises. Les entreprises sondées possèdent des caractéristiques très variées : les PME sont présentes au même titre que les grandes multinationales, et ces entreprises appartiennent à différents secteurs industriels.

A partir des réponses à l'enquête, il est possible de dégager des variables d'intérêt permettant de caractériser la R&D dans chaque entreprise. Ainsi, les montants investis dans la R&D interne et dans la R&D externe sont accessibles ainsi que la taille et le secteur technologique de l'entreprise. Ceci nous donne une vision globale de sa stratégie de R&D.

L'un des axes d'étude de cette stratégie concerne les liens qui existent entre la coopération entre acteurs et les efforts internes consacrés à la recherche. S'agit-il de deux aspects complémentaires ou bien antagonistes d'une stratégie de R&D ? Quels caractéristiques déterminent le choix de la coopération, quelles variables influent sur le montant des dépenses propres en recherche ? Une vaste littérature, aussi bien théorique qu'empirique existe sur le sujet. Nous avons voulu confronter différentes hypothèses à la réalité décrite par l'enquête CIS3.

L'objectif de ce travail est d'élaborer à partir de ces données un modèle quantitatif mettant différentes variables en corrélation et valable pour toutes les entreprises sondées. L'écart de chaque entreprise à la courbe moyenne générale permet de retrouver sa stratégie de R&D : recherche exclusivement faite en interne ou collaboration. Cette étude économétrique est très proche de celle réalisée par Reinhilde Veugelers [1] sur des données hollandaises, datant de 1998. Ces résultats constituent donc une base intéressante de comparaison des caractéristiques de la R&D dans les entreprises de ces deux pays.

# 1 Une revue de la littérature : les bases de la théorie de la coopération en R&D

Notre étude statistique se base sur une ample littérature portant sur la R&D en entreprise, qui tente de répondre à deux questions :

- Quels facteurs influent sur la capacité de recherche d'une entreprise ?
- Quels facteurs influent sur la décision d'une entreprise de coopérer ?

Ainsi, le modèle que nous proposons, très proche de celui de Veugelers, repose sur une série d'hypothèses que des études précédentes, théoriques ou empiriques, ont tenté de confirmer. Afin de comprendre en détail ce modèle, nous proposons ici de revenir rapidement sur ces quelques hypothèses. De nombreuses revues de littérature existent (voir Veugelers [1], ou Manant ([11])). Celle-ci ne prétend pas être exhaustive, mais synthétique, et résumant uniquement les faits capturés dans notre modèle. Elle permettra au lecteur de se référer aux articles fondamentaux sur le sujet.

## 1.1 Concentration du marché et incitation à innover

La question de l'influence de la taille sur la capacité et l'incitation à innover est une question qui a été largement débattue, avec des vues souvent contradictoires.

**Hypothèse 1 : L'entrepreneur recherche des monopoles temporaires par le processus d'innovation.**

Selon Schumpeter [6], l'innovation permet à une firme de réaliser un profit par le monopole qu'elle lui confère. La recherche des parts de marchés est une incitation forte à l'innovation.

**Hypothèse 2 : Le monopole a une faible incitation à innover.**

En effet, Arrow [5] introduit le concept d'effet de remplacement. Un monopole est moins incité à innover, car en innovant, il ne gagne rien de plus que ce qu'il a déjà : un monopole.

## 1.2 Coopération et spillovers

L'étude théorique de la coopération en R&D s'est depuis longtemps fondée sur la notion de spillover (transfert de connaissance). Ces "fuites" peuvent avoir plusieurs origines : les mouvements de personnel entre les entreprises, les conférences, les échanges avec les fournisseurs et les clients. Fondamentalement, ils représentent une appropriabilité imparfaite des résultats de la recherche. Il en résulte une externalité positive pour les entreprises, qui sont moins incitées à investir en recherche en présence de spillovers, car elles bénéficient de l'innovation des concurrents.

**Hypothèse 3 : Plus les effets des spillovers sont importants, moins les firmes investissent en R&D.**

Le mécanisme, mis en lumière par D'Aspremont et Jacquemin dans [7] est double : un investissement plus important profite aussi à la firme concurrente, qui devient

plus compétitive. Anticipant cet effet, les entreprises retiennent leurs investissements. On parle d'effet de free-riding. De plus, les dépenses des autres entreprises remplacent les dépenses propres. Il s'agit d'un effet de substitution.

**Hypothèse 4 : Si le taux de spillover est en-dessous d'une valeur critique, une coordination des dépenses en R&D réduit les investissements en R&D par rapport à la situation de compétition. Il les augmente sinon.**

Dans le cas d'une coordination des dépenses entre firmes (d'un cartel de recherche), l'effet précédent est internalisé. Il disparaît. En revanche, apparaît un autre effet : les firmes retiennent leurs dépenses en R&D, car elles ne sont plus en compétition à ce stade. Pour de forts spillovers, la disparition du premier compense l'apparition du deuxième. Le résultat est globalement positif.

Il est possible d'introduire une autre forme de collaboration en R&D. Deux entreprises peuvent en effet décider de partager les résultats de leur recherche. Cette situation est désignée dans la littérature par le terme Research Joint Venture (RJV). Elle correspond théoriquement à un très fort spillover, c'est à dire une non-appropriabilité parfaite de la recherche. On est alors conduit à distinguer 4 cas, selon que les firmes coordonnent leurs dépenses de recherche ou non, et partagent leurs résultats ou non. Le résultat suivant, du à Kamien, Müller et Zang [8], est alors compréhensible étant donné ce qui précède :

**Hypothèse 5 : Quel que soit le taux de spillover, la dépense en recherche des firmes est toujours supérieure dans le cas d'un RJV avec cartel R&D, comparé aux autres situations. Elle est toujours inférieure dans le cas d'un RJV avec compétition au niveau des dépenses de recherche.**

En effet, un RJV-cartel implique des dépenses supérieures à un cartel R&D simple, car l'effet spillover (internalisé) est plus important dans un RJV. Un RJV-compétition implique des dépenses inférieures à une situation de compétition simple, car le spillover est plus important pour le RJV. Le résultat intéressant est qu'une situation de RJV-cartel est meilleure que la situation de compétition simple.

### 1.3 La notion de capacité d'absorption

Une notion complémentaire à celle de spillover a considérablement transformé la perspective traditionnelle concernant les liens entre dépenses de recherche et coopération. Il s'agit de la notion de capacité d'absorption. Celle-ci représente la capacité d'une entreprise à assimiler le flux d'information provenant des autres entreprises, c'est à dire les spillovers. La capacité d'absorption définit une certaine "productivité" des spillovers : plus elle est grande, plus les spillovers ont d'effet sur la productivité de l'entreprise. En ce sens, introduire cette capacité revient à faire de la recherche externe et de la recherche interne des substituts imparfaits. La manière classique pour introduire cette capacité d'absorption est de considérer qu'elle augmente avec les dépenses internes en R&D : ces dépenses propres permettent d'augmenter le stock de savoir de la firme, et donc d'avoir les bases pour comprendre les recherches des firmes concurrentes.

**Hypothèse 6 : La capacité d'absorption, lorsqu'elle est liée aux dépenses in-**

ternes de R&D :

- **Introduit un effet positif des spillovers sur les dépenses internes en R&D**
- **Diminue l'effet négatif de free-riding dû aux spillovers sur les dépenses internes en R&D**

Ce résultat, exhibé par Cohen et Levinthal dans [9] est assez clair. Si les spillovers augmentent, la firme a intérêt à dépenser plus en R&D pour augmenter sa capacité d'absorption et mieux les assimiler. De plus, une firme a moins d'intérêt au free-riding, puisque diminuer sa propre recherche diminue sa capacité d'absorption.

#### 1.4 Réseaux de collaboration en R&D

Il est nécessaire, pour avoir une compréhension plus globale des phénomènes de coopération, d'introduire des éléments stratégiques, quand au choix des partenaires de recherche. Ceci peut être fait en partant de la notion de réseau de coopération en recherche, c'est à dire considérer que les firmes peuvent décider du nombre de leurs partenaires. Goyal et Moraga (dans [10]) obtiennent des résultats intéressants en considérant des collaborations bilatérales de type RJV-compétition au sens de Kamien.

**Hypothèse 7 : Un réseau complet (toutes les firmes collaborent avec toutes les autres) minimise le bien-être social, ainsi que les dépenses en R&D.**

Cette situation est la même que celle de Kamien lorsqu'il étudie le modèle de RJV-compétition. Le résultat est donc cohérent

**Hypothèse 8 : Un nombre de collaborations intermédiaires maximise le bien-être social.**

Ce résultat est contre-intuitif, si l'on se base sur ceux de Kamien. Il y a toutefois un aspect que Kamien ne saisit pas dans son étude : il s'agit de la non-duplication des efforts en R&D qu'implique les spillovers. Cet effet positif compense les diminutions globales du niveau des dépenses en R&D dû au free-riding, lorsque le niveau de collaboration n'est pas trop élevé.

**Hypothèse 9 : Le réseau complet est stable.**

Ceci est problématique : en effet, les firmes pourraient préférer collectivement ne pas former trop de liens, comme le suggère l'hypothèse 6. Dans une telle configuration, deux firmes gagneraient un avantage concurrentiel à entamer une collaboration. Ainsi, il y a donc une incitation pour les firmes à former de trop nombreuses collaborations.

#### 1.5 Les limites de la théorie

Il existe un certain nombre d'aspects qui n'ont pas été traités correctement par la théorie, et qui pourtant jouent vraisemblablement un rôle important dans les décisions des entreprises relatives aux investissements en recherche.

Le premier aspect concerne celui de la spécificité des actifs mis en oeuvre dans une coopération en R&D, ou, plus largement, le partage de compétences entre

entreprises. En effet, il peut être socialement intéressant pour deux entreprises de collaborer si leurs compétences dans un domaine sont complémentaires. Dans ce cas, la non-duplication des coûts associés à l'acquisition des compétences peut plaider en faveur de collaboration accrues.

Le deuxième aspect concerne le rôle essentiel de l'innovation dans les stratégies de différenciation des entreprises. La théorie traite la R&D sous l'angle de l'innovation de processus qui réduit les coûts de production. L'innovation de produit est généralement conçue comme une innovation verticale, qui augmente la demande pour un produit. Les deux aspects sont, sous l'angle des équations, totalement équivalents. Or, l'innovation peut aussi être horizontale (peut viser à établir un produit destiné à une fraction différente du marché), et donc comporter des implications stratégiques importantes. De ce point de vue, il peut y avoir une désincitation forte à coopérer dans la mesure où la coopération permet aux entreprises concurrentes de réaliser des produits similaires.

## 2 Description des variables

Afin d'évaluer les intérêts des entreprises à adopter une stratégie coopérative ou au contraire compétitive, il est nécessaire d'étudier leur structure et leurs orientations en matière d'innovation et de recherche. Les évolutions technologiques atteignent un rythme tel qu'il est indispensable pour les entreprises d'engager des dépenses dans la recherche et le développement. Cependant, si certaines entreprises possèdent les moyens et les structures capables de supporter une recherche propre, et de la protéger, d'autres entreprises préféreront bénéficier d'un savoir-faire externe ou partager les coûts d'un laboratoire avec d'autres. Il apparaît ici que la taille de l'entreprise ou sa capacité à absorber des connaissances externes, par exemple, sont des variables potentiellement influentes sur la stratégie de R&D adoptée. Les caractéristiques susceptibles d'influer sur la stratégie R&D sont listées dans cette première partie et récapitulées dans le tableau 1. Certaines ont été présentées dans l'article de référence de R. Veugelers [1], d'autres semblent pertinentes compte tenu du sondage soumis aux entreprises [2].

L'un des objectifs de cette étude est d'établir les liens entre les stratégies de recherches et les caractéristiques propres à l'entreprise. Le cœur de l'étude repose donc sur l'aptitude d'une entreprise à développer des produits ou à innover. Malheureusement, ces attributs ne sont pas quantifiables : il est nécessaire d'effectuer des hypothèses afin de les évaluer. Pour mesurer les performances technologiques d'une entreprise, et sa propension à soutenir une recherche interne propre, la variable **IRD** (*Internal R&D*) est introduite. **IRD** représente les investissements en R&D effectués au sein de l'entreprise, hors subventions ou apports extérieurs. C'est un montant en euros. Si certaines entreprises consacrent une partie de leur budget dans de la R&D externe, i.e. effectuée hors de ses locaux, la part de ce budget accordée à la recherche propre semble refléter la volonté de l'entreprise à garder une structure innovante à part entière. Cette donnée est accessible dans l'enquête [2], question 4.1, dans laquelle toutes les dépenses en R&D sont détaillées.

Intuitivement, si une entreprise est de petite taille, elle n'aura pas la capacité d'entretenir un département de R&D et aura plutôt tendance à favoriser soit une coopération sous forme d'un laboratoire commun, soit une acquisition des savoir-faire d'autres entités sans mener elle-même de recherche interne propre. Par contre une grande entreprise dispose de plusieurs alternatives. Elle peut soutenir un laboratoire interne fermé ou bien coopératif, ou encore adopter le choix de ne pas faire de recherche interne mais de profiter des savoirs extérieurs. Par ailleurs, si les petites entreprises sont plus réactives que les grandes, ces dernières ont une stabilité qui leur permet de prendre des risques en matière d'innovation. La taille de l'entreprise influe donc potentiellement sur la stratégie de R&D. Elle est chiffrée à l'aide de la variable **S** (*Sales*) égale au chiffre d'affaire en euros, donné en question 0.5 de l'enquête [2].



De manière similaire une petite entreprise nationale n'aura pas le même comportement en matière d'innovation qu'une multinationale qui peut déléguer le travail de R&D à certaines de ses filiales. Dans le cas présent, l'influence du caractère multinational est chiffré à l'aide d'une variable binaire, un *dummy*, **M** (*Multinational*) valant 1 lorsque l'entreprise est une multinationale dont le siège est situé hors de France, et 0 sinon. Cette donnée répond à la question 0.1 de l'enquête [2].

Il relève d'un choix de comparer les stratégies en matière de R&D d'entreprises de secteurs différents. Il est en effet possible de supposer que la stratégie est en partie influencée par le domaine d'activité et de comparer les comportements de toutes les entreprises entre elles. Il est également envisageable de reproduire l'étude sur tous les domaines d'activités indépendamment les uns des autres. L'enquête [2] vise des entreprises innovantes dans les domaines de la chimie, dont le *dummy* associé est **CHEM** (*Chemistry*), des technologies de l'information : **IT** (*Information Technology*), du développement de programmes informatiques : **INFO** (*Software services*) ou enfin de l'électronique : **MVEN** (*Metals manufacturing and electronics*). Certaines de ces entreprises ne répondant pas à tous les critères exigés pour l'étude, seuls les domaines **CHEM** et **MVEN** sont retenus. La détermination du domaine d'activité des entreprises sondées dans l'enquête s'effectue à l'aide de leur numéro NAF-APE700 caractérisant le secteur [3]. Toutefois, dans l'échantillon que nous avons considéré, il n'y avait aucune entreprise du secteur IT ou des services informatiques. Pour la pertinence de nos résultats, nous avons supprimé ces variables du modèle.

Les efforts entrepris en R&D peuvent être encouragés par des aides gouvernementales. L'existence de subvention conditionne donc potentiellement la stratégie de R&D. Si ces dépenses sont données sans condition par le gouvernement français, les subventions européennes sont conditionnées par l'obligation d'adopter une stratégie coopérative. La variable **SUB** (*Subvention*) est un *dummy* valant 1 si l'entreprise reçoit des subventions régionales, gouvernementales ou européennes (réponse à la question 7.1 de [2]).

Afin d'évaluer précisément le montant des investissements en recherche à l'extérieur de l'entreprise, on introduit la variable **ERD**. La variable **ERD** est une valeur numérique correspondant à la somme directement investie par l'entreprise en recherche externe, ce qui correspond à la question 4.1 de [2]. Comme la variable **IRD**, il s'agit d'un montant évalué en euros. Ce montant ne prend pas en compte les autres moyens de coopérer entre entreprises comme la mise en partage de brevets et licences ou la mise en commun de personnel de recherche. Cette variable nous permet toutefois une bonne évaluation de l'effort entrepris par l'entreprise dans sa recherche externe.

Une autre variable à prendre en compte afin d'estimer la recherche externe d'une entreprise est l'existence d'une recherche en coopération. La réponse donnée à la

question 8.1 de [2] permet de définir la variable adaptée **COOP**. La variable **COOP** est un *dummy* qui a pour valeur 1 si l'entreprise répond qu'elle mène des activités de recherche en collaboration avec d'autres entreprises, ou bien 0 si l'entreprise répond négativement à cette question. C'est une variable très importante pour le modèle que nous allons développer dans la prochaine partie, et elle y sera détaillée plus avant.

Un paramètre important pour évaluer la politique de recherche d'une entreprise est sa capacité à maintenir une veille technologique sur son domaine de recherche et développement. La réponse à la question 9 de [2] permet de savoir si l'entreprise met en oeuvre cette veille ou non. Cette question porte sur les différentes sources d'informations dont dispose l'entreprise pour orienter son effort de recherche : articles publiés, etc... Nous avons alors introduit le paramètre **MONI** qui est un *dummy* dont la valeur est 1 si l'entreprise réalise de la veille technologique et 0 si elle n'en maintient pas. Il s'agit d'un paramètre n'intervenant pas dans l'article de R. Veugelers [1] mais qui doit apporter un éclairage pertinent sur la politique de recherche et sur les éventuelles collaborations mises en oeuvre par les entreprises sondées dans l'enquête [2].

Un élément important pour mesurer l'activité de recherche et développement d'une entreprise est sa capacité à développer des produits innovants. L'enquête [2] sur laquelle nous travaillons permet de savoir, grâce à la réponse donnée par l'entreprise à la question 1.4, la part de chiffre d'affaire réalisée par l'introduction sur le marché de produits innovants. La variable **DIFF** a donc été introduite pour prendre en compte ce montant. Il s'agit d'une somme évaluée en euros, comme précédemment. Cette variable traduit une orientation de la politique de recherche : une somme importante montre que l'entreprise mise sur la différenciation de ses produits. On peut alors imaginer qu'une entreprise faisant de la recherche pour introduire des produits différents, plutôt que pour améliorer des processus cherchera moins à coopérer, puisque tout l'avantage qu'elle peut en retirer vient précisément des spécificités de son produit. Mais on peut aussi imaginer que l'élaboration de nouveaux produits représente un coût trop important pour être supporté par une seule entreprise, ou que les produits innovants sont conçus avec des partenaires en dehors du secteur de l'entreprise, et donc sans risque de concurrence. Cette variable **DIFF** peut ainsi jouer dans un sens ou dans l'autre, et il sera intéressant de comparer nos intuitions aux résultats du modèle que nous avons développé.

Un autre facteur à prendre en compte dans cette étude est la volonté de l'entreprise de protéger légalement ses recherches. Une entreprise menant des recherches de R&D a en effet deux choix quant à la protection du résultat de ses recherches. Elle peut soit garder le résultat de ses recherches secret, sans le divulguer ni le protéger ou bien déposer des brevets afin de garantir le respect de ses droits sur ses inventions, mais dans ce cas ses travaux sont publiés et accessibles à tous, bien que leur utilisation soit sujette à l'obtention d'une licence d'utilisation. La variable

**PRO** prend en compte la protection juridique des recherches de l'entreprise sondées. C'est un *dummy* dont la valeur vaut 1 si l'entreprise brevète ses recherches, 0 sinon. Cette variable correspond à la réponse à la question 11.2 de l'enquête [2].

On peut imaginer qu'une entreprise qui protège ses recherches par des moyens publics juridiques est plus en mesure de collaborer avec d'autres entreprises pour mener une recherche en commun, puisqu'elle sait que ses droits sont protégés. A l'inverse, une entreprise qui fait le choix de ne pas déposer ses brevets pour garder ses résultats secrets ne se lancera pas dans une recherche en coopération avec l'extérieur. Ce raisonnement est sans doute valable pour les grandes entreprises, mais il faut considérer qu'une petite entreprise n'a pas les moyens de se protéger légalement, mais peut tout de même entreprendre une recherche en collaboration. Par ailleurs, la protection juridique des recherches d'une entreprise peut être un indicateur de sa méfiance vis à vis des autres entreprises. Il sera donc intéressant de regarder quel rôle notre modèle fait jouer à ce paramètre, s'il pousse ou dissuade une entreprise à mener des recherches en collaboration.

<b>Variables</b>	<b>Signification</b>	<b>Type de variable</b>
<b>IRD</b>	Investissement en R&D interne	donnée en euros
<b>S</b>	Taille de l'entreprise, chiffre d'affaire	donnée en euros
<b>M</b>	Multinationale	<i>Dummy</i>
<b>CHEM IT INFO MVEN</b>	Secteur d'activité	<i>Dummy</i>
<b>SUB</b>	Existence de subvention	<i>Dummy</i>
<b>ERD</b>	Dépenses en recherche externe	donnée en euros
<b>COOP</b>	Existence de recherche en collaboration	<i>Dummy</i>
<b>MONI</b>	Capacité à faire de la veille technologique	<i>Dummy</i>
<b>PRO</b>	Volonté de protéger la recherche interne légalement	<i>Dummy</i>
<b>DIFF</b>	Capacité à se différencier	Pourcentage investi dans des produits innovants

FIG. 1 – Tableau récapitulatif des variables utilisées

### 3 Le modèle

Nous utilisons un modèle de régression à deux équations simultanées. On peut consulter l'article de Veugelers [1], pour une explication détaillée des raisons pour lesquelles ceci est intéressant. En substance, ce modèle suppose qu'il existe une influence réciproque entre les dépenses internes en R&D et la coopération, et que ces deux variables peuvent être par ailleurs expliquées par des modèles qui leurs sont propres. Ainsi, certaines variables dépendantes (explicatives) de ces équations peuvent avoir un effet sur les dépenses internes et sur la coopération. Utiliser une seule équation n'est donc pas satisfaisant, car il ne nous permet pas de séparer leur influence directe, et leur influence "indirecte" sur chacune des variables indépendantes. Pour prendre un exemple, on remarquera que la taille de l'entreprise peut avoir un lien avec ses dépenses en R&D (pour des questions de budget), mais aussi sur la coopération (avantage de la taille pour mettre en oeuvre les coopérations). Une grande entreprise pourrait ainsi avoir tendance à avoir de grandes dépenses internes en R&D, et une forte probabilité de coopérer. Mais comme la présence d'une coopération influe elle-même sur les dépenses, on voit que la taille de l'entreprise à deux influences de natures différentes sur la dépense. Le modèle nous permet donc de séparer ces deux effets et d'évaluer leur importance.

Les équations du modèle sont les suivantes :

$$\begin{aligned} \log IRD = & x_0 + \log S(x_1 + x_2M) + x_3M + x_4 \log DIFF \quad (3.1) \\ & + x_5SUB + \log ERD(x_6 + x_7MONI) \\ & + COOP(x_8 + x_9MONI) \end{aligned}$$

$$\begin{aligned} COOP = & x_0 + \log S(x_1 + x_2M) + x_3M + x_4 \log DIFF \quad (3.2) \\ & + x_5PRO + x_6 \log IRD \end{aligned}$$

Takeshi Amemiya, dans [4], a décrit une méthode en deux étapes pour estimer des modèles mélangeant variables discrètes et continues. Les quatre étapes de résolution sont les suivantes :

- le modèle

$$\log IRD = \Pi_1 X + \sigma_1$$

est évalué par OLS, avec X comprenant toutes les variables indépendantes expliquant IRD et COOP.

- Le modèle

$$COOP = \Pi_2 X + \sigma_2$$

est évalué par Probit ML, avec X comprenant toutes les variables indépendantes expliquant IRD et COOP.

- Le modèle 3.1 est estimé par OLS en remplaçant COOP par la valeur donnée par le modèle estimé en deuxième étape.

- Le modèle 3.2 est estimé par Probit ML en remplaçant IRD par la valeur donnée par le modèle estimé en première étape.

On trouvera dans l'annexe de plus amples détails sur les procédures économétriques mises en oeuvre.

## 4 L'analyse des résultats

### 4.1 Les résultats

Les résultats obtenus à partir des données de l'étude CIS3 sont présentés dans le tableau 2.

Parmi les 7000 entreprises ayant répondu au questionnaire nous avons éliminé celles qui n'avaient pas répondu aux questions qui nous servent à la définition des variables du modèle (cf l'annexe page 24). C'est ainsi qu'il nous reste un ensemble statistique de 168 entreprises.

les résultats successifs sont présentés dans le tableau ci joint.

	OLS 1	Probit 1	OLS 2	Probit 2
Intercept	0.713 (0.388) 0.068	-0.065 (1.015) 0.949	0.726 (0.393) 0.066	0.211 (0.909) 0.816
log S	0.015 (0.114) 0.899	-0.281 (0.294) 0.339	0.062 (0.125) 0.620	-0.408 (0.275) 0.138
M*log S	0.110 (0.151) 0.467	0.467 (0.386) 0.227	0.031 (0.172) 0.858	0.360 (0.373) 0.334
M	-0.516 (0.758) 0.497	-2.008 (1.911) 0.293	-0.175 (0.840) 0.836	-1.567 (1.849) 0.397
SUB	0.332 (0.097) 0.001	0.847 (0.256) 0.001	0.188 (0.182) 0.304	
log ERD	0.549 (0.065) <.0001	0.050 (0.155) 0.748	0.541 (0.067) <.0001	
MONI* logERD	-0.046 (0.041) 0.262	0.131 (0.094) 0.164	-0.069 (0.055) 0.213	
log DIFF	0.225 (0.092) 0.016	0.283 (0.228) 0.215	0.176 (0.107) 0.101	0.128 (0.235) 0.585
PRO	0.095 (0.094) 0.316	0.556 (0.235) 0.018		0.561 (0.229) 0.014
MONI* COOPn			0.010 (0.209) 0.963	
COOPn			0.163 (0.225) 0.470	
log IRDn				0.515 (0.248) 0.038
N	168	168	168	168

FIG. 2 – Résultats des régressions. Le premier chiffre est la valeur estimé, le chiffre entre parenthèse la variance. Le dernier est le pourcentage de significativité.

## 4.2 Leur pertinence

Avant de commencer l'analyse de nos résultats, il est important d'en comprendre la pertinence. En effet, les valeurs retournées correspondent à la réalisation la plus probable au sens des moindres carrés de variables aléatoires. Ces résultats sont associés à des intervalles de confiance. L'annexe en page 20 explique comment le test de Student (ou de façon similaire le test du  $\chi^2$ ) permet d'estimer la pertinence du résultat trouvé. La valeur de ce test est à comparer à 2 : si elle est inférieure à 2, la valeur trouvée n'est pas significative pour un risque donné de 5%. Or nous observons qu'à l'exception de quelques variables comme *log ERD* ou *SUB*, les résultats retournés sont peu significatifs. Ils sont donc à prendre pour l'instant avec prudence.

D'autres problèmes entrent en jeu dans l'exploitation de ces résultats : les 168 entreprises finalement sélectionnées ne représentent pas un panel représentatif de l'ensemble des entreprises ayant une activité de recherche et développement. On s'aperçoit assez nettement que la plupart des entreprises de notre panel ont un chiffre d'affaire supérieur à 1 million d'euros. Les entreprises de taille plus modeste n'ont donc pas pris le temps de répondre à toutes les questions, dont certaines demandaient des recherches comptables au sein même des entreprises.

Par ailleurs certaines réponses étaient constituées d'une simple appréciation sous la forme d'une case à cocher du type : "fort", "moyen" ou "faible". Ce type de question sur des données subjectives pose deux problèmes majeurs :

- comment s'assurer que chaque personne ayant rempli un questionnaire avait en tête la même grille d'évaluation ?
- comment s'assurer que la personne ayant rempli le questionnaire avait une vision globale et non faussée de sa propre entreprise ?

Ce type de problème rend la définition de variable subjective particulièrement délicate. C'est le cas par exemple de la variable **MONI** qu'il nous a semblé intéressant d'introduire mais qui souffre d'un manque de définition objective. Nous avons choisi de considérer que des entreprises ayant affirmé se baser beaucoup sur une des sources d'information pour sa R&D faisait de la veille technologique. La variable **PRO** est également dans ce cas : les entreprises disposent de plusieurs moyens de se protéger, formels ou informels. Nous considérons que les entreprises qui se protègent sont celles qui utilisent au moins une protection de chaque type.

Pour ces raisons, l'exploitation des résultats que nous allons faire est à prendre avec beaucoup de prudence. Ils devraient être confrontés à d'autres expériences avant d'être validés.

## 4.3 Leur analyse

Conscients des limitations précédentes, nous allons faire ressortir les grandes tendances pour chacune des variables de notre modèle et mettre en parallèle ces résultats avec les intuitions développées précédemment.

**variable S :**

A l'issue du deuxième OLS, la variable **S** est estimée inférieure à 1, mais elle a un écart-type important. Cette valeur confirme que plus une entreprise a un chiffre d'affaires important, moins la part de son chiffre d'affaires consacrée à la R&D interne est importante. Ce résultat semble logique, en effet une grande entreprise peut faire des économies d'échelle en matière de R&D. Même si les tests ne sont pas très concluants pour valider cette valeur, elle va néanmoins dans le sens d'études précédentes, comme celle de Veugelers en Belgique [1]. On note cependant que les résultats du probit montrent que les petites entreprises semblent avoir plus intérêt à coopérer que les grandes (avec une bonne certitude pour les tests).

**variable M :**

Une autre question que l'on se posait était de savoir si les entreprises multinationales avaient un comportement différent des autres en matière de R&D interne et de coopération. En ce qui concerne la coopération, on obtient des valeurs non nulles des coefficients, mais les tests ne sont pas concluants et ne permettent pas d'affirmer avec certitude que cette variable a un réel impact. On peut rajouter que les entreprises de l'échantillon considéré ont des tailles à peu près comparables et que les effets des deux occurrences de **M** dans l'équation ont tendance à s'annuler. Les résultats sont un peu plus probants dans la première équation. Ils semblent montrer que les multinationales dépensent globalement moins en R&D interne, mais que la taille est un facteur plus déterminant pour les multinationales que pour les entreprises locales.

**variable SUB :**

Dans notre modèle, la variable **SUB** est un dummy, donc ne reflète qu'incomplètement le montant des subventions reçues par les entreprises. Toutefois, les résultats obtenus sont relativement concluants. Les subventions semblent bien influencer positivement sur les dépenses de R&D interne des entreprises. L'absence de valeur pour ces subventions empêche cependant de quantifier cet apport.

**variable ERD :**

Comme on pouvait s'en douter, les dépenses externes et internes en matière de R&D évoluent dans le même sens. Pour cette variable, la méthode OLS permet d'obtenir des résultats très probants, l'écart-type du coefficient étant faible. On peut observer que les dépenses internes de R&D varient comme la racine carrée des dépenses externes (coefficient de 0.54). Pour les entreprises qui font de la veille technologique, (variable **MONI**), les dépenses internes croissent légèrement plus lentement avec les dépenses externes. Cette différence n'est toutefois pas réellement significative, et pourrait provenir de la petite taille de l'échantillon choisi.



### **variable DIFF :**

Les entreprises émettant un produit innovant dépensent plus en R&D interne que les autres, ce qui s'explique aisément par les moyens mis en oeuvre pour créer ces nouveaux produits. Le coefficient associé à  $\log(DIFF)$  dans la première équation est d'environ 0.2, et même en prenant en compte les erreurs liées au modèle, cette valeur reste presque sûrement entre 0 et 0.5. Là encore, pour les nouveaux produits, même en prenant en compte le coefficient de  $S$ , les entreprises bénéficient d'économies d'échelle, même si elles sont moins importantes.

La deuxième équation ne donne pas de résultat probant en ce qui concerne la variable **DIFF**. Les entreprises ayant une volonté d'innovation semblent dans l'ensemble un peu plus enclines à coopérer en matière de R&D, cependant on ne peut pas considérer cela comme une règle.

### **variable PRO :**

Malgré les incertitudes liées à sa définition, la variable **PRO** permet de dégager un résultat intéressant : le coefficient qui lui est associé est significativement positif, ce qui tend à prouver que les entreprises qui se protègent efficacement ont plus tendance à coopérer, ou que réciproquement les entreprises qui coopèrent sont plus soucieuses de protéger leurs résultats. Là encore, l'écart-type associé à ce dummy est important. Plus que de règle mathématique, on peut parler de tendance globale pour ce phénomène.

## **4.4 La coopération se substitue-t-elle aux dépenses de R&D ? Les variables COOP et IRD**

Le Probit permet d'obtenir un résultat intéressant sur ces deux variables. En effet, le coefficient obtenu pour **IRDn** est positif, même si la dispersion des valeurs est forte. Cela signifie que les entreprises ayant un fort département R&D coopèrent globalement plus que les autres. La coopération semble donc être le fruit d'une volonté d'investir dans la R&D en trouvant de manière efficace les compétences dont une entreprise a besoin, et non un moyen de partager des dépenses considérées comme secondaires par la direction. On peut également considérer que les entreprises qui font beaucoup de R&D ne semblent pas se sentir lésées dans la coopération, même si elles apportent plus que leur partenaire.

En revanche, l'autre équation montre que l'influence de la coopération sur les dépenses R&D est plus réduite avec les variables prises en compte dans cette équation. Si le coefficient associé à **COOPn** est positif, on a une forte dispersion des valeurs. Cela pourrait s'expliquer par le fait que **COOP** est un dummy, et que **COOPn** ne quantifie pas forcément efficacement la coopération. On peut également signaler que l'influence conjointe de la coopération et de la veille technologique sur les dépenses internes en R&D est nulle.

## 4.5 Comparaison avec une autre étude.

Comme signalé précédemment, notre étude se veut le pendant français d'une étude réalisée par Veugelers [?] en Belgique il y a une dizaine d'années.

Avant toute autre chose, si les deux modèles utilisés sont proches, certaines variables sont différentes car les questionnaires servant de sources étaient différents. Certaines variables continues sont devenues des dummies, certaines sont apparues et d'autres ont disparu. Il semble donc difficile de comparer quantitativement les résultats des deux études. Nous pouvons cependant tirer quelques conclusions sur le plan qualitatif.

- Un résultat partagé par les deux études, et que Veugelers présente comme étant confirmé par d'autres sources, est l'économie d'échelle réalisée par les grandes entreprises dans le domaine de la R&D. Elles semblent simplement beaucoup plus fortes dans notre cas.
- De la même manière, dans les deux études, les subventions font augmenter les dépenses internes de R&D.
- Les dépenses internes et externes de R&D apparaissent dans les deux cas comme complémentaires, les coefficients qui les relient étant positifs de manière significative. Chez Veugelers, la veille technologique joue un rôle important sur cette relation, que notre étude ne permet pas de déterminer. Notre étude suggère toutefois un lien bien plus important entre ces recherches interne et externe.
- Dans les deux cas, la coopération est avantageuse pour des entreprises de petite taille ayant un budget important alloué à la R&D. Ceci dénote une volonté de s'orienter vers la recherche, et la coopération peut-être un moyen efficace dans cette optique.
- Pour les multinationales, les économies d'échelle sont moins importantes que pour des entreprises locales, et ce d'après les deux études. Cependant, dans notre étude, ce résultat a plus valeur de tendance générale que de règle.
- Notre modèle permet de dégager une influence non négligeable de la volonté d'innovation sur les dépenses internes de R&D, ce qui contredit les conclusions de Veugelers.

## Conclusion

A la vue d'un certain nombre de publications, l'idée selon laquelle l'externalisation des dépenses de R&D se substitue aux efforts internes de développement s'avère simpliste. En effet, la littérature actuelle tend d'avantage à démontrer que R&D interne et externe sont complémentaires. Dans notre étude, nous nous sommes intéressés à cette problématique en nous fondant sur les résultats de l'étude CIS-3 France.

Pour ce faire, nous avons introduit un modèle d'équations couplées faisant intervenir un certain nombre de variables pertinentes. Ce modèle prend en compte les interactions mutuelles entre dépenses internes et coopération. Il fait intervenir la notion de capacité d'absorption, conformément aux intuitions de la théorie.

L'étude que nous avons menée a donc permis de mettre en évidence certains faits caractéristiques de la coopération dans le domaine de la R&D, en accord avec la littérature théorique. Elle a tendance à confirmer l'étude faite par Veugelers [1] sur des entreprises flamandes tout en introduisant de nouvelles variables pertinentes dans le cadre de notre analyse.

Toutefois, nos résultats sont à prendre avec réserve, dans la mesure où nous avons dû faire face à un certain nombre de difficultés pratiques. L'échantillon d'entreprises sur lesquelles nous avons travaillé est de taille réduite, et les tests statistiques servant à la validation des résultats ne permettent pas d'aboutir à des conclusions tranchées.

## Références

- [1] Reinhilde Veugelers, "Internal R&D Expenditures and External Technology Sourcing", *Research Policy* 26, pp. 303–315, 1997.
- [2] The Third Community Innovation Survey : CIS3, Eurostat France, 1998–2000.
- [3] <http://www.dr-belair.com/dic/Commerce/Ressources/ape-naf.htm>
- [4] Amemiya, T., 1979, The Estimation of a Simultaneous Equation Tobit Model, *International Economic Review*.
- [5] Keneth Arrow, 1962, Economic Welfare and the allocation of resources for invention
- [6] J. Schumpeter, 1912, *Théorie de l'innovation économique*
- [7] C. d'Aspremont, A. Jacquemin, 1988, Cooperative and Noncooperative R&D in Duopoly with Spillovers
- [8] M.I. Kamien, E. Müller, I. Zang, 1992, Research Joint Ventures and R&D Cartels
- [9] W.M. Cohen, D.A. Levinthal, 1989, Innovation and Learning : The Two Faces of R&D
- [10] Sanjeev Goyal, José Luis Moraga, 2000, R&D Networks
- [11] Mathieu Manant,

## 5 Annexe : Quelques mots d'économétrie...

### 5.1 Moindres carrés ordinaires et modèle linéaire multiple

Le modèle linéaire multiple postule que la grandeur à expliquer est une expression linéaire des variables explicatives retenues, perturbée par un aléa.

Ainsi le modèle de variable à expliquer :  $P$ , et à trois variables explicatives :  $X$ ,  $Y$  et  $Z$  (ou plutôt quatre en comptant explicitement la constante), s'écrit :

$$P = a + b.X + c.Y + d.Z + \epsilon$$

où  $a$ ,  $b$ ,  $c$  et  $d$  sont les coefficients, inconnus, du modèle, et  $\epsilon$  la perturbation aléatoire ou aléa.

Le problème est alors d'estimer les paramètres du modèle à partir des  $N$  observations des variables :  $P_i$ ,  $X_i$ ,  $Y_i$  et  $Z_i$ .

La méthode des mco (pour "moindres carrés ordinaires", ou *olsq*, pour "ordinary least squares", chez les Anglo-Saxons) revient à chercher l'expression :

$$P = a + b.X + c.Y + d.Z$$

qui minimise la somme des carrés des erreurs d'ajustement :  $u_i = P_i - (a + b.X_i + c.Y_i + d.Z_i)$ , ou résidus, pour les valeurs observées.

Les valeurs estimées pour  $a$ ,  $b$ ,  $c$  et  $d$  sont les estimations des moindres carrés des coefficients du modèle (les logiciels économétriques les calculent par des méthodes matricielles appropriées).

La méthode permet de calculer le coefficient  $R^2$ , appelé coefficient de corrélation multiple. Si une constante figure parmi les explicatives, les résidus sont de somme et de moyenne nulle, tandis que la variable ajustée a même moyenne que la variable observée. Le coefficient  $R^2$  est compris entre 0 et 1, la proximité de 1 traduisant toujours une forte liaison linéaire multiple, mais d'un seul point de vue descriptif, l'introduction de variables explicatives quelconques le faisant croître mécaniquement...

#### Propriétés des estimateurs des mco

On suppose que les aléas :  $\epsilon_i$ , suivent la même loi normale :  $N(0, \sigma)$ , centrée, et d'écart-type :  $\sigma$  (inconnu), et sont indépendants, ces présupposés sont généralement appelés les hypothèses des moindres carrés. L'hypothèse de normalité n'est pas déraisonnable si l'on admet que ces aléas résultent de l'addition de nombreux effets secondaires mineurs indépendants (cf. théorème de la limite centrale). Sous ces conditions :

- Les résidus calculés :  $u_i$ , approchent les aléas inconnus :  $\epsilon_i$ , et la quantité :

$$\frac{1}{2} \frac{\sum_i u_i^2}{(N - 2)}$$

liée à la somme des carrés des résidus, est une bonne estimation de l'écart-type :  $\sigma$ , de l'aléa. Elle est appelée : écart-type résiduel.

- Les estimateurs de  $a$ ,  $b$ ,  $c$  et  $d$  suivent des lois normales dont les espérances sont les quantités à estimer : ces estimateurs sont sans biais.
- Les écart-type des estimateurs de  $a$ ,  $b$ ,  $c$  et  $d$ , peuvent être estimés.

### Test de significativité des coefficients

Ces estimateurs suivant donc des lois normales, et leurs écart-type pouvant être estimés, on peut opérer le test de Student de significativité pour chaque coefficient.

Il s'agit de tester si, pour un niveau de confiance donné (en général 95%), l'intervalle de confiance peut ou non contenir la valeur : 0. En effet si la valeur véritable du coefficient peut être 0, il n'est même pas certain que la variable explicative (ou le terme constant) intervienne réellement dans le modèle.

Sachant que pour un risque :  $\alpha$ , l'intervalle de confiance pour l'estimé de  $a$  suivant une loi  $N(\mu, v)$  est :

$$[\mu - t_\alpha * v; \mu + t_\alpha * v]$$

le test revient à examiner si le rapport :

$$\frac{|\text{coefficient estimé}|}{\text{écart-type estimé}}$$

dépasse ou non  $t_\alpha$ .

On fait en général ce test au risque  $\alpha = 5\%$ , ce qui donne, en utilisant la valeur approchée :  $t_{0,05} = 1,96 \approx 2$  :

- $\frac{|\text{coefficient estimé}|}{\text{écart-type estimé}} < 2$  : coefficient non significatif au risque 5%
- $\frac{|\text{coefficient estimé}|}{\text{écart-type estimé}} > 2$  : coefficient significatif au risque 5%.

Lors d'une étude économétrique, le test de Student sur chacun des coefficients est beaucoup plus important que l'examen du coefficient de corrélation.

Un "bon" test de Student doit toutefois être regardé avec une certaine modestie, ce test suppose en effet la pertinence du modèle, mais il n'a pas vocation à la confirmer ; en fait, il sert essentiellement à mettre en doute ou à écarter les variables d'influence incertaine.

## 5.2 Du modèle exponentiel au modèle linéaire

Une relation non linéaire usuelle entre des variables explicatives d'un modèle est la liaison exponentielle. Ainsi la classique fonction de production de type Cobb-Douglas :

$$P = a.X^b.Y^c.Z^d$$

où  $X, Y$  et  $Z$  sont les explicatives, et  $a, b, c$  et  $d$  en sont les paramètres inconnus est un modèle non linéaire.

Ce modèle peut néanmoins être linéarisé par passage aux logarithmes. Le modèle, par exemple, se transforme en :

$$\ln(P) = \ln(a) + b.\ln(X) + c.\ln(Y) + d.\ln(Z)$$

qui est un modèle linéaire équivalent ; la variable à expliquer étant à présent  $\ln(P)$ , les explicatives  $\ln(X)$ ,  $\ln(Y)$  et  $\ln(Z)$ , et les coefficients  $\ln(a)$ ,  $b$ ,  $c$  et  $d$ .

Le modèle ainsi linéarisé peut alors être estimé par régression linéaire. On remarque toutefois que la perturbation aléatoire du modèle initial (que l'on n'avait pas écrite explicitement) doit avoir une forme multiplicative particulière, dite log-normale, pour que celle du modèle transformé vérifie les hypothèses des moindres carrés ordinaires.

Il faut également noter que les coefficients sous leur forme initiale n'étant fréquemment pas des expressions linéaires des coefficients transformés, leurs estimations indirectes n'ont plus nécessairement les propriétés d'optimalité des estimations par les moindres carrés ordinaires de ces derniers.

Ainsi pour conclure sur ce modèle, on peut se demander pourquoi choisir un modèle exponentiel plutôt qu'un modèle linéaire. La principale raison est que ce modèle a une élasticité constante de  $P$  en fonctions de  $X$ ,  $Y$  et  $Z$ .

### 5.3 Modèle Probit

Si on cherche à modéliser de façon linéaire une probabilité ou une variable binaire notée  $P$ , par définition comprise entre 0 et 1, en fonction de deux variables explicatives. Une première idée est d'étudier un modèle tel que

$$P = a + b.X + c.Y + d.Z$$

où  $a$ ,  $b$  et  $c$  sont les variables à estimer. Par définition, on appelle la variable  $a$  l'*intercept*.

Il est clair qu'un tel modèle est naïf et peu adapté, les valeurs ajustées de  $P$  pouvant a priori parcourir l'ensemble des réels et non simplement l'intervalle  $[0, 1]$ .

Une idée destinée à remédier à l'inconvénient précédent est de passer par l'intermédiaire d'une fonction  $F$  dont l'ensemble des valeurs est l'intervalle  $[0, 1]$ , et d'estimer un modèle de la forme :

$$P = F(a + b.X + c.Y + d.Z)$$

où encore

$$F^{-1}(P) = a + b.X + c.Y + d.Z$$

dans le cas de deux explicatives.

Les fonctions de répartition associées aux distributions de probabilité conviennent. Si on prend pour  $F$  la fonction de répartition associée à la loi normale  $N(0, 1)$ , le modèle est le modèle Probit. La fonction  $F$ , ni son inverse ne s'expriment sous une forme résolue, mais peuvent être calculées numériquement.

On va présenter ce modèle d'une manière quelque peu différente, quoique mathématiquement équivalente, dans le cadre général des modèles à variable latente, non observable.

### Modèle Probit à variable latente

Conservant pour fixer les idées deux variables explicatives,  $X$  et  $Z$ , on suppose que la réalisation de l'événement étudié ( $P = 1$ ) dépend d'une variable latente :  $P^*$ , non observable :

$$P^* = a + b.X + c.Y + d.Z + \epsilon$$

dépendant elle-même des variables explicatives et d'une perturbation aléatoire  $\epsilon$ , et telle que

$$\begin{cases} P = 1 & \text{si } P^* > 0 \\ P = 0 & \text{sinon} \end{cases}$$

par suite

$$\begin{aligned} \mathbb{P}(P = 1) &= \mathbb{P}(P^* > 0) \\ &= \mathbb{P}(\epsilon > -a - b.X - c.Y - d.Z) \\ &= \mathbb{P}(\epsilon < a + b.X + c.Y + d.Z) \\ &= F(a + b.X + c.Y + d.Z) \end{aligned}$$

en supposant la loi de la perturbation  $\epsilon$  centrée et symétrique, et en notant encore  $F$  sa fonction de répartition.

La variable  $P^*$  n'intervenant que par son signe, on voit que les coefficients  $a$ ,  $b$ ,  $c$ , et  $d$  ne sont définis qu'à un facteur d'échelle près, qu'on peut fixer librement, par exemple via l'écart-type de  $\epsilon$ .

Comme il a été dit, si  $\epsilon$  suit la loi normale centrée réduite, on a le modèle Probit.

### Estimation, interprétation des résultats

Le modèle précédent est estimé par la méthode du maximum de vraisemblance.

Les coefficients étant définis à un facteur multiplicatif près, ce n'est pas leur valeur mais leur significativité et leur signe (qui indique le sens de l'influence de la variable considérée) que l'on examine.

Pour conserver l'exemple précédent, on voit que si l'incidence de la variable  $X$  sur  $P^*$  est linéaire (et mesurée par l'effet marginal  $b$ ), celle sur la probabilité,  $\mathbb{P}(P = 1)$ , ne l'est nullement et dépend des valeurs des variables explicatives au point considéré. Pour cette raison, on calcule fréquemment l'élasticité de la probabilité en des points donnés, par exemple au point moyen de la population.

On a également coutume de calculer le taux de bien classés : pour chaque observation, on détermine sa probabilité théorique par le modèle estimé, puis on examine toutes les paires associant un événement positif ( $P = 1$ ) et un événement négatif ( $P = 0$ ), et on calcule enfin la proportion de paires pour lesquelles les probabilités calculées sont dans le sens logiquement attendu, c'est à dire la plus grande pour l'événement observé. Il est encourageant de trouver un bon taux de bien classés, mais il serait déraisonnable d'attendre un nombre très proche de 1, compte tenu notamment du caractère aléatoire du modèle.



## **6 Annexe : Origine des données (Intitulés des questions posées)**

### **Variable : IRD ( Question 4.1)**

Did your enterprise engage in the following innovation activities in 2000 ?

- Intramural Research & experimental development

Total innovation expenditure in 2000 ?

### **Variable : S ( Question 0.5)**

Total turnover market sales of goods and services in year 2000, including export and taxes except VAT ?

### **Variable : CHEM, MVEN) ( NAF-APE700)**

Provient de l'identifiant NAF-APE700 de chaque entreprise.

### **Variable : M ( Question 0.1)**

Is the enterprise part of an enterprise group ?

If yes, what is the country of head office of the group ?

### **Variable : SUB ( Question 7.1)**

Did your enterprise receive any public financial support for innovation activities during the period 1998- 2000 ?

Local or regional authorities, Central government (including institutions working on behalf of central government) or The European Union

### **Variable : ERD ( Question 4.1)**

Did your enterprise engage in the following innovation activities in 2000 ?

- Acquisition of R&D (extramural R&D)

Total innovation expenditure in 2000 ?

### **Variable : COOP ( Question 8.1)**

Did your enterprise have any co-operation arrangements on innovation activities with other enterprises or institutions during 1998-2000 ?

### **Variable : MONI ( Question 9)**

The main sources of information needed for suggesting new innovation projects or contributing to the implementation of existing projects are asked in this question. Please indicate the degree of importance attached to various alternative information sources :

- Internal sources
- Market sources
- Institutional sources
- Other sources

**Variable : PRO ( Question 11.2)**

During the period 1998-2000, did your enterprise, or enterprise group, make use of any of these other methods to protect inventions or innovations developed in your enterprise ?

- Formal methods (Registration of design patterns, Trademarks, Copyright)
- Strategic methods (Secrecy, Complexity of design, Lead-time advantage on competitors)

**Variable : DIFF ( Question 1.4)**

During the period 1998-2000, did your enterprise introduce new or significantly improved products (goods or services) not only new for your enterprise, but also new for your enterprise's market ?

If yes, estimate the contribution of these products in total turnover in 2000.