UNIVERSITÉ DE MONTRÉAL

FIRM-SPECIFIC CHARACTERISTICS IMPACTING COLLABORATIVE BEHAVIOR: THE CASE OF THE CANADIAN BIOTECHNOLOGY INDUSTRY

ALESSANDRO CESCHIA
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INDUSTRIEL)
AVRIL 2008

© Alessandro Ceschia, 2008.

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-41549-8
Our file Notre référence
ISBN: 978-0-494-41549-8

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

FIRM-SPECIFIC CHARACTERISTICS IMPACTING
COLLABORATIVE BEHAVIOR:
THE CASE OF THE CANADIAN BIOTECHNOLOGY INDUSTRY

présenté par: CESCHIA Alessandro
en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées
a été dûment accepté par le jury d'examen constitué de:

Mme DE MARCELLIS-WARIN Nathalie, Doctorat
Mme BEAUDRY Catherine, D.Phil.
M. McNIVEN Chuck, B.A. (honours), M.A.

A Marta

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor, Prof. Catherine Beaudry for giving me the opportunity to work on this project. Her expertise, patience and dynamic character assisted me in tackling this challenging task.

I would like to thank Prof. Nathalie de Marcellis-Warin whose contagious passion acts as a stimulus for all of the members of the research group, *InnovaRisQ*.

I am grateful to the staff of Statistics Canada, in particular to Charlene Lonmo and Chuck McNiven. Without their help, this work would not have been possible.

This research would have hardly been possible without the financial assistance of CRSH and I express my gratitude to this organization.

I would like to thank the staff of the Industrial Engineering department, particularly the members of *InnovaRisQ*. A very special thank you to Rudy for his valuable and sincere friendship.

Finally, I am grateful to my family and friends for the support they have provided me my entire life.

Ai miei genitori, grazie per la vita.

RÉSUMÉ

Le rôle de la collaboration et de la coopération dans le domaine de la biotechnologie au Canada est primordial au développement et à la commercialisation de ces technologies. De nombreuses recherches étudient donc la collaboration. Ce projet adopte quant à lui une nouvelle approche et cherche à différencier les caractéristiques des entreprises qui collaborent de celles qui ne le font pas. Les enquêtes de Statistique Canada sur l'utilisation et de développement de la biotechnologie (1999, 2001, 2003 et 2005) sont utilisées afin de répondre à plusieurs questions : existe-t-il une différence entre les PME et les grandes entreprises, tant du point de vue de la propension à participer à des alliances que du nombre de ces ententes de collaboration ? Existe-t-il une différence entre les entreprises de biotechnologie qui collaborent afin d'avoir accès à des connaissances non disponibles à l'intérieur et celles qui collaborent pour la production et commercialisation de nouveaux produits et processus ? Existe-t-il une différence entre les entreprises de biotechnologie qui collaborent avec une autre entreprise et celles qui collaborent avec une institution publique ? Quelles sont ces caractéristiques distinctives ?

À l'aide de modèles logit, nous séparons les firmes qui collaborent de celles qui ne collaborent pas de façon à expliquer cette dichotomie par certaines de leurs caractéristiques : taille, âge, formation par essaimage, contrats, brevets, orientation vers la R&D en biotechnologie, financement et import/export. Dans un deuxième temps, l'utilisation de régressions binomiales négatives nous permet d'expliquer l'effet d'un certain nombre d'attributs des firmes sur leur propension à collaborer et sur le nombre d'ententes de collaboration qui en résultent. Nous analysons la propension à collaborer en distinguant deux différents types d'alliances (notamment d'une part les alliances visant à obtenir l'accès aux connaissances externes et d'autre part les alliances pour la production et la commercialisation des nouveaux produits et processus) et deux types de partenaires (institutions publiques et entreprises).

Nous montrons que, parmi les petites et moyennes entreprises de biotechnologie, la taille et l'orientation vers la biotechnologie sont les facteurs qui jouent un rôle déterminant, et influencent positivement la propension et l'intensité de la collaboration. Aussi, les entreprises issues de l'essaimage ont une plus grande propension à collaborer et à avoir un plus grand nombre d'ententes de collaboration. D'autres caractéristiques, comme par exemple les stratégies de protection de la propriété intellectuelle, ont une influence sur la propension à participer à des alliances de type spécifique, ou avec un partenaire spécifique. Toutefois, des recherches ultérieures sont nécessaires pour en comprendre leur influence, dans un encadrement plus complexe.

En connaissant la dynamique et les relations cause-effet qui influencent la propension à collaborer et en intégrant ce travail avec d'autres recherches, il sera possible de comprendre et d'évaluer l'impact des politiques publiques visant à soutenir les entreprises de biotechnologie, et en favoriser la croissance.

ABSTRACT

The collaboration and cooperation within the Canadian biotechnology industry is integral to the development and commercialization of innovations. The perspective adopted in this work aims at identifying firm-specific characteristics that influence collaborative behavior. We used data collected in the « biotechnology use and development surveys » (1999, 2001, 2003 and 2005) to answer the following questions: is there a difference in the propensity and in the intensity of collaboration among small, medium and large biotechnology firms? Is there a difference between firms collaborating for knowledge-related reasons and those collaborating for production or commercialization? Is there a difference between firms collaborating with another firm and those collaborating with a public institution? What are the distinguishing characteristics?

Through the logit analysis, we detect the differences between collaborative and non-collaborative biotechnology firms in Canada, with respect to the following relevant characteristics: size, age, generation through a spin-off, contracts, licensed and obtained intellectual property rights, biotechnology and R&D orientation, capital financing, and import/export. Through a negative binomial analysis, we explain the effect of the aforementioned characteristics on the number of collaborative arrangements a firm is involved in. We distinguish partnerships by the reason leading to their formation (alliances related to knowledge and alliances related to production and commercialization) and by the type of partner (a firm or a public institution).

We show that size, biotechnology orientation and generation through a spin-off are the most important determinants in explaining a higher propensity to collaborate and a higher intensity of alliances among small and medium firms. In addition, other characteristics such as IP protection strategies play a role in shaping collaborative behavior with a specific partner or for a specific reason, and need further in-the-field investigation to fully asses their impact.

By integrating this work with additional research aimed at tracing a complete description of the cause-effect relationships between collaborative behavior and firm-specific characteristics, it would be possible to analyze and evaluate public policies supporting small biotechnology firms and fostering research and development of innovations.

CONDENSÉ EN FRANÇAIS

1. INTRODUCTION

Ce travail vise à distinguer les caractéristiques des entreprises innovantes en biotechnologie qui collaborent de celles collaborent peu ou pas du tout, tout en contrôlant les variations provinciales de l'environnement économique. La part des petites et moyennes entreprises de biotechnologie au Canada qui participent à des ententes de collaboration a diminué, passant de 60% en 2001 à environ 50% en 2003 (Statistics Canada, 2005). Pourtant, Oliver (2004) associe l'absence d'alliance à la mort des entreprises de cette industrie, corroborant ainsi l'idée que les alliances interorganisationnelles sont essentielles à la survie des entreprises. En effet, en biotechnologie, le processus de développement, de production et de commercialisation d'un nouveau produit est extrêmement complexe, et requiert plusieurs compétences et connaissances spécifiques et très diverses. Il est donc préférable que chaque agent du système d'innovation soit spécialisé dans un domaine particulier pour qu'il puisse ajouter de la valeur à une étape du processus de façon efficace et efficiente. Ainsi, le réseau de valeur en biotechnologie tend à prendre la forme d'une chaîne d'alliances (Stuart et al., 2007) qui suit les étapes du développement du produit, depuis la recherche fondamentale jusqu'à la mise sur le marché. En se focalisant sur les petites entreprises qui sont au cœur de l'innovation dans cette industrie, nous reconnaissons plusieurs types d'alliances qui sont formées. Par exemple, dans une optique stratégique consistant à obtenir l'accès à de nouvelles idées et connaissances à l'état embryonnaire, les petites entreprises peuvent participer à des ententes de collaboration avec

des institutions publiques (universités, hôpitaux ou laboratoires gouvernementaux). La finalité de ce type d'alliance est de s'approprier une technologie nouvelle pour la performer en un produit ou un processus économiquement rentable. Par la suite, les petites entreprises ne disposent très souvent pas des ressources et des compétences nécessaires pour la production, l'expérimentation et la commercialisation d'un produit. A travers une alliance avec une grande compagnie pharmaceutique (dans le cas des biotechnologies liées au secteur de la santé humaine, le secteur majoritairement impacté par les biotechnologies), une petite entreprise innovante peut obtenir l'accès aux compétences nécessaires. Du point de vue des grandes firmes, collaborer avec des petites firmes orientées vers la recherche et développement (et qui donc innovent plus) est une activité nécessaire pour transformer, adapter ou renouveler le savoir-faire technologique et fournir de nouveaux produits ou processus commercialisables.

2. OBJECTIFS ET HYPOTHÈSES

A partir de la littérature existante sur le sujet, nous avons déterminé quelles sont les caractéristiques qui potentiellement ont un effet sur la propension à collaborer et sur l'intensité de cette collaboration, mesurée par le nombre d'ententes de collaboration auxquelles une firme participe. Ces caractéristiques sont les suivantes :

2.1. <u>Produits et processus de biotechnologie</u> : nous visons à évaluer l'effet des produits (à différentes étapes du développement) et des processus développés par la firme sur sa propension à collaborer.

2.2. Taille de l'entreprise : nous supposons que la taille d'une entreprise de biotechnologie n'influence pas la propension à collaborer. Néanmoins, nous supposons que cette caractéristique a un effet positif sur l'intensité de la collaboration.

2.3. Age de l'entreprise : L'effet de l'âge de l'entreprise sur la propension et l'intensité de la collaboration est difficile à évaluer. En fait il faudrait une modélisation dynamique qui analyse l'évolution de l'entreprise dans le temps pour prendre en compte cette caractéristique.

2.4. Orientation vers la biotechnologie : nous supposons que plus une entreprise est orientée vers la biotechnologie, notion que nous mesurons à travers le pourcentage d'employés en recherche et développement ayant une responsabilité liée à la biotechnologie, plus elle est enclin à collaborer et à participer à davantage d'alliances.

2.5. Stratégie d'appropriation de la propriété intellectuelle : bien que la littérature ne montre pas une forte corrélation entre le nombre de brevets obtenus par une entreprise et les alliances, nous supposons qu'une firme de biotechnologie qui collabore est plus enclin à protéger sa propriété intellectuelle. En fait, la « direction » de l'effet de cette caractéristique n'est pas claire et certaine, et plus de recherches seront nécessaires pour mieux comprendre cette dynamique.

2.6. Orientation vers l'exportation des produits/processus de biotechnologie : d'après la littérature existante sur le sujet, l'intensité d'exportation de produits et processus de biotechnologie est un facteur déterminant pour la propension à collaborer dans certains pays. Quand il est significatif, son effet est positif. Nous supposons donc

qu'une activité d'exportation majeure est associée à une plus grande propension à la collaboration et à un nombre supérieur d'alliances.

2.7. Aide publique : souvent, les projets d'aide publique demandent que l'entreprise participe à des alliances avec d'autres agents du système d'innovation. Comme il est déclaré dans un rapport gouvernemental du 2007, le Canada vise à augmenter la productivité des entreprises innovantes en les encourageant à former des ententes de collaboration. Nous supposons donc que les entreprises qui reçoivent du financement public montrent une propension supérieure à la collaboration.

2.8. Formation par essaimage : Nous supposons que les entreprises de biotechnologie formées par essaimage ont une plus grande propension à collaborer par rapport aux autres types d'entreprises. Néanmoins, il n'est pas clair que cette caractéristique a un effet sur le nombre d'alliances auxquelles l'entreprise participe.

En testant ces hypothèses selon les méthodes présentées dans la section suivante, nous voulons donc distinguer les entreprises qui collaborent de celles qui ne le font pas. Nous considérerons également différents types d'alliances selon la raison de la collaboration (accès aux connaissances ou pour la production et commercialisation) et selon les partenaires (institutions publiques ou autres entreprises). En outre, comme nous l'annoncions précédemment, nous estimerons l'effet des caractéristiques présentées sur le nombre d'ententes de collaboration formées par une firme.

3. MÉTHODOLOGIE

Les données de Statistique Canada (enquêtes sur l'utilisation et le développement de la biotechnologie 1999, 2001, 2003 et 2005) seront notre source de données, en particulier les questions spécifiques ayant trait à la collaboration ainsi que les questions traitant des caractéristiques présentées dans la section précédente. Afin d'atteindre nos objectifs de recherche, nous analyserons les données selon trois approches différentes. La première consiste à examiner les différences entre diverses sous populations des firmes de biotechnologie au Canada (divisées par taille et localisation) à l'aide des statistiques descriptives et des tables de contingence. La deuxième consiste en une analyse logit de la propension à établir des ententes de collaboration. Cette analyse comporte plus précisément trois volets distincts : les entreprises qui collaborent ou pas, les partenaires des ententes de collaboration (institutions publiques ou autre firme), et le type de raison des ententes de collaboration (alliances pour avoir accès à la connaissance et alliances pour la production et commercialisation). La troisième analyse vise à examiner le nombre d'ententes de collaboration établies par diverses firmes à l'aide de régressions binomiales négatives. Cette analyse comprend les mêmes volets utilisés pour l'analyse logit : nombre total d'alliances, nombre d'alliance par type de partenaire (institutions publiques ou autres entreprises), nombre d'alliances par type de raison (pour accéder à des connaissances externes ou pour la production et la commercialisation).

Pour l'analyse des régressions, nous utilisons plusieurs modèles pour déterminer le groupe de variables indépendantes qui décrivent mieux la propension des entreprises à collaborer et l'intensité de collaboration.

4. RÉSULTATS

4.1. STATISTIQUES DESCRIPTIVES

Bien que le nombre total des entreprises a constamment cru (elles étaient 358 en 1999, 375 en 2001, 490 en 2003 et 532 en 2005), la part de celles qui collaborent a chuté, entre 2001 et 2003 d'environ 10% (les firmes qui participaient à des ententes de collaboration en 2001 étaient le 60.3%, et en 2003 le 51.2%). Malheureusement, en 2005 cette crise de collaboration n'a pas rebondi. En distinguant les alliances par type de partenaire, en 1999 le 47.8% des firmes ayant au moins une entente de collaboration, avait comme partenaire une autre entreprise. Ce pourcentage était égal à 61.5% en 2001, à 47.4% en 2003 et à 62.3% en 2005. La diminution de l'intensité de collaboration avec un partenaire public est anticipée de deux ans par rapport à la collaboration avec une autre entreprise. En 1999, 54.5% des firmes qui collaboraient avaient un partenaire public, alors qu'en 2001 elles n'étaient que 36.7%, en 2003 31.5% et en 2005 49.3%. Il est donc évident que la collaboration avec un partenaire privé est beaucoup plus diffusée et commune. En considérant la taille de l'entreprise, il est intéressant de noter que la collaboration avec une autre entreprise est, en proportion, de plus en plus commune parmi les petites firmes. Quant aux raisons qui déterminent le besoin de collaborer, l'accès aux connaissances non disponibles à l'intérieur de l'organisation est la motivation la plus importante pour participer à des alliances. En 1999, le 96.0% des firmes qui avait au moins une alliance, collaborait pour cette raison. Cette part était égale à 91.6% en 2001, 65.7% en 2003 et 95.0% en 2005.

4.2. ANALYSE LOGIT

Les résultats de l'analyse logit révèlent que les facteurs qui ont une influence significative et positive sur la propension à collaborer sont :

- la taille de l'entreprise, mesurée à travers le nombre total d'employés
- l'orientation vers la biotechnologie, mesurée à travers le pourcentage d'employés ayant une responsabilité liée à la biotechnologie
- le degré de nouveauté des produits développés, mesurée à travers les variables binaires `dprod` et `dproc`, décrites dans l'annexe A
- la création par essaimage
- le nombre de droits de propriété intellectuelle obtenus par l'entreprise.

Les caractéristiques qui influencent la propension à collaborer avec une autre entreprise sont les mêmes, mais leur effet est sensiblement plus faible (négligeable dans le cas du degré de nouveauté des produits). Quant aux partenaires publics, la formation par essaimage a un impact très fort sur la propension à collaborer, surtout en 2005. Cette caractéristique, avec la taille de l'entreprise et le nombre de droits de propriété intellectuelle obtenus, est significative aussi pour déterminer la propension à collaborer pour obtenir accès aux connaissances externes (enquête de 2005 seulement).

4.3. RÉGRESSION BINOMIALE NÉGATIVE

La comparaison entre l'analyse logit et la régression binomiale négative montre que, bien que la formation par essaimage impacte de façon significative la propension à collaborer, elle n'a pas un effet sur l'intensité de collaboration, mesurée à travers le nombre total d'ententes de collaboration formées pas une entreprise. Le nombre de droits de propriété intellectuelle influence positivement l'intensité de collaboration ; néanmoins, il est surprenant de noter que le nombre de droits accordés à d'autres entreprises a une influence négative.

Nous remarquons aussi un effet positif de la nouveauté des produits développés sur l'intensité de collaboration avec une autre entreprise en 2005, alors qu'en 2003 les firmes qui développaient des processus de biotechnologie participaient à un nombre supérieur d'ententes de collaboration.

5. CONCLUSION

Parmi les caractéristiques que nous avions supposées comme influençant la propension et l'intensité de la collaboration, parmi les entreprises de biotechnologie au Canada, seulement la taille, l'orientation vers la biotechnologie, la formation par essaimage et le nombre de droits de propriété intellectuelle obtenus ont un effet significatif. En utilisant différentes variables dépendantes pour les modèles de régression, nous avons trouvé que l'effet de ces caractéristiques varie en fonction du type de raison à l'origine de la décision de participer à des ententes de collaboration et en fonction du type de partenaire. En les intégrant avec d'autres recherches, il sera possible de déterminer quelles sont les po-

litiques publiques les plus efficaces pour soutenir, favoriser et améliorer la productivité de l'industrie de biotechnologie au Canada.

TABLE OF CONTENTS

DEDICATION iv
ACKNOWLEDGMENTS v
RÉSUMÉ vi
ABSTRACT viii
CONDENSÉ EN FRANÇAIS x
TABLE OF CONTENTS xix
LIST OF FIGURES xxxi
LIST OF ABBREVIATIONS xxxii
LIST OF APPENDICES xxxiii

INTRODUCTION 1

CHAPTER 1 : LITERATURE REVIEW 4
1.1 THE NEED TO COLLABORATE 7
1.1.1 THEORETICAL FRAMEWORK 8
1.1.2 MONEY 12
1.1.3 HUMAN CAPITAL 14
1.1.4 KNOWLEDGE AND SKILL 17
1.2 CHARACTERISTICS INFLUENCING COLLABORATIVE BEHAVIOR 27
1.2.1 SECTOR-SPECIFIC CHARACTERISTICS 28
1.2.2 FIRM-SPECIFIC CHARACTERISTICS 29
1.3 SUMMARY 36

CHAPTER 2 : METHODOLOGY 37
2.1 DATA 38
2.2 SUMMARY STATISTICS AND CONTINGENCY TABLES 42

2.3 REGRESSION MODELS 45
2.3.1 DEPENDENT VARIABLES 48
2.3.2 INDEPENDENT VARIABLES 50

CHAPTER 3 : RESULTS - SUMMARY STATISTICS 55
3.1 GENERALITIES 55
3.2 PARTNERS 61
3.3 REASONS TO COLLABORATE 64
3.4 SUMMARY 71

CHAPTER 4 : RESULTS - REGRESSION MODELS 72
4.1 LOGIT ANALYSIS 72
4.1.1 OVERALL PROPENSITY TO COLLABORATE 74
4.1.2 MODELS 79
4.1.3 COLLABORATION AND PARTNERS 85
4.1.4 TYPE OF COLLABORATION: KNOWLEDGE AND PRODUCTION/COMMERCIALIZATION 89
4.2 NEGATIVE BINOMIAL ANALYSIS 92
4.2.1 OVERALL INTENSITY OF COLLABORATION 92
4.2.2 INTENSITY OF COLLABORATION AND PARTNERS 95
4.2.3 INTENSITY OF COLLABORATION AND REASONS 99
4.3 SUMMARY 103

CONCLUSIONS AND FURTHER RESEARCH 104

REFERENCES 109

APPENDICES 126

LIST OF TABLES

Table 1.1 : Reasons to collaborate and main risks 26
Table 2.1 : Sampling weights variables. 41
Table 2.2 : Contingency table for variable ec : total number of partnerships, 2005..... 44
Table 2.3 : Dependent variables for logit analysis, 1999 to 2005 surveys..... 49
Table 2.4 : Dependent variables used in Negative Binomial analysis, 1999 to 2005 surveys. 49
Table 3.1 : Changes in biotechnology firms in Canada by size, 1999 to 2005..... 55
Table 3.2 : Changes in collaborative biotechnology firms in Canada by size, 1999 to 2005 and by-size percentages of collaborative firms..... 57
Table 3.3 : Proportion of small collaborative firms by province, 1999 to 2005..... 58
Table 3.4 : Change in the number of alliances by firm size, 1999 to 2005..... 59
Table 3.5 : Collaborative firms by type of partner, 1999 to 2005..... 62
Table 3.6 : Collaborative firms and proportions by size and type of partner, 1999 to 2005..... 64
Table 3.7 : Collaborative firms by reason, 1999 to 2005..... 65
Table 3.8 : Proportion of firms having at least one knowledge-related partnership, by size, 1999 to 2005. 67
Table 3.9 : Distribution of alliances by reason, 2001 to 2005. 69
Table 4.1 : Fourth logit model, dependent variable ec, 2003 and 2005..... 81
Table 4.2 : Seventh logit model, dependent variable ec, 2003 and 2005. 84
Table 4.3 : Fourth logit model, dependent variable ecepri, 2003 and 2005..... 87
Table 4.4 : Fourth logit model, dependent variable ecipub, 2003 and 2005..... 88
Table 4.5 : Fourth logit model, dependent variable ecc, 2003 and 2005..... 90
Table 4.6 : Fourth logit model, dependent variable ecpc, 2003 and 2005..... 91

Table 4.7 : Fourth negative binomial model, dependent variable nec, 2003 and 2005 94
Table 4.8 : Fourth negative binomial model, dependent variable necepri, 2003 and 2005. 97
Table 4.9 : Fourth negative binomial model, dependent variable necipub, 2003 and 2005. 98
Table 4.10 : Fourth negative binomial model, dependent variable necc, 2003 and 2005 101
Table 4.11 : Fourth negative binomial model, dependent variable necpc, 2003 and 2005. 102
Table A.1 : Collaborative behavior-related variables. 127
Table A.2 : Variables related to reasons leading to collaborative behavior 134
Table A.3 : Location variables and province codes. 135
Table A.4 : Location dummy variables. 136
Table A.5 : Size-related variables. 137
Table A.6 : Age-related variables. 137
Table A.7 : Variables related to the financial situation 138
Table A.8 : Variables related to firm type 138
Table A.9 : Variables related to biotechnology products 139
Table A.10 : Contracts-related variables 140
Table A.11 : Variables related to Intellectual Property 143
Table A.12 : Variables related to capital financing 147
Table A.13 : Variables related to import/export. 149
Table A.14 : Variables related to business strategy. 150
Table A.15 : Variables related to population by province. 151
Table A.16 : Biotechnology R&D personnel by province. 151
Table A.17 : Number of universities, by province. 152
Table A.18 : Biotechnology patents, by province. 153
Table C.1 : Total number of biotech firms by size and province, 1999 to 2005 164

Table C.2 : Total number of collaborative biotech firms by size and province, 1999 to 2005 164
Table C.3 : Percentage of biotechnology firms involved in at least one collaborative arrangement, by size and province, 1999 to 2005 165
Table C.4 : Number of collaborative arrangements, by firm size and province, 1999 to 2005 165
Table C.5 : Number of collaborative arrangements per collaborative firm by size and province, 1999-2005 166
Table C.6 : Contingency table for variable nec : total number of partnerships, 1999 166
Table C.7 : Contingency table for variable nec : total number of partnerships, 2001 167
Table C.8 : Contingency table for variable nec : total number of partnerships, 2003 167
Table C.9 : Contingency table for variable nec : total number of partnerships, 2005 167
Table C.10 : Total number of firms collaborating for knowledge-related reasons (variable ecc), by size and province, 1999. 168
Table C.11 : Total number of firms collaborating for knowledge-related reasons (variable ecc), by size and province, 2001. 168
Table C.12 : Total number of firms collaborating for knowledge-related reasons (variable ecc), by size and province, 2003. 168
Table C.13 : Total number of firms collaborating for knowledge-related reasons (variable ecc), by size and province, 2005. 169
Table C.14 : Total number of knowledge-related alliances (variable necc), by size and province, 2001. 169
Table C.15 : Total number of knowledge-related alliances (variable necc), by size and province, 2003 survey. 169

Table C.16 : Total number of knowledge-related alliances (variable necc), by size and province, 2005. 170
Table C.17 : Total number of firms collaborating for production/commercialization-related reasons (variable ecpc), by size and province, 1999. 170
Table C.18 : Total number of firms collaborating for production/commercialization-related reasons (variable ecpc), by size and province, 2001. 170
Table C.19 : Total number of firms collaborating for production/commercialization-related reasons (variable ecpc), by size and province, 2003. 171
Table C.20 : Total number of firms collaborating for production/commercialization-related reasons (variable ecpc), by size and province, 2005. 171
Table C.21 : Total number of production/commercialization-related alliances (variable necpc), by size and province, 2001. 171
Table C.22 : Total number of production/commercialization-related alliances (variable necpc), by size and province, 2003. 172
Table C.23 : Total number of production/commercialization-related alliances (variable necpc), by size and province, 2005. 172
Table C.24 : Total number of firms collaborating with another biotechnology firm (variable eceb), by size and province, 2001 172
Table C.25 : Total number of firms collaborating with another biotechnology firm (variable eceb), by size and province, 2003 173
Table C.26 : Total number of firms collaborating with another biotechnology firm (variable eceb), by size and province, 2005 173
Table C.27 : Total number of firms collaborating with another non-biotechnology firm (variable eceab), by size and province, 2001. 173

Table C.28 : Total number of firms collaborating with another non-biotechnology firm (variable eceab), by size and province, 2003. 174
Table C.29 : Total number of firms collaborating with a pharmaceutical firm (variable ecep), by size and province, 2005 174
Table C.30 : Total number of firms collaborating with another non-biotechnology and non/pharmaceutical firm (variable ecabp), by size and province, 2005 174
Table C.31 : Total number of firms collaborating with another firm (variable ecepri), by size and province, 1999. 175
Table C.32 : Total number of firms collaborating with another firm (variable ecepri), by size and province, 2001. 175
Table C.33 : Total number of firms collaborating with another firm (variable ecepri), by size and province, 2003. 175
Table C.34 : Total number of firms collaborating with another firm (variable ecepri), by size and province, 2005. 176
Table C.35 : Total number of firms collaborating with a hospital/university (variable ecuh), by size and province, 1999 176
Table C.36 : Total number of firms collaborating with a hospital/university (variable ecuh), by size and province, 2001 176
Table C.37 : Total number of firms collaborating with a hospital/university (variable ecuh), by size and province, 2003 177
Table C.38 : Total number of firms collaborating with a hospital/university (variable ecuh), by size and province, 2005 177
Table C.39 : Total number of firms collaborating with a government laboratory (variable eclg), by size and province, 1999 177
Table C.40 : Total number of firms collaborating with a government laboratory (variable eclg), by size and province, 2001 178
Table C.41 : Total number of firms collaborating with a government laboratory (variable eclg), by size and province, 2003 178

Table C.42 : Total number of firms collaborating with a government laboratory (variable eclg), by size and province, 2005 178
Table C.43 : Total number of firms collaborating with a public institution (variable ecipub), by size and province, 1999. 179
Table C.44 : Total number of firms collaborating with a public institution (variable ecipub), by size and province, 2001. 179
Table C.45 : Total number of firms collaborating with a public institution (variable ecipub), by size and province, 2003. 179
Table C.46 : Total number of firms collaborating with a public institution (variable ecipub), by size and province, 2005. 180
Table C.47 : Contingency table for variable neceb : total number of partnerships with another biotech firm, 2001. 180
Table C.48 : Contingency table for variable neceb : partnerships with another biotech firm, 2003. 180
Table C.49 : Contingency table for variable neceb : partnerships with another biotech firm, 2005. 181
Table C.50 : Contingency table for variable eccc : collaboration to gain access to external knowledge/skill, 1999. 181
Table C.51 : Contingency table for variable eccc : collaboration to gain access to external knowledge/skill, 2005. 181
Table C.52 : Contingency table for variable eccnd : collaboration to gain access to external knowledge, 2003. 182
Table C.53 : Number of firms collaborating for knowledge-related reasons, 1999. 182
Table C.54 : Contingency table for variable ecc : knowledge-related collaboration, 1999. 182
Table C.55 : Number of firms collaborating for knowledge-related reasons, 2001. 183
Table C.56 : Contingency table for variable ecc : knowledge-related collaboration, 2001 183
Table C.57 : Number of firms collaborating for knowledge-related reasons, 2003. 183

Table C.58 : Contingency table for variable ecc : knowledge-related collaboration, 2003 184
Table C.59 : Total number of firms collaborating for knowledge-related reasons, 2005 184
Table C.60 : Contingency table for variable ecc : knowledge-related collaboration, 2005 184
Table C.61 : Number of firms collaborating for production/commercialization-related reasons, 1999 185
Table C.62 : Contingency table for variable ecpc : production/commercialization-related collaboration, 1999 185
Table C.63 : Number of firms collaborating for production/commercialization-related reasons, 2001 185
Table C.64 : Contingency table for variable ecpc : production/commercialization-related collaboration, 2001 186
Table C.65 : Number of firms collaborating for production/commercialization-related reasons, 2003 186
Table C.66 : Contingency table for variable ecpc : production/commercialization-related collaboration, 2003 186
Table C.67 : Number of firms collaborating for production/commercialization-related reasons, 2005 187
Table C.68 : Contingency table for variable ecpc : production/commercialization-related collaboration, 2005 187
Table C.69 : Total number of knowledge-related collaborative arrangements by province and firm size, 2001 187
Table C.70 : Total number of knowledge-related collaborative arrangements by province and firm size, 2003 188
Table C.71 : Total number of knowledge-related collaborative arrangements by province and firm size, 2005 188

Table C.72 : Total number of manufacturing/commercialization-related collaborative arrangements by province and firm size, 2001. 188
Table C.73 : Total number of manufacturing/commercialization-related collaborative arrangements by province and firm size, 2003. 189
Table C.74 : Total number of manufacturing/commercialization-related collaborative arrangements by province and firm size, 2005. 189
Table D.1 : First logit model, small firms, dependent variable ec, 1999 to 2005..... 191
Table D.2 : First logit model, small and medium firms, dependent variable ec, 1999 to 2005..... 192
Table D.3 : Second logit model, small firms, dependent variable ec, 1999 to 2005. ... 193
Table D.4 : Second logit model, small and medium firms, dependent variable ec, 1999 to 2005..... 194
Table D.5 : Third logit model, small firms, dependent variable ec, 1999 to 2005. 195
Table D.6 : Third logit model, small and medium firms, dependent variable ec, 1999 to 2005..... 196
Table D.7 : Fourth logit model, small firms, dependent variable ec, 2001 to 2005. 197
Table D.8 : Fourth logit model, small and medium firms, dependent variable ec, 2001 to 2005..... 198
Table D.9 : Fifth logit model, small firms, dependent variable ec, 1999 to 2005. 199
Table D.10 : Fifth logit model, small and medium firms, dependent variable ec, 1999 to 2005..... 200
Table D.11 : Sixth logit model, small firms, dependent variable ec, 1999 to 2005. 201
Table D.12 : Sixth logit model, small and medium firms, dependent variable ec, 1999 to 2005..... 202
Table D.13 : Seventh logit model, small firms, dependent variable ec, 1999 to 2005..... 203
Table D.14 : Seventh logit model, small and medium firms, dependent variable ec, 1999 to 2005..... 204

Table D.15 : First logit model, dependent variable ecepri, 2003 and 2005. 205
Table D.16 : Third logit model, dependent variable ecepri, 2003 and 2005.....206
Table D.17 : Fourth logit model, dependent variable ecepri, 2003 and 2005. 207
Table D.18 : Fifth logit model, dependent variable ecepri, 2003 and 2005.....208
Table D.19 : Sixth logit model, dependent variable ecepri, 2003 and 2005. 209
Table D.20 : Seventh logit model, dependent variable ecepri, 2003 and 2005.....210
Table D.21 : Fourth logit model, dependent variable ecipub, 2003 and 2005. 211
Table D.22 : Fourth logit model, dependent variable ecc, 2003 and 2005. 212
Table D.23 : Seventh logit model, dependent variable ecc, 2003 and 2005. 213
Table D.24 : Fourth logit model, dependent variable ecpc, 2003 and 2005.....214
Table D.25 : Seventh logit model, dependent variable ecpc, 2003 and 2005.....215
Table E.1 : First negative binomial model, dependent variable nec, 2003 and 2005.....217
Table E.2 : Third negative binomial model, dependent variable nec, 2003 and 2005.....218
Table E.3 : Fourth negative binomial model, dependent variable nec, 2003 and 2005.....219
Table E.4 : Fifth negative binomial model, dependent variable nec, 2003 and 2005.....220
Table E.5 : Seventh negative binomial model, dependent variable nec, 2003 and 2005.....221
Table E.6 : Fourth negative binomial model, dependent variable necepri, 2003 and 2005.....222
Table E.7 : Seventh negative binomial model, dependent variable necepri, 2003 and 2005.....223
Table E.8 : Fourth negative binomial model, dependent variable necipub, 2003 and 2005.....224

Table E.9 : Seventh negative binomial model, dependent variable necipub, 2003 and 2005. 225
Table E.10 : Fourth negative binomial model, dependent variable necc 2003 and 2005. 226
Table E.11 : Seventh negative binomial model, dependent variable necc 2003 and 2005. 227
Table E.12 : Fourth negative binomial model, dependent variable necpc 2003 and 2005. 228
Table E.13 : Seventh negative binomial model, dependent variable necpc 2003 and 2005. 229
Table F.1 : Correlation coefficients, 2005 survey. 231
Table F.2 : Correlation coefficients, 2003 survey. 234
Table F.3 : Correlation coefficients, 2001 survey. 237
Table F.4 : Correlation coefficients, 1999 survey. 240

LIST OF FIGURES

Figure G.1 : Likelihood function and its log, with respect to the parameter σ^2.....260
Figure G.2 : Effect of the intercept on the BRM curve.....275
Figure G.3 : Effect of the slope on the BRM curve 275
Figure G.4 : Effect of μ on the conditional probability distribution 293

LIST OF ABBREVIATIONS

ANOVA : ANalysis Of the VAriance

BRM : Binary Response Model

BUDS : Biotechnology Use and Development Survey

DBF : Dedicated Biotechnology Firm

GLM : Generalized Linear Model

GMM : Generalized Method of Moments

IP : Intellectual Property

IPO : Initial Public Offering

LM : Lagrange Multipliers

LPM : Linear Probability Model

LR : Likelihood Ratio

ML : Maximum Likelihood

MLE : Maximum Likelihood Estimation

NBR : Negative Binomial Regression

NSI : National System of Innovation

OLS : Ordinary Least Squares

p.d.f : probability density function

PRM : Poisson Regression Model

R&D : Research and Development

ROI : Return On the Investment

SME : Small and Medium Enterprises

SSI : Sectoral System of Innovation

LIST OF APPENDICES

ANNEX A : VARIABLES 126

ANNEX B : REGRESSION MODELS 154

ANNEX C : SUMMARY STATISTICS TABLES 163

ANNEX D : RESULTS – LOGIT MODELS 190

ANNEX E : RESULTS – NEGATIVE BINOMIAL MODELS 216

ANNEX F : COEFFICIENTS OF CORRELATION 230

ANNEX G : ECONOMETRIC METHODS 243

INTRODUCTION

Collaborative behavior is now recognized as an essential element of the business model of the biotechnology industry. The development, production and commercialization of a new product require a set of skills and competences extremely diversified, which the same agent of the biotechnology innovation system can hardly hold. For this reason, firms need to specialize and focus on a precise core competence, in order to efficiently and effectively add value to a specific stage of the development of a biotechnology product. New knowledge and technology at the embryonic stage is usually developed in universities, hospitals and other public institutions; then it is transferred to the industry, and particularly to small biotechnology-oriented firms, whose core competence is research and development. Nevertheless, small firms do not have the necessary resources to manufacture, test and commercialize a new product, and they need to collaborate with other agents of the biotechnology industry. In the case of human health biotechnology (which accounts for the majority of this industry), large pharmaceutical companies collaborate with small R&D oriented firms in order to appropriate new potentially profitable technologies, to perform clinical tests and, finally, to produce and commercialize them. Thus, according to this framework, collaborative arrangements are a necessary means for successfulness in the biotechnology industry, rather than a strategy among others. Corroborating this view, Oliver (2004) makes a strong point, associating the absence of partnership to a firm's death.

These considerations, along with the drop in the proportion of collaborative firms in the last years, claim for a complex and articulate research project whose final aim is providing an analysis and an evaluation of the public policies supporting the biotechnology innovation system through encouraging the formation of partnerships. This work represents the first step of this project, and aims at identifying firm-specific characteristics that influence the collaborative behavior. We use data collected in the « biotechnology use and development surveys » (1999, 2001, 2003 and 2005) to answer the following

questions: is there a difference in the propensity and in the intensity of collaboration between small, medium and large biotechnology firms? Is there a difference between firms collaborating for knowledge-related reasons and those collaborating for production or commercialization? Is there a difference between firms collaborating with another firm and those collaborating with a public institution? What are the distinguishing characteristics?

The first chapter is a review of the existing literature on collaborative arrangements in the biotechnology industry. It considers the three main needs that lead a firm to involve in a partnership with another agent: access to money, to knowledge and to human capital. We also present the "chain of alliances", the established pattern for innovation development in this industry. Finally, we focus on the firm-specific characteristics that we suppose to have an impact on the collaborative behavior, and we formulate the hypothesis.

In the second chapter, we provide an en explanation of the methodology we followed in analyzing the "biotechnology use and development surveys". Data issues, summary statistics, logit and negative binomial regression are introduced, and the variables employed are discussed in detail.

The third chapter presents the results of the analysis. In the first section, we provide the summary statistics to describe collaborative behavior within the Canadian biotechnology industry, distinguishing firms by their size and, where particular interest exists, by their location. Secondly, the results of the logistic analysis are presented, showing evidence on what are the firm-specific characteristics that influence the propensity to collaborate. We also distinguish between alliances related to knowledge and alliances related to production and commercialization. In addition, partnerships formed with another firm or with a public institution are considered. Finally, the third section presents the results of the negative binomial regression, and the differences between the firms are analyzed with respect to the number of alliances in which they are involved.

In the annexes, the reader will find additional information on the variables, models, and results. In particular, Annex G provides necessary theoretical basis to tackle the analysis of a complex survey using models for binary outcomes and for count variables.

CHAPTER 1

LITERATURE REVIEW

This work aims at describing the characteristics of collaborative versus non-collaborative firms, applying the econometric models presented in the third chapter. Firm characteristics are supposed to influence the propensity to collaborate and the number of collaborative arrangements in which an enterprise is involved. As the absence of alliances has proved to be a determinant of a biotech firm's death (Oliver, 2004), it is particularly relevant to deepen our understanding of this topic. Further work will be required to investigate the cause-effect relationships between firm-specific characteristics and the lack of collaborative arrangements. This way, it will be possible to determine which firms are at high risk, and to provide means for preventing their exit. Moreover, determining the characteristics of collaborative firms by type of alliance and partner, we will provide the basis for further research aimed at reducing the variability introduced by the risk, which is intrinsic to alliances and takes various forms, as will be discussed in the following.

Ernst&Young's last biotech report (2007) reveals that the biotechnology sector[1] is witnessing the beginning of a new stage of its evolution. Patterns of new knowledge creation are now more established and understood, the industry is more profitable, and collaboration between small biotech firms and large biopharmaceutical companies is now a well-acknowledged win-win strategy to turn new ideas into new successful products. In fact, during the first "exploratory" stage of the industry, the potential of biotechnology was not clear and the dynamics of the new emerging industry was not established. Due to the specific characteristics of (1) knowledge creation, (2) protection of the intellectual

[1] While most of the existing literature refers to biotechnology as an industry, we rather suggest employing the term "sector". In fact, biotechnology is a set of cross-economic activities, and it would be more appropriate to refer to it collectively as a sector.

property and (3) commercialization of new products in this sector, biotechnology firms were struggling to figure out what the appropriate business model to market biotech products could be. It was the "disruption stage" (Christensen, 1997) of the emerging technology, the typical stage of any emerging industry (Giarratana, 2004). It was the phase in which the agents of the new industry had to find out the sources of competitive advantage, and learn how to apply biotechnologies to living targets in order to develop and commercialize new products and have a return on the initial investments. In fact, a disruptive technology opens up radically new perspectives for the incumbent industry, but often requires new organizational practices and industrialization patterns. Dramatically improved efficiency and effectiveness, reduced costs, higher potential benefits and better end-user products are available on the market once the agents of the emerging industry learn how to do this and get familiar with the new techniques. On the other hand, an emerging and disrupting technology represents a threat for incumbents. Large, established firms that commercialize high-volume products whose characteristics are well understood and standardized and whose markets are well defined, are at one extreme of the continuum of innovating firms (Abernathy and Utterback, 1978). Their innovations[2] are rather incremental, aimed at satisfying customers' existing, well-known and understood needs. Incumbents often tend to oppose and resist to an emerging disruptive technology in order to avoid losing current investments; in fact switching to a new technology requires an important and costly effort in terms of acquiring new competences, adopting new internal procedures and organization, new patterns for appropriating, exploiting and turning new knowledge into marketable products. At the opposite, small, high technology and R&D oriented firms are more flexible, nimble and dynamic; these characteristics make them able to identify new needs or new ways to meet existing ones. In other words, within an innovative context, a radical, disruptive technology is more likely to come from a small, science-oriented enterprise. This is not to say that large es-

[2] The Oslo Manual (OECD, 2005) provides extensive details on the various types of innovation. The reader should revise this manual in order to understand the fundamental definitions employed throughout this work. It is also strongly recommended that the reader consult the Frascati Manual (OECD, 2002), which deals with the notion of R&D and the techniques to measure input and output of R&D activities. This work implicitly employs the definitions and procedures described in this manual.

tablished firms are not able to introduce breakthroughs, and the literature indicates the major role that "absorptive capacity" (Nicholls-Nixon, 1993) plays to foster innovativeness and the ability to learn from the external environment. Rather, the lack of "willingness to cannibalize" (Chandy and Tellis, 1998) existing investments has been proved to be a major factor leading to a much lower level of innovativeness of large firms, and a bad strategy in the long run for market-oriented incumbents.

Applying these concepts to the case of pharmaceutical biotechnologies, we note how the drug development process has been dramatically improved, from a "trial and error" pattern before the advent of modern biotechnology, to a more effective and "targeted process". Therefore, biotechnologies introduced a dramatically different and more effective way to produce a drug, requiring a set of competences that large established biopharmaceutical firms did not have. In this perspective, it is not surprising that small firms, often start-ups, introduced this set of new abilities in the industrial context and shed light on the beneficial effect these skills could have to the drug industry. These benefits and the way to achieve them are now clearer than in the past, the procedures and innovation patterns are now better understood. As stressed by the 2007 Ernst&Young biotech report, there are signs that the biotechnology industry is entering a new, more mature stage, in which the resistance from old incumbent firms has mostly disappeared, and a deeper understanding of the biotech innovation model has made larger investments by large firm possible with a lower level of risk. Furthermore, a net of relationships with small biotechnology oriented firms is a necessary condition for a big biopharmaceutical company to remain competitive in the drug industry, as (Oliver, 2003) shows. In other words, it is acknowledged that the opportunities from biotechnologies are immense and, at present, not only affordable and achievable, but represent indeed a better way to approach the research for the industrialization of new drugs and of many other types of products.

Although companies and Dedicated Biotechnology Firms (DBF's) operating in the Human Health sector account for more than a half of the whole biotechnology industry[3], they are not the only source of innovations. Evidence exists proving that a large amount of biotechnology products and processes have emanated from research projects performed by universities and public laboratories (Edwards et al., 2003), which are often closely linked to enterprises. As we will see below, collaboration between DBF's and public institutions is aimed at increasing the efficiency and the effectiveness of the innovation process in biotechnology, as they foster the transfer of new fundamental knowledge from universities, laboratories etc. to firms, providing them with new ideas for applied research projects. Public research is in fact more fundamental and is not aimed at commercializing an innovation; it requires thus further research to be turned into a marketable product.

1.1 THE NEED TO COLLABORATE

Alliances, partnerships and any form of collaboration between high-technology firms are acknowledged to be one of the determinants of success. In the biotechnology industry in particular, alliances play a central role in the innovation process (Barley et al., 1992), are a necessary condition for growth (Baum et al; 2000; Mytelka, 1999; Niosi, 2003) and can thus be considered as an essential element of the business model specific to this industry. Moreover, statistical evidence has proved that collaboration is indispensable for a firm to survive (Oliver, 2004). The need to collaborate comes from a variety of motives, and a variety of theories exists trying to explain why and how firms collaborate[4]; nevertheless, none of them provides an exhaustive and complete framework to understand

[3] Human health biotechnology accounts for the 54% of the whole biotechnology industry in terms of number of firms, and for more than 50% in terms of biotechnology revenues. Of the 532 Canadian biotech firms in 2005, 397 (74.6%) were small-sized, 83 (15.6%) medium-sized and just 52 (9.8%) were large (Lonmo and McNiven, 2007); yet large firms generated 2,465 millions of dollars from biotech activities in 2003 which represents the 64.2% of the total biotech revenues. (StatisticsCanada, 2005)

[4] For an introduction to the different theories, see Child and Faulkner (1998).

competitive alliances. In this work, we will not follow one particular theory: elements and concepts from different frameworks are used to shed light on the empirical findings. From a resource-based perspective, a firm may need to have access to (1) financial resources, (2) human capital and (3) new knowledge and skills (Aiken and Hage, 1968), depending on its characteristics and core competences. Although these three broad classes of motives can provide a useful and coherent understanding of collaboration among biotech firms, other dimensions are to be taken into account. Firms get involved in partnerships in order to reduce the risk intrinsically embedded in innovation, to gain access to distribution channels, to cut costs, for regulatory affairs, to improve productivity, to achieve economies of scale and/or scope, to block competitors or to gain competitive advantage on them. These new dimensions partially overlap with the three preceding, and do not constitute a conceptual framework; rather, they allow a deeper investigation, setting collaboration in a broader economic model aimed at conceptualizing the strategic behavior of a firm. In fact, collaboration is an element of a much broader system, composed of a variety of agents carrying on market and non-market activities aimed at creating, developing and commercializing new knowledge. In this respect, some authors (Cowan and Jonard, 2003; Powell and Brantley, 1992) recognize that the dyadic forces of competition and cooperation (in general, external relationships) among the agents within an innovative context are the necessary premise for creating new knowledge.

1.1.1 THEORETICAL FRAMEWORK

In 1952, Galbraith realized that innovations could be achieved only by undertaking costly activities requiring large resources that are available only to large companies: "There is no more pleasant fiction than that technical change is the product of the matchless ingenuity of the small man forced by competition to employ his wits to better his neighbor" (Galbraith, 1952, pp. 91-92). In other words, the patterns of technological change shifted towards ones that are more complex and the appearance of the competitive environment had altered, reshaping the structure of the investments, the competitive

strategies and all the economic activities in which a high-technology firm is involved. In fact, a leading high-tech firm cannot be isolated from the external environment, as innovation requires a broad system in which a variety of actors of different nature (private companies, public institutions, laboratories etc.) act in order to create, transfer and commercialize new knowledge.

Such a system, when focused to a specific country, is often referred to as National System of Innovation (NSI). Scientific literature on NSI's is wide, and provides a key level of analysis for scholars interested in the economics of innovation; the unit of analysis of the present work is the biotechnology firm, whose characteristics intervene in effecting its behavior and relationships with other agents in its system. In the present work, this system is identified as a fraction of the Canadian NSI[5], whose definition is intrinsically imprecise and whose boundaries are naturally blurry. The Unit of Observation of the econometric analysis is the single Canadian firm involved in activities requiring the use of biotechnologies. These firms can collaborate and have links with other agents that do not necessarily belong to the same NSI (e.g. collaboration with US biotechnology firms, agreements with European hospitals and so on) or to the same Sectoral System of Innovation (collaboration or links with a non-biotechnology firms). By Sectoral System of Innovation (SSI) we imply a "[...] set of new and established products for specific uses and the set of agents carrying market and non-market interactions for the creation, production and sale of these products" (Malerba, 2002, p. 248). The adoption of this unit of analysis is justified by the differences in the characteristics (knowledge base, demand, production processes, etc.) across industrial sectors. In other words, the industry-specific characteristics of a given industrial sector drive the dynamics of innovation in that sector. For example, there is a strong evidence that the knowledge base of a given sector has an influence on its rate of innovativeness (Breschi et al., 2000) and shapes the patterns of economic activities. This fact confirms the effectiveness of taking the SSI as an aggregate unit of observation.

---

[5] The econometric analysis is based on the sample of the Canadian Biotechnology Use and Development Surveys (1999, 2001, 2003 and 2005)

Among the different theoretical traditions dealing with the notion of system of innovation, the evolutionary theory is nowadays one of the most popular and interesting. It is a behavioral approach to the firm, stating the existence of two main processes leading to technological change and, therefore, to economic growth: selection and variation. As opposed to neoclassical models of economic growth, described in terms of maximization criteria where the behavior of the agents is deduced, in an evolutionary perspective, firms are at the center of the analysis, and their behavior is taken as given (Nelson and Winter, 1974). According to this framework, learning assumes a central role for economic growth, and is closely related to links among the agents of the system. Powell et al. (1996) suggest that networks of learning, rather than the individual firm, are the actual locus of innovation, which is coherent with what has been said above, observing that an isolated firm cannot induce a major technological change, and that innovation comes from a complex system of interacting actors. This obviously implies a non-linearity in the innovation process: technological change and economic growth arise from a multidimensional system composed by a myriad of agents linked together, whose connections imply bi-directional knowledge (both codified and informal) flows (Cowan and Jonard, 2003).

Focusing on the biotechnology industry, a common pattern can be recognized in alliances and partnerships formation among firms and with public institutions. In particular, a vertical alliance chain pattern is identified in the human health biotechnology sector, as evidenced by Stuart et al. (2007). Alliances follow the steps that the invention, development and commercialization a biotechnology product requires in the drug industry. Universities and public labs are often the source of basic, fundamental knowledge. At this early stage, technology needs more development and large financial resources to be turned into a marketable product, which is beyond the scopes of public institutions. Small DBF's, as it has been said above, are on average more innovative, have the ability to develop more radical innovations and disruptive technologies, but in general they lack financial resources. Conversely, large biopharmaceutical firms have the resources (both financial and non-financial) to manufacture, test and commercialize new products, but

usually lack the ability to produce breakthroughs, as they are more market-oriented. A partnership between these two types of agents obviously produces a benefit for both, providing the large firm with a promising disruptive technology to assure an advantage on the competitors, and the DBF with the financial resources to perform R&D and have a return on the investment (Powell et al., 2005). In other words, a small firm gains access to market without losing its focus on R&D, and without the large investments required to acquire complementary assets (Pisano, 1997).

One could suggest that a large company should improve its ability to produce disruptive technologies in house, without turning to alliances with a small DBF; literature confirms that this is possible, and it has been shown how a firm's internal characteristics can foster its ability to introduce radical innovations. Commitment to innovative activities has indeed been proved to be the key to the success for small firms (Baldwin, 1995), but also for large biopharmaceutical companies. While there are several factors affecting the ability to produce radical innovations and to commercialize them (the way information flows within a firm (Moorman and Miner, 1997), the way in which a firm is organized (Olson et al., 1995), and the willingness to cannibalize existing investments (Chandy and Tellis, 1998) to name a few) firm size is still considered one of the most relevant (Cohen, 1995). In this respect, an important topic concerns the effect of the R&D commitment on the likeliness to produce breakthroughs. As Soete (1979) shows, expenditures for R&D activities grow more than proportionately with firm size. Therefore, despite the fact that large firms spend much more on R&D activities they introduce less radical innovations on the market. In fact, large firms are more market-driven and demand-oriented than small firms. Thus, we can suggest that the processes leading to the creation and commercialization of breakthrough innovations in a large company are, on average, less effective. In addition, size can have an effect on the other drivers of innovation: for example, organizational flexibility plays a crucial role in an innovative context, and a small-sized firm is more likely to be flexible in its activities. More flexible organizational characteristics of small firms facilitate their ability to develop radical in-

novations: small size allows a firm's culture to be nimble, to adapt to major changes and to be proactive in its behavior (Riolli-Saltzman and Luthans, 2001).

The next part of this review deals with the three reasons leading to collaboration: money, manpower and competences. Access to these resources is fundamental for dedicated-biotechnology firms (DBF's): the core (and often the only one) competence of small firms (the engine of innovation in the biotechnology industry), is knowledge creation, which is a skill that usually larger firms lack (they usually cannot reach the same level of innovativeness, as it has been shown above). By contrast, small firms usually do not have the necessary financial resources to carry on R&D projects and to commercialize innovations; conversely, large biopharmaceutical firms have the necessary capital and the competences to market a new product. A collaboration between small, R&D oriented biotechnology-firms and large biopharmaceutical companies is therefore a win-win strategy, a powerful means enabling performance enhancement of economic activities of both partners through exploiting economies of scale and scope, cutting transaction costs, focusing on core competences and reducing duplications.

1.1.2 MONEY

The question of financing biotechnology deserves a deeper attention as it represents an important reason leading to the decision to participate in a partnership. In this work, we focus particularly on small enterprises, which are the engine of innovation in biotechnology and usually need financial resources from an external source. For a small DBF, access to capital is the main concern (Niosi, 2003), and failing to reach this objective leads to the firm's death. As it has been anticipated above, developing a new product in the human health sector is a long and expensive process and a small firm just cannot afford to go it alone. According to Shan et al. (1994) the capital required to commercialize a new drug is estimated to be between 125 and 250 millions dollars, an amount of money that is hardly affordable by a small firm.

Access to capital and financial resources is thus a challenging task for inventive small and medium enterprises (SME's), especially at the beginning of their existence, and different ways can be followed to achieve this target. Different sources of capital can be employed, and for each source, a firm can use several instruments to reach its goal. Firm-specific characteristics, in particular R&D orientation, are proven to have a strong influence on financing patterns (Baldwin et al., 2002). For example, firms operating in a R&D-oriented and high-technology industry, which is intrinsically risky and faces a constant market uncertainty, must rely more on internal sources of financing (Hache, 2005) and less on debt. Applying this concept to our subject, we can consider that at the beginning of its operations, a biotechnology small firm does not have any financial resource, and must rely on external capital: venture capital, debt, angel investors, governmental capital and alliances. Nevertheless, due to the high level of uncertainty embedded in a new but embryonic technology, it is hard for a new small biotechnology firm to raise capital from external agents. Also, Robbins-Roth (2001), Barley et al. (1992), Stuart et al. (2007), Kim et al. (2007), show that alliances are important for small and medium firms in order to raise capital. Even though one of the main concerns of a small biotechnology-oriented firm is access to capital, and an alliance with a larger company could provide it, many other aspects of a partnership must be considered. In fact, an alliance between a small biotech firm and a larger biopharmaceutical company is multifaceted, and the flow of money is just one of the elements to be considered, and could be confused and hidden depending on the perspective adopted.[6] In the next sections, we will further examine these details describing the other main reasons leading to collaboration and how these shape the process of new knowledge creation and new product commercialization.

Let us now examine the entry of a new biotechnology firm, in order to deeper understand the different needs of this type of agent during the first stages of its operations.

[6] e.g. the evolutionary theory places learning at the center of the analysis, leaving a secondary role to money; the same happens considering complementarities in competences.

The case of a small team of scientists who decide to start a company to develop a new biotechnology product is realistic in the biotechnology industry, as the birth of a biotechnology firm often follows this pattern, particularly in a geographic area where other biotechnology firms operate (Prevezer, 1997). The very first source of financing for a start-up is, most likely, the founders' personal capital: they provide the financial resources to carry on the preliminary activities to start the R&D project. At this "concept stage", the intervention of a venture-capitalist is very rare due to the high risk and incertitude of the outcome of the R&D project, notwithstanding the rate of return on the investment in the biotechnology industry can reach the very high value of 25 % (Hache, 2005). After the preparatory phase, the required investments are usually too onerous to be undertaken by the founders, and other sources and instruments must therefore be used. Among others, business incubators are particularly useful sources not only of financial resource, but also of a variety of services which help the start-up to carry on its activity and become profitable (Joseph et al., 2005).

Once this preliminary stage is overcome and only if the new technology developed by the start-up is promising, then the company can rely on other financing sources. One possibility at this stage is venture capital, and in the case of biotechnology returns are high: 15 – 25 % in the U.S. (Hache, 2005). At this stage, even though access to capital is a necessary condition to avoid exit, the need of external competences, skills and manpower gets more and more important, and the collaboration with the other agents of the innovation system needs to be reshaped.

Sometimes the start-up fails in reaching its goals, or simply the outcomes of the R&D project are not satisfying. In this case the firm can be acquired by a larger company (which takes advantage of the competences of the scientists and increases its R&D orientation) or, in the worst case, can go through a bankruptcy.

1.1.3 HUMAN CAPITAL

So far, access to financial resources for R&D activities has been investigated, which is one of the two necessary (but not sufficient in themselves) conditions enabling the de-

velopment of innovations. Another necessary factor is skilled human capital availability. Creating new knowledge requires the work of researchers, scientists and engineers with a high degree of specialization, acquired through a high-level education. Therefore, investigating the availability of these types of workers provides insight on the propensity to innovate. Human capital and financial commitment to R&D are the premises of the ability to innovate; without these two factors, innovation cannot take place. In addition, the lack of highly skilled workers is one of the motives to involve in a partnership. When focusing on partnerships and relations among biotechnology firms, it is necessary to investigate geographical agglomeration and knowledge spillovers to understand the importance of the set of connections among scientists. In fact, within a cluster, the relationships between biotechnology firms' researchers and scientists working in universities or public labs are, in general, close, and the level of innovativeness is highly influenced by the structure of the web of formal connections among firms and, on another level, by the set of personal connections among scientists.

Just to give some examples, it is a common practice among professors to take a sabbatical in a biotech firm[7]; this behavior fosters and increases informal links between the agents of the innovation system. The same pattern can be acknowledged for doctoral and postdoctoral students who, working on different projects in their careers, widen their links with other researchers in their domain. Moreover, Wolfe and Gertler (2004), Niosi and Bas (2001) and Prevezer (1997) provide strong evidence that the presence in a given geographical area of star-scientists in biotechnology is a key determinant for the formation of a new cluster (Zucker et al., 1998). They show that the availability of outstanding competence and new knowledge creation ability is the key to gain a competitive advantage. Conversely, the lack of highly skilled scientists in a geographic area is an obstacle to developing biotechnology.

The fact that scientists in universities, in public laboratories and in the firms are highly interconnected within a cluster and form a tight web of relationships suggests a new perspective to explain the pattern of innovation, namely the collective dimension. In this re-

[7] Audretsch and Stephan (1996) provide an extensive analysis of company-scientist links.

spect, Cowan and Jonard (2003) have found evidence that the communication network within a cluster has a great positive effect on the level of innovativeness. It is to be noticed, however, that when a cluster is too closed upon itself, and the connections with the external environment are not strong enough, a lower level of innovativeness is observed: new fresh ideas from outside are important for new R&D projects. As operating in a technology-intensive industry, biotechnology firms tend to agglomerate spatially in small geographical locations (Audretsch, 2001), and the case of Canada, with its three main biotechnology clusters in Toronto, Montréal and Vancouver (Beaucage and Beaudry, 2006), provide further empirical evidence.

The higher rate of innovativeness of the firms located within a cluster can be explained by the presence of knowledge spillovers. Knowledge spillovers are defined as the way tacit knowledge flows between agents. Audretsch and Feldman (1996), Jaffe (1998), Jaffe et al. (1993), Beaudry and Berschi (2001), just to cite some, attribute a strong importance to knowledge spillovers as a key factor leading to a higher level of innovativeness; another stream of literature considers spillovers as unmeasurable or spatially unbounded (Krugman, 1991), and therefore not deserving a deeper attention. The perspective assumed in this work follows the former stream of literature: in a knowledge-based economy, the importance of non-codified knowledge exchanges through informal channels assumes a relevant role and defines indeed the knowledge spillovers. In fact, codified knowledge does not require geographical proximity to be transmitted, while non-codified knowledge does. It is worth noting that some authors directly contradict Krugman's opinion, providing a measure of knowledge spillovers through, for example, patent citations (Jaffe et al., 1993). A generally acknowledged result among the scientists who attribute a strong importance to the streams of non-codified knowledge among agents is that capturing knowledge spillovers can lead to a boost to the rate of innovativeness and eventually to a higher social benefit (Jaffe, 1998) which is proven to be influenced by the introduction of innovations (Trajtenberg, 1989).

In this scenario, collaboration among the agents of the system of innovation can be divided into two main categories: formal and informal[8]. A formal agreement aimed at gaining access to human capital in the biotechnology industry is usually stipulated between a firm and a university or public institution. A sabbatical year, a doctoral or post-doctoral project in a biotechnology firm is an example of this collaboration[9]; in this case, geographical proximity has been proven to foster the flow of knowledge between the agents (Audretsch and Feldman, 1996). The myriad of links between scientists belonging to different agents in the innovation system represents the informal dimension of collaboration, and the traces they leave under the form of patent citation can capture, as we said above, the knowledge spillovers (Jaffe et al., 1993). Finally, informal collaboration aimed at getting access to human capital does not show a typical pattern with respect to the agents involved, but it is rather related to the characteristics of the cluster and the environment.

In conclusion, a knowledge-intensive and innovative industrial cluster can be seen as an intricate web of overlapped interconnections of different nature among the different agents of the cluster at different levels. The set of all the informal agreements between the agents of a cluster forms a net which is overlapped to the net of formal collaboration among the scientists working within the same cluster, and belonging to different agents.

1.1.4 KNOWLEDGE AND SKILL

Within an innovation system, new knowledge stems from a variety of agents involved in R&D activities, and the importance of the concept of the "collective dimension" of innovation has already been stressed. New technologies are not generated by the individual firm (or, more generally, the individual agent within a system of innovation), but rather are the product of the whole system of interconnected entities (Cowan and Jonard,

[8] In this work, we will analyze formal collaborative arrangements only: informal link between scientists are not captured by the Biotechnology use and development surveys.

[9] Literature on informal collaboration is extensive and many authors performed analyses on this topic, especially with respect to knowledge flows. For example, Cockburn and Henderson (1998) provide further details on the forms of interactions between the public and the private sectors in the case of R&D in drug discovery.

2003). The agents perform different activities, which generate new ideas. This pattern allows each agent to specialize in the activities it is better at, and to have access to complementary assets through a web of relationships. Thus, the value chain is fragmented, and every agent is associated with a specific stage of development in this chain. While some years ago the discussion on whether to internalize complementary assets or to gain access to them through an alliance was still open (Pisano and Teece, 1989), it is nowadays clear, especially in the human-health biotechnology sector, that the vertical chain of alliances (Stuart et al., 2007) is a more efficient way to create, develop and commercialize a new technology. Given the length, the complexity and the large-capital-requiring nature of a new drug development process, it is hard, or even impossible, for a firm to undergo the whole process alone (Baum et al., 2000). A strong R&D orientation, the ability to produce radical innovations and the skills to commercialize a new product would be necessary. Through a partnership, a firm can gain access to complementary assets (Pisano, 1991) and concentrate on its key competences; this way the overall efficiency of the whole system of innovation is increased. In the following the typical alliance chain observed in the human health biotechnology sector is analyzed in greater detail. At the centre of the chain is the small biotechnology firm, which is involved in upstream partnerships aimed at having access to basic knowledge and in downstream partnerships, to gain access to capital and complementary assets to produce and commercialize a new product.

1.1.4.1 FIRM-UNIVERSITY PARTNERSHIPS

The role that universities have in the biotechnology industry is to provide new, fresh ideas for the developing of new technologies to be eventually commercialized. In fact, the research performed in universities and public institutions is basic and fundamental, requiring further investments to be developed and turned, in a marketable product (Colyvas et al., 2002). It has been suggested that biotechnology inventions arise from new knowledge held by a group of scientists who decide, recognizing the potential embedded in it, to involve in further research to commercially exploit new knowledge to

gain financial returns from it (Zucker et al., 1998). Empirical evidence confirms this fact, as about a half of small biotechnology firms are start-ups created by groups of university researchers (Stuart et al., 2007) that maintain close ties with universities (Audretsch and Stephan, 1996). These ties, as it has been mentioned, can be both formal and informal. Informal linkages can leave a trace in the form, for example, of coauthorships among researchers in public and private organization (Stuart and Ding, 2006). In this respect, collaboration aims both at getting access to human capital and to external knowledge and skills, held by university scientists. Formal linkages take often the form of R&D projects involving one or more university researchers during a certain period. This way, new fresh ideas from public institutions can flow to the industry, acting as a source of information for new research projects. In this respect, universities are indeed considered by firms as the most valuable source of inventions (Rosenberg and Nelson, 1994).

Formal licensing through the technology transfer offices has acquired in the last years a prominent importance in the biotechnology domain, but a number of issues must be addressed to make this process more efficient, effective and fast. Universities usually protect new knowledge through patents or copyrights, aimed at exclusive licensing to a firm. This way, new technologies are more attractive to companies, and can be more efficiently exploited. However, a major issue in technology transfer from universities to industry is the slowness of the process, often due to the difficulty to pinpoint the university's interests as evidenced by Colyvas et al. (2002). They observe how the role of Intellectual Property (IP) protection within public institutions and its effect on industrial R&D is not clear: a broader analysis must be performed to shed light on the process.

Contrarily to the notion that universities and, more generally, public research institutions, only generate new ideas for industrial R&D projects, Cohen et al. (2002) suggest that other sources of new knowledge are in some cases more relevant for high technology companies. In fact, empirical evidence shows that in the high technology industry, flows of knowledge are not linear, and the inputs for new radically new products and processes can stem from both upstream (supply-side) and downstream (demand-side)

sources. Basic science provides more supply-side ideas, for products that do not exist yet and are new to the market. New ideas do not come only from universities and public institutions; the 1999, 2003 and 2005 surveys of innovation[10] made by Statistics Canada provide interesting insights on the use of sources of information for innovative activities. While the results for the manufacturing sector are not directly comparable due to the slightly different questions asked, it is none the less possible to assert that they show the same pattern: the most important internal source of new ideas for innovative projects is management personnel, followed by production personnel, sales and marketing personnel and, finally, R&D personnel. The 2003 survey of innovation reveals the same pattern for service enterprises, with the exception, as it is to be expected, of production personnel, whose value is significantly lower. This is coherent with a customer-oriented behavior, confirmed and emphasized by the surveys' results on the factors leading to success: existing customers' satisfaction is always indicated as the first and most important factor of success. This means that inventive firms are more oriented to incremental innovations than to radical innovations. The latter are more risky, aim at creating new markets through creating new needs and require cannibalization of current allocated resources (Chandy and Tellis, 1998). On the other hand, incremental innovation is demand-side (Ryans and Shanklin, 1989), responds to current customers' changing needs, aims at maintaining a competitive advantage on the competitors. It is less fraught with short-run risk in its nature, but it can't respond to new, disruptive, radical innovation introduced in the market by other competitors. It is well acknowledged that an inventive firm must have a portfolio of products at different stages of the S-curve. That is to say, a firm should introduce both incremental innovations (improvements) for existing products and radical innovations. Incremental innovation is associated with a more mature product, with a declining rate of growth: these products are "cash-cows" (Henderson, 1998), generate large revenues (depending on their market share), and provide financial resources to fund R&D projects for radical innovation. The latter class of products is char-

[10] 1999 and 2005 surveys were addressed to the manufacturing sector, while 2003 survey was addressed to the service sector. For further detail on populations, samples and methodology, see www.statcan.ca.

acterized by a high level of uncertainty, requires large investments and does not generate revenues in the short run. At this stage, a diversified and balanced product portfolio is therefore the ideal situation for a successful inventive firm.

In conclusion, the process for the creation of a breakthrough innovation stemming from fresh new knowledge often follows the same steps. Basic research performed in universities and public institutions is transferred through a certain type of collaboration to a small dedicated-biotechnology firm, which provides further research to turn the new embryonic knowledge into a quasi-marketable product. Finally, a large biopharmaceutical company intervenes, providing the complementary assets to accomplish the commercialization of the new product. Thus the stream of knowledge starts in the universities, passes first through R&D-oriented firms, then through a large biopharmaceutical company and, finally, approaches and diffuses into the market.

In this respect, the innovation process for a breakthrough innovation in biotechnology is said to follow a linear pattern. According to a linear perspective of the innovation process, new ideas come mainly from upstream sources (upstream research, suppliers) and brand new products are introduced into the market. While it has been shown that the linear model does not apply in an innovative context (Cohen et al., 2002), in the case of biotechnology and pharmaceuticals, the impact of public research is substantial, providing industry with embryonic technologies requiring additional research. In this respect, the flow of knowledge in the biotechnology industry seems to be more linear compared to other high-technology sectors, where downstream sources, e.g. customers, are relevant sources of new ideas for new products. If public R&D is one of the most important sources of new projects for biotechnology companies, it also contributes significantly to project completion. In this second case, the partnerships between firms and universities does not take the form of an exclusive or semi-exclusive licence (as it was in the case of the Technology Transfer Office); rather, a more complex contract is generally required, and the main concern for the firm is not to let knowledge spill over.

Another point to be stressed relates to fundamental research performed by firms. It is well acknowledged, in fact, that basic research is not only performed by universities and

public institutions, but also by private companies, although it does not lead to immediate financial revenues. So why do firms fund and perform fundamental research? The benefits of basic research are higher for the society as a whole than for the single firm. Why do firms, regardless to the non-appropriability of basic knowledge and to the impossibility of deriving financial return from it on the short-run, perform basic research? Rosenberg (1990) provides a wide discussion on this topic, arguing that basic research is a long-run investment, an entry means to a specific information network, and represents a solid base for evaluating the characteristics of more applied knowledge. An alliance with a university aimed at improving the quality of fundamental research can take the form of a contract for a specific period of time or for the completion of a specific R&D project or, as Cohen et al. (1998) show, hybrid (university-industry) research centers can be created. What is remarkable, is that not only large biopharmaceutical firms perform basic research, but so do small biotechnology firms: in this field new technology is more readily appropriable than in other high technology sectors (Rosenberg, 1990). Another reason leading to basic research performance is that its output provides a first-mover advantage and can be considered as an access key to an information network. This motive often leads a large company to create links with universities (Arora and Gambaradella, 1990).

1.1.4.2 FIRM-FIRM PARTNERSHIPS

Inter-firm alliances[11] aimed at gaining access to external skill and knowledge are the most important form of collaboration in the biotech industry and are considered by many authors as the key to successfulness in developing and commercializing a new product. As it has been said, small firms are the engine of innovation in the biotechnology industry, they have the ability to perform research aimed at creating technologies that are new to the market; however, they lack the complementary assets (in terms of capital and

[11] In the literature, as well as in the survey questionnaires, no differences are introduced between alliances, partnerships and collaborative arrangements; therefore, we will use these three terms indifferently. In fact, the focus of this work is on collaboration as a strategic behavior, rather than on the forms that a collaborative arrangement between two agents of the system of innovation in biotechnology can take.

competences) necessary for the manufacture, test and commercialization of a new product. For a high-technology small firm, and particularly in the biotechnology field, it is essential to form alliances to overcome the lack of competences, capital and knowledge. Pisano (1991) outlines four categories of capabilities a small biotech firm needs in order to exploit its potential: manufacturing, clinical testing, regulatory processes and distribution.

An alliance helps a firm to have access those commercial and technical resources it would take several years to build in-house (Ahuja, 2000). Baum et al. (2000) show how alliance patterns at the funding affects future performance and pattern of alliances, and Oliver (2004) links the lack of partnership with a firm's death. Given the need to form alliances for a biotechnology firm to overcome its resource scarcity, vertical alliances allow each partner to focus on their core competences. Incumbents provide the upstream partner with capital to fund R&D projects and usually are involved with subsequent production, clinical test and commercialization of the new product. In fact, in the pharmaceutical industry, established companies have the expertise and experience in the product testing process, which is time-consuming and requires large investments. They also have competences in marketing and sales which are necessary for the last steps of the commercialization (Kim and Higgins, 2007).

In this respect, a small biotech firm, whose core competence is R&D, acts as a *technology broker* (Stuart et al., 2007), an intermediary between the institutions performing basic research and the large biopharmaceutical companies, usually unable to perform R&D aimed at generating a radically new technology. The resulting vertical structure is specific to the biotechnology industry compared to other high-technology domains (e.g., semiconductors), in which there is the tendency to integrate different competences (research and manufacturing) in order to avoid knowledge to spill over and to erode the competitive advantage. This consideration does not imply that in the biotechnology industry knowledge does not spill over among partners; it rather means that firms have to face this risk, since a partnership is simply necessary. Literature on this topic is wide and rapidly evolving, following the development of new industrial sectors, and shows how in

general relational risk has a multidimensional nature[12] (Delerue, 2004). Leak of information and tacit knowledge is perceived as the most relevant source of risk for a firm involved in an alliance (Kale et al., 2000) and goes together with the risk of conflict (Ariño and De La Torre, 1998) and loss of independence (Hagedoorn and Sadowski, 1999). In this respect, the biotechnology sector introduces a peculiarity: the founders of a biotech start-up can expect being acquired by a larger company (sometimes the partner itself), once the R&D project is complete and a new technology or product is ready to undergo the commercialization process.

From the point of view of the large biopharmaceutical company, alliances with R&D-oriented DBF's are essential to have access to the latest advancements and maintain an advantage on competitors. These linkages are facilitated if the partners' strategies and competences are complementary (Arora and Gambaradella, 1990), which justifies the dynamics of collaboration described above. R&D agreements, minor participations (investments in capital stock) and acquisitions are the linkages a large biopharmaceutical firm aims at forming with a DBF. Through the acquisition of a small biotech firm, a company can specialize in a certain area of research, and, in general, improve its R&D orientation. As the risk of failure in this field is quite high, large firms have often a diversified portfolio of agreements, each of which is usually product-specific. This way, a company can increase the probability to gain financial revenues and return on the initial investment. Arora and Gambaradella (1990) also suggest that a large firm should diversify the type of alliances it is involved in, investing both in long and short term projects, in order to minimize the effects of risk.

We have seen how the agents in a biotechnology innovation system have usually different but complementary competences, and how through forming a web of alliances, their activities can enhance the value of new knowledge, focusing on the activities they are

[12] Baum et al. (2000) provide a complete overview on the various risks that a small dedicated-biotechnology firm faces.

better at. We have also examined how the characteristics of a drug development process explain for the characteristics of the pattern of innovation in this industry, and for the vertical structure of the alliances. The prominent role of fundamental research as a source of new embryonic knowledge is a key determinant of the close links between universities and DBF's. The lack of complementary assets (capital, resources, manufacturing and distribution competences) and the relevance of the time-to-market (Gulati, 1998) for a new technology obliges a DBF to collaborate with a larger company in order to gain have access to those resources. In addition, a biopharmaceutical company must diversify its partnerships portfolio in order to minimize the effects of the risk, which is naturally embedded in a R&D project for a brand new technology.

Table 1.1 resumes the advantages a firm hopes gain from forming a partnership and the risk that the alliance introduces, distinguishing by agent and by partner. This way, we adopt a strategic point of view, where the decision to involve in a collaborative arrangement comes from the evaluation of the advantages with respect to the risks[13]. As it has been observed, universities collaborate with small biotech-dedicated firms in order to gain financial revenue and to build a set of useful connections, at the risk of losing work force, which could migrate from the public sector to the industry. Conversely, a DBF collaborates with a university in order to gain access to human capital, to new fresh knowledge and in order to build a web of connections, at the risk of spilling over internal knowledge[14]. Although this is perceived as a risk from the point of view of the single agent involved in the partnership, it is considered as an overall benefit for the industry as a whole (Mowery, 1998). In addition, transferring a new technology from a university or a public laboratory requires a complex process involving hard-to-solve appropriability

[13] Although the focus of this work is on collaborative behavior, it is useful to provide some details on this aspect in order to get a more complete view of the forces that play a role in the dynamics of collaboration. We do not purport to provide a complete analysis of the risks, as this would be beyond the scopes of this work. If interested, the reader should consult the literature cited in the text.

[14] It is worth noting that, in the case of alliances involving a university and a firm, the former agent usually aims at publishing the results in an academic journal, while the latter is usually interested in keeping the secret in order to maintain an advantage on the competitors. This could be the cause of conflictual relationships.

issues. Finally, a small firm faces the risk of acquiring a technology whose outcomes do not match the expectations.

Table 1.1 : Reasons to collaborate and main risks

Agent	Partner	Reasons	Risks
Universities	DBF's	Financial revenue Building a network	Losing star scientist
DBF's	Universities	Access to human capital Access to new knowledge Building a network	Knowledge leak Appropriability issues Acquiring non-promising technology
DBF's	Large biopharma	Access to capital Access to market Manufacturing Clinical testing Distribution Risk reduction	Knowledge leak Loss of independence Appropriability issues Conflict Opportunism
Large biopharma	DBF's	Access to new technology Acquiring familiarity with new technology	R&D project failure No R.O.I.

In the downstream side, a DBF usually collaborates with a large biopharmaceutical company to gain access to capital and to those complementary assets (marketing competences, manufacturing, regulatory affairs and clinical testing) that are necessary to the commercialization of the new technology. Also in this type of alliances, a DBF faces the risk of spilling over its knowledge and has to deal with complex appropriability issues. Many other factors, furthermore, could lead to a turbulent relationship marked by conflict and opportunism that could induce a loss of independence.

Through a partnership with a DBF, a large biopharmaceutical group gains access to new potentially disruptive knowledge, allowing it to keep a competitive edge among the

other agents in the industry. The risk in this case is related to the potential of the new knowledge, which could not match the initial expectations and generate no return on the investments. As the focus of this work is on small and medium dedicated-biotechnology firms, Table 1.1 does not report alliances between large companies and public institutions. Nevertheless, these two agents do form collaborative arrangements; especially for testing (a large company needs the expertise to conduct tests on living organisms) and to have access to fundamental knowledge.

1.2 CHARACTERISTICS INFLUENCING COLLABORATIVE BEHAVIOR

In the preceding section, the nature and the reasons leading to collaboration within the biotechnology industry have been introduced and the typical vertical pattern of alliances of this sector has been presented. We now focus on the objectives of this work: what are the characteristics of the firms deciding to involve in a collaborative arrangement with another agent of the biotechnology system of innovation? How do those characteristics influence the propensity to be involved in a partnership? Given the importance of collaboration and the central role of new biotechnology firms in the innovation process, it is necessary to understand what are the firm-specific characteristics having an effect on the propensity to collaborate, in order to identify those that have a detrimental impact on the collaborative behavior. In fact, although partnerships are vital for biotech firms to reach their goals, a significant portion of them do not collaborate with other agents, as the preliminary results of the Canadian Biotechnology Use and Development Surveys show (Statistics Canada, 2005). Tracing a picture of non-collaborative firms is the necessary premise to understand why they do not collaborate and to provide means to avoid this behavior. Further stressing the importance of collaboration among the agents belonging to the biotechnology system of innovation, Oliver (2003) shows that a DBF should engage in at least two collaborative arrangements. Conversely, a large biopharmaceutical company should set at least ten partnerships with so many DBF's in

company should set at least ten partnerships with so many DBF's in order to reduce the risk of negative ROI caused by R&D project failure.

Before examining the firm-specific characteristics affecting the propensity to collaborate, we briefly focus on the sector-specific characteristics in order to fix our understanding of the dynamics of collaboration in the biotechnology industry.

1.2.1 SECTOR-SPECIFIC CHARACTERISTICS

So far, the need of collaboration and the motives leading to the decision to form a partnership have been analyzed, taking the single agent as the unit of observation. From a different perspective, the biotechnology sector-specific characteristics leading to collaboration have been introduced, and the way the process of innovation shapes the pattern of alliances among different agents has been considered. The development of a new biotech product requires a multi-stage research effort, which is performed by universities at a more basic level and subsequently by DBF's. The outcome of public research-oriented institutions is usually a new, embryonic technology requiring additional development, investment and time to generate, eventually, a marketable product. Formal agreements are common for transferring new knowledge form universities to the industry, and can take the various forms described above. A small biotech firm's core competence is R&D performance, but a set of other resources are required. An alliance with a large, established biopharmaceutical company is usually the best way to overcome these difficulties. The downstream partner provides the R&D-oriented firm with capital and manages testing, manufacturing and distribution. The competences of these two types of agents are therefore complementary and, considered individually, insufficient to develop a new product through all the stages. This makes an alliance between incumbents and small biotech firms a win-win strategy, whose benefits are captured by both partners and the overall efficiency of the innovation process is improved.

In a high technology industry, as in the case of biotechnology, a higher level of collaboration is more likely to be observed (Dogson, 1994), due mainly to the degree of complexity of the innovation process (Malerba and Orsenigo, 1993). However, if competi-

tion is very intense, Von Hippel (1989) shows that the risk of knowledge leak retains firms from collaborating.

Appropriability conditions within a specific industry also have an influence on the overall propensity to collaborate: a higher degree of appropriability, as in the case of biotechnology, leads to a higher intensity of collaboration among firms, as a formal agreement is the only way to have access to external knowledge (Pyka, 2002). None the less, as we will see in the following, the use of formal methods of protection does not always have an effect on a firm's propensity to collaborate.

1.2.2 FIRM-SPECIFIC CHARACTERISTICS

We now focus on the characteristics of the firms involved in collaborative agreements, which represents the core of the present work. Existing literature sheds light on those characteristics that affect a firm's propensity to collaborate, but the results are sometimes in contradiction. The choice of the relevant characteristics (and the variables used to measure them), presented in chapter 2 is justified by the existing literature. In particular, it is recommended to consult Dachs et al. (2004) and Schmidt (2007).

1.2.2.1 BIOTECHNOLOGY PRODUCTS AND DEGREE OF NOVELTY

The measure of the novelty of an innovation is neither straightforward nor easy to assess, as the distinction between radical innovation and incremental innovation is not always clear. One should rather consider a continuum in the degree of innovativeness of a new product or process. None the less, it is acknowledged that a higher degree of innovativeness is associated with the introduction of a new product, while process innovation has usually less radical qualities. This perspective could apply to the case of the biotechnology industry: a new product has often a disruptive effect on the existing technologies and creates new markets (it is supply-side). For instance, the discovery of insulin is a valid example to understand how biotechnology can induce major changes in a certain market. Discovered and first isolated by Banting and Best in 1922 (Bliss, 1982), insulin to treat diabetes was extracted from bovine liver until recombinant DNA technology was

developed in 1978 by Genentech. This second stage of the history of insulin is a perfect example of the pattern of innovation in biotechnology: Eli Lilly was the leader in bovine insulin production, but in the 1970's recombinant DNA started a new battle for this market. Competition between Eli Lilly and Genentech, an R&D-oriented biotech firm, was intense and the latter, in August 1978, won the race. Nevertheless, Genentech lacked the resources and the skills to produce insuline, and entered into an agreement with Eli Lilly (Gans and Stern, 2002). However, we must be careful in assessing a higher level of innovativeness to new products rather than new processes, as exceptions may exist.
Results provided by Dachs et al. (2004) suggest that the degree of innovativeness positively affects collaboration when R&D is perceived as the core competence of the firm. In this perspective, diversification in innovative activities (research ranging from basic to applied, for example) predicts the likelihood to be involved in a partnership. These findings, however, are proved to vary across country, implying that other regional factors play a significant role in explaining collaborative behavior.

Hypothesis 1: a higher degree of novelty of the innovations a firm introduces in the market has a positive effect on the likelihood to be involved in one or more collaborative arrangements. The measure of the innovativeness of a new product or process is not straightforward. In Chapter 2, we discuss in detail the choice of the independent variables employed in the analysis.

1.2.2.2 SPIN-OFFS

Following the definition provided by Statistics Canada, a spin-off is "a new firm created to transfer and commercialize inventions and technology developed in universities, firms or laboratories", we can intuitively associate this type of firms with a higher propensity to collaborate. In fact, when a portion of a larger organization (firm or public institution) spins off and becomes independent, it usually maintains tight links with the originating agent in order to take advantage of common skills, competences and expertise. While we

hypothesize that a spin-off firm is more likely to collaborate, no evidence is provided on the effect that this characteristic may have on the intensity of collaboration.

Hypothesis 2: A firm generated through a spin-off is more likely to be involved in at least one collaborative arrangement. Nevertheless, we hypothesize that this characteristic does not influence the number of alliances in which a firm is involved.

1.2.2.3 PUBLIC SUPPORT

A positive impact on the propensity to collaborate is found by Abramovsky et al. (2005) and Negassi (2004) among those firms that receive public funding. Ambiguous results in this respect are provided by Dachs et al. (2004), who find a country dependence for this characteristic. This diversity could be explained considering that a certain country is more or less concerned with fostering university-industry cooperation and this leads to different forms of public funding programs and different requirements for a firm willing to receive public financial support. However, especially when a funding program does not require a firm to form a partnership with a public research institution, public support (especially fiscal incentives) can indirectly affect the propensity to collaborate, increasing the level of innovativeness (Hall and Van Reenen, 2000).

Hypothesis 3: The Canadian policy is intended to foster collaboration among the agents of its innovation systems ("Mobilizing science and technology to Canada's advantage", 2007). We want to investigate whether public funding has a positive effect on the likelihood of being involved in a partnership for firms in the sample of the Canadian Biotechnology Use and Development Surveys. This will provide useful result for comparison with future work aimed at evaluating the impact of the Canadian strategy to foster innovation.

1.2.2.4 SIZE

The way size influences the likelihood to cooperate has been shown to be positive for inventive firms (Miotti and Sachwald, 2003). This tendency has been confirmed by many authors, in particular for R&D cooperation (Cassiman and Veugelers, 2002). For the biotechnology industry, however, the industry-specific characteristics shape the likelihood to collaborate in a different and particular way: small firms do not have the resources necessary to carry on R&D projects, and do not have the competences to commercialize new products. For these reasons, alliances are vital for this class of biotech firms (Oliver, 2004), whose biggest concern is access to capital (Niosi, 2003). On the other hand, larger firms do not often have the characteristics to create new, disruptive products; however, disruptive technologies are the major threat to these firms, whose leadership position is constantly at risk. Therefore they need to be one step forward and to appropriate these technologies; one way to do so is through a partnership with smaller, more dynamic and radically inventive biotech firms, in order to have access to new, disruptive knowledge. Thus, size is probably not that relevant explaining the likelihood to participate in an alliance. One could argue that it has an effect on the number of collaborations in which a firm is involved. Let us consider the case of a biotechnology start-up created to take commercial advantage of the potential embedded in a new embryonic technology. Most probably, the firm will be involved in one downstream alliance to have access to capital and to complementary asset. Conversely, as it has been said, a large firm needs to involve in a variety of upstream partnerships in order to minimize the negative effect of the risk and the incertitude intrinsic to innovations.

Hypothesis 4: As collaboration is necessary for a DBF to survive and for a large biopharmaceutical company to keep a competitive advantage, we hypothesize that size does not have an effect on the propensity of being involved in a collaborative arrangement. By contrast, we expect size to have a positive effect on the number of collaboration a firm is involved in.

1.2.2.5 PROTECTION OF INTELLECTUAL PROPERTY

The influence that the system of IP protection has on the propensity to collaborate is not clear, and the literature provides conflicting results (Schmidt, 2007). It can be argued that when a firm uses mostly strategic (non formal) means of protection, it is less willing to collaborate, to avoid the loss of competences and knowledge through spillovers. Conversely, when new knowledge is protected through a formal method (in the biotechnology sector, typically a patent), a firm is more willing to collaborate, as any use of the new knowledge by others is legally protected. Mowery (1998) observes that, in the biomedical industry, the strength of the intellectual property protection, once obtained, is considerable, and fosters the transfer of a new technology from one agent to another. However, a large company could collaborate with a small dedicated-biotechnology firm which does not hold any patent, hoping to appropriate new knowledge. For this reason, we do not formulate any hypothesis on the effect of the number of patents held on the propensity of being involved in at least one collaborative arrangement, as well as on the number of partnerships in which a firm is involved.

1.2.2.6 BIOTECHNOLOGY R&D ORIENTATION

In the literature on alliances R&D orientation is one of the major factors explaining the propensity to collaborate for R&D activities (Bayona et al, 2001). Biotechnology is a knowledge intensive industry, where R&D plays a fundamental role and is the engine and the essence of this industry. This industry-specific characteristic emphasizes the primary role of collaboration among firms, and states the importance of the alliances within the biotech industry.

At the firm level, R&D orientation could be argued having a positive effect on the likelihood to be involved in one or more collaborative arrangements. However, we can conceptually link a firm's R&D orientation with its level of innovativeness, even though the variables employed in the analysis are not formally correlated. This opens up an important issue regarding the use of an econometric analysis. A negligible formal correlation

between two variables does not necessarily imply that in reality the two aspects are not linked. The results must therefore be conscientiously interpreted and integrated with a more qualitative analysis of the industry.

Hypothesis 5: a higher R&D orientation has a positive effect on the propensity to be involved in a collaborative arrangement with another agent of the innovation system. R&D orientation is measured through the percentage of expenditures for R&D.

1.2.2.7 EXPORT ORIENTATION

Although no empirical evidence is found in this respect, a link between export orientation and the propensity to collaborate is argued by Dachs et al. (2004). For the biotechnology industry, international collaboration is a common practice, and the usual pattern (small firms collaborate for R&D activities with larger pharmaceutical firms to have access to capital) can be recognized, as Kang and Sakai (2000) explain.

Hypothesis 6: measured as the percentage of the total revenues coming from exportation, we hypothesize that this characteristic has a positive effect on the propensity to collaborate.

1.2.2.8 AGE

In the literature, no evidence is provided proving that the age of a firm has an influence on its propensity to form collaborative arrangements with the other agents of the innovation system. An explanation can be provided considering the biotechnology innovation process and the vertical structure of the alliances: universities, new biotechnology firms and incumbents collaborate in order to take advantage of their partners' complementarities and thus improve the efficiency of innovative activities. In addition, a strong relationship exists between age and size, especially for small firms, which undergo a growth process. In order to evaluate these effects, a cross-sectional analysis is not enough, and a more complex framework following the evolution of a firm over time is required. For

this reason in our analysis, it will be difficult to assess the effect of a firm's age on the propensity and the intensity of collaboration.

1.3 SUMMARY

This chapter began considering that patterns of new knowledge creation in the biotechnology sector are now more established and understood, the industry is more profitable, and collaboration between small biotech firms and large biopharmaceutical companies is now a well-acknowledged win-win strategy to turn new ideas into new successful products. The complexity of the business model leading to the commercialization of a new product in the biotechnology industry is extremely high, requiring a wide variety of very different skills and competences that a single agent can hardly hold. For this reason, collaboration between universities, biotechnology-dedicated firms and large biopharmaceutical companies is a necessary condition for success for new product and process development. In particular, this industry is characterized by a specific structure of collaboration: a vertical alliance chain is identified in the human health biotechnology sector. Alliances follow the steps that the invention, development and commercialization a biotechnology product require in the drug industry. Universities and public labs are often the source of basic, fundamental knowledge. This way, new fresh knowledge created in universities, hospitals and government laboratories, is often transferred to R&D-oriented small DBF's. They provide further research in order to turn fundamental knowledge into a potentially marketable product; however, they lack the resources to perform clinical test, to produce and to commercialize a new product. For this reason, small DBF's, after R&D project completion, need to transfer the technology they developed to a large company, which has the necessary resources to test, manufacture and market a new product.
Finally, in the second section of the chapter, we introduced the characteristics affecting collaborative behavior in the biotechnology industry, building on the existing literature. The impact of size, orientation towards biotechnology, IP protection and a number of other factors is examined, and the hypotheses we will test in the following of this work are introduced.

CHAPTER 2

METHODOLOGY

In this chapter, we detail the procedures employed to analyze data collected by Statistics Canada in the Biotechnology Use and Development Surveys (BUDS). After a brief overview of the techniques used to perform the analysis, we focus on data issues, in particular imputation and data manipulation.
The analysis is performed employing the survey package implemented in STATA 10, whose relevant commands are briefly discussed in the following. For a deeper understanding of the features of this software, it is highly recommended to consult the STATA 10 user's guide.

In this work, three main techniques are used in order to achieve the ends we introduced in the first chapter:

1. *Summary statistics and contingency tables*: computing means and proportions of the variables, we intend to provide a description of the characteristics of the Canadian biotechnology industry. Moreover, calculating summary statistics and contingency tables for subgroups, we investigate on the differences among small, medium and large firms and among provinces, particularly focusing on collaborative behavior.

2. *Logit regression*: the effect of the relevant firm-specific characteristics on the propensity to be involved in at least one collaborative arrangement is evaluated using different logit models. As we will see below, several different dependent variables are used, and different independent variables to measure the relevant factors are employed.

3. *Negative Binomial Regression*: the effect of the relevant firm-specific characteristics on the number of collaborative arrangements a firm is involved in is evaluated. As in the logit analysis, several dependent variables are used, and different ways of measuring the relevant factors are employed.

As it is explained in Annex G, methodologists at Statistics Canada use the so-called post-stratification sampling technique in order to take account of the response rate. In fact, stratification is performed on three dimensions (industrial code, size and province/territory). This way, the resulting number of strata is extremely high[15]. In this situation, of course, it is not possible to maintain a formal definition of strata for estimation: consider, for example, that the overall mean of a variable is given by the weighted mean within each stratum; if a formal stratum has less than two observations, estimation cannot be performed on that specific stratum. This requires the introduction of sampling weights in order to take into account non-respondents (Traoré, 2004), and the use of a single stratum: stratification is therefore a technique to compute the sampling weights *a-posteriori*.

2.1 DATA

We used data gathered in the Biotechnology Use and Development surveys. We do not provide here a detailed description of the questionnaires[16] to avoid prolixity.

Data collected in the surveys is provided under the form of a `.dta` file, which can be directly used as an input for STATA 10. In order to be analyzed, data require, in addition to renaming, some manipulation and transformation. Here we present the most common cases encountered:

[15] For example, the 1999 survey has 468 different strata and just 223 observations, which would make the adoption of formal stratification nonsensical.

[16] The pdf version of the questionnaires is availabe on the Statistics Canada website: http://www.statcan.ca/english/sdds/indexai.htm

1. For certain dichotomous questions, the respondent is required to check if the answer is "*yes*", and to leave a blank otherwise. Data are thus recorded under the form of a positive integer if the answer was "*yes*", and of a missing value otherwise. In this case, "*missing value*" has a precise meaning, and the replacing criterion is straightforward. In order to calculate proportions, we replaced the missing value with a zero, and the integer value with a 1 using the command `replace`.

2. For certain dichotomous variables, the respondent is required to choose between two options, generally "*yes*" or "*no*". In this case, the variable is recorded using three possible values. A "*no*" is generally reported as 3, a "*yes*" as a 1 and a blank is reported as a missing value. In this case, the meaning of a missing value is different from the preceding case, and a different logic must be adopted. Nevertheless, the response rate for this type of questions in the surveys is usually equal to 100%, except for nested questions. In this case, replacing is straightforward: if the preceding question was a "*no*", as blank in the nested question can be considered as a "*no*" as well.

3. The answer to certain questions is an integer: for example, the number of collaborations in which a firm is involved. In this case, the answer is reported under the form of an integer or a missing value. Replacing is done following a case-by-case procedure, considering the relationship between the answer to other linked questions. In this type of situations, a specific logic criterion drives us to a solution; to avoid prolixity, we do not examine the details of every choice.

4. Some questions require a value to be introduced. For example, total revenues of the firm, the percentage of the expenditures for biotechnology-related activities… For the variables used to fit the logit and the negative binomial models, the rate of response was always very high (in almost all cases equal to 100%). In these cases, the choice was between non-replacing (only non-missing value

would be used, reducing the size of the sample), substitution with the average (missing values would be replaced by the average of non-missing observations) or substitution with zero. We decided to use the last of these techniques, despite the non-consistent meaning of a zero. In fact, we prefer to treat this type of missing values as outliers, rather than fixing them at the mean value of non-missing observations.

Some of the variables used in our analysis directly coincide with the variables collected in the surveys. Sometime, however, some manipulation is required to obtain the variables we need. We give the example of the variable `ecipub` which is equal to one if the firm is involved in a collaborative arrangement with a public institution, zero otherwise. The questionnaires do not directly address the question, and a new variable must be created, using the command `generate`. The variable `ecipub` can be thus calculated as follows:

$$\texttt{ecipub} = \texttt{ecuh} \cup \texttt{eclg}$$

where the symbol $\cup$ has the meaning of logic union[17]. This way, the variable `ecipub` takes the value zero only if the two variables on the right side of the equal sign are zero. If at least one of them is equal to one, then also `ecipub` is equal to one.

A complete list of the variables drawn from the BUDS is provided in Annex A. A synthetic description of each variable is provided, along with its type (only for binary, integer and Lickert-scale variables), coding (when not evident) and in which survey it appears. In fact, as the questionnaire evolves continuously following the needs and the most recent understanding of the biotechnology industry, not all the variables can be found in each survey. Unfortunately, also some of the independent variables used in the logit and negative binomial regression do not appear in all of the surveys; this limits the

[17] Adopting this framework, the value 1 is considered as "*true*", while the value 0 denotes a "*false*".

possibility to compare some results from different surveys, and causes some models, initially though as different, to collapse into the same one.

Only after data are organized as desired, the analysis can be performed. First, the sampling design needs to be considered. We must take account of the specificity of the way population is divided (if ever) and of the way the sample is drawn. In STATA 10, a special package for survey analysis is implemented, so that the sample design can be taken into account and estimates can be calculated in the form introduced in Annex G. The software requires survey data to be set through the command `svyset`. Strata, clusters, sampling weights and other possible sampling characteristics must be declared at the beginning of each program. Post-stratification (or stratification through sampling weights) implies the use of a single stratum that does not need to be declared[18]. Sampling weights need to be declared through the command `svyset` as follows:

```
svyset [pweight=name_var]
```

In each survey, sampling weights are collected in a specific variable. Table 2.1 shows the name of the variables used by Statistics Canada to gather the sampling weights.

Table 2.1 : Sampling weights variables.

Survey	1999	2001	2003	2005
Weights	`Weight99`	`Weight01`	`weight03`	`weight05`

In addition to summary statistics, also logit and negative binomial models (and a wide variety of other models) can be fitted using the survey package; moreover, survey-specific post-estimation and model diagnostic techniques can be employed. In this respect, the most important aspect for our work concerns scalar statistics for testing the fit

[18] We obtain a stratified sample by dividing the entire population in mutually exclusive subdomains, within each of which a sample is drawn using a specific technique. Therefore, a non-stratified population can be considered as a single-stratum population.

of a model: if sampling weights are used, it is not possible to evaluate some of the pseudo-R^2 statistics, as we will see below in this chapter, when we deal with logit and negative binomial models.

2.2 SUMMARY STATISTICS AND CONTINGENCY TABLES

Through summary statistics and contingency tables, we aim at outlining the main characteristics of the Canadian biotechnology industry, particularly focusing on collaborative behavior. Rich summary statistics are provided in the "*Canadian Trends in Biotechnology*" (2005) and by (Lonmo and McNiven, 2007) Of the wide variety of variables used in the BUDS, we focus on alliances: for example, proportions and average number of per-firm partnerships are given by province and firm size. We also consider different types of alliances and different type of partners. Among others, partnerships aimed at increasing the firm's R&D capacity, production and manufacturing, or aimed at gaining access to external capital, skill and knowledge are considered. Subgroup means and totals are computed for all the variables listed in Annex A; nevertheless in the following we will only provide the most interesting and relevant results concerning collaborative behavior.

To compute means and totals respectively, we use the following commands

```
svy: mean
svy: total
```

Employing the option `over` it is possible to compute these estimates for subpopulations, taking account of sampling weights, as it has been said above. For example, for the variable `ec` we obtain the following command lines:

```
svy: mean ec, over (province3 taille)
svy: total ec, over (province3 taille)
```

The output is the list of the means and totals of the variable `ec` within each of the 15[19] subgroups. The use of the variable `province3` (see Annex A) is necessary to avoid the identification of a specific biotechnology firm. In fact, considering the small number of large biotechnology firms, the use of the non-aggregate variable for the province would lead to obtain some subgroups with just one element. In order to observe the Canadian law on statistics, we cannot provide any result that could allow the identification of a specific firm.

In addition to the command `svy: total`, the command `svy: tab` can be used to obtain contingency tables. Contingency tables provide two-way tables for frequency counts, along with a Pearson statistic to test for independence among different subsamples. This way it is possible to detect the differences among subgroups (as usual, the population is classed by province and size) of firms in order to outline the geography of the characteristics of the biotechnology industry. The usefulness of the command `svy: tab` resides in its flexibility and in the options that can be used. In this work the following syntax for `svy: tab` is employed:

```
svy: tab province3 taille, tab(name_var) se ci row
```

The output is a two-way table in which the sum of each row equals to 100%, as the option `row` is used. Each cell is the proportion of the counts of the variable `ec` for the subgroup identified by its raw and column. Therefore, Table 2.2, which is a two-way contingency table for variable `ec` (2005 survey), must be interpreted as follows.

[19] five aggregated provinces and three firm sizes.

Table 2.2 : Contingency table for variable `ec` : total number of partnerships, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	69.8%	18.9%	11.3%
Ontario	71.8%	18.8%	9.4%
Prairie Provinces	50.0%	26.1%	23.9%
British Columbia	89.4%	7.3%	3.3%
Total	72.1%	16.9%	11.0%

For example, the value 69.8% (Small firms, Québec) represents the proportion of small collaborative firms in Québec among all collaborative firms in Québec. In other words, among all collaborative biotechnology firms located in Québec, the 69.8% has less than 50 employees. Clearly, every line sums up to 100%. Contingency tables are thus a simple but powerful tool to investigate on two-way non-homogeneities of a certain variable. Comparing the values of the same column, it is possible to detect differences in the collaborative behavior (or in any other observed variable) among different provinces with respect of a certain firm size. For example, while in British Columbia the 89.4% of the collaborative firms are small, in the Prairie Provinces this proportion is just 50.0%.

Standard error and confidence interval for each subgroup estimate are displayed using the options `se` and `ci`, respectively. However, to avoid prolixity, these estimates are not reported in the tables. For partnership-related variables (Table A.1 in Annex A) we also used the option "`count`" for the command `svy: tab` in order to obtain the weighted cell counts. The default test for independence is based on a Pearson chi-squared statistic converted into an F-statistic using the second order correction proposed by Rao and Scott (1984). This test works well in every situation, and there is no need to adopt a different one, although in the literature we can find a variety of other tests (see STATA 10 user's guide).

It is important to note that through contingency tables, it is only possible to conduct an analysis on subdomains, and the conclusions should not be intended as describing the behavior of a single firm. Rather, the results will allow us to depict the Canadian biotechnology industry with a particular focus on collaboration and to contextualize in a precise scenario the results arising from logit and NBR analysis.

2.3 REGRESSION MODELS

The firm-specific characteristics having an influence on the propensity to be involved in at least one collaborative arrangement with another firm or public institution are identified employing the logit regression. The variables taking into account the collaborative behavior assume the value one if the firm is involved in at least one collaborative arrangement, zero otherwise. We will employ the Negative Binomial Regression to investigate the effect that firm-specific characteristics have on the number of collaborative arrangements a firm is involved in. In this case, the dependent variable is a count, which justifies the use of NBR.

Three variations of each model are tested:

- *Split by size for small firms*: the model is fitted on the subgroup of small firms. Through the option `subpop` in STATA 10, it is possible to restrict the population to the desired subpopulation. Implementing this restriction requires modifying the syntax of the command as follows: `svy,subpop(dpetite):logit` *`dep_var indep_vars`*. This way, only small firms are included in the subgroup.
- *Split by size for small and medium firms*: the same considerations as in the preceding case apply, with the difference that the subgroup includes both small and medium firms. The syntax of the option is the following: `svy,subpop(dpme):logit` *`dep_var indep_vars`*.

- *Whole sample*: the analysis is performed using the whole sample of small, medium and large biotechnology firms.

Unfortunately, it is not possible to perform the analysis just for medium firms and for large firms, as the number of observations in the subsample would be too exiguous.

Through the command `svy: logit` it is possible to estimate the parameters of the logit model, along with their standard deviation, t-statistic and p-value and confidence interval. The syntax of `svy: logit` is the following:

```
svy: logit depvar indep_var_list
```

where `depvar` denotes the dependent variable, and `indep_var_list` denotes the list of the independent variables. For each parameter of the model, the output of `svy: logit` lists the value of the coefficients associated with the variables, their standard error, the t-statistic and p-value and, finally, the 95% confidence interval. In addition, an F-test for the overall significance of the model is automatically performed. The degrees of freedom of the F-statistic are equal to the number of independent variables used in the model and to the number of observations minus the number of independent variables. The output of `svy: logit` also displays the number of observations, the population size (calculated on through the sampling weights) and the overall design degrees of freedom (equal to the number of observation minus the number of strata).
As we are interested in evaluating the effect of each factor on the propensity to collaborate (and on the number of alliances), we do not need to employ tests for complex hypotheses, like Wald, Likelihood Ratio or Lagrange Multipliers. For simple hypotheses testing, it suffices the default t-test automatically performed using `svy: logit`. In fact, as the sample grows in size, the t-distribution tends towards a normal distribution; therefore the t-test tends to be equivalent to a chi-squared test, on which the Wald test is based.

To fit a negative binomial model for survey data, we employ the command `svy: nbreg` in STATA 10, whose syntax is the same as `svy: logit`, and the same aforementioned considerations apply. The output layout is similar to the one provided by `svy: logit`. An F-test on overall significance is performed; the F-statistic is reported, along with the associated p-value. Coefficients, standard error, t-statistic and associated p-value and confidence interval are listed for each parameter. Also, the value of α, its standard error and confidence interval is displayed. We recall that α is introduced in the negative binomial regression to take account of data overdispersion (see Annex G).

Measures of fit deserve a particular attention when using the survey package, as most of the usual scalar measures of fit cannot be computed when, as in our case, sampling weights are employed. With STATA 10, it is possible to use the command `fitstat` to get a variety of scalar measures of fit: R^2 and Adjusted-R^2, Efron R^2, McFadden R^2, Cragg and Uhler R^2 and McKelvey and Zavoina R^2. Unfortunately, of all these coefficients, only McKelvey and Zavoina R^2 can be calculated if sampling weights are used. This pseudo-R^2 compares the estimated variance of the latent variable and the estimated variance of the error:

$$R^2_{MZ} = \frac{Var(\hat{y}^*)}{Var(\hat{y}^*) + Var(\varepsilon)}$$

In addition, the variance of the latent variable and the variance of the error are displayed. We recall that the latent variable is a conceptual artefact, is not directly observed and therefore its variance could not be estimated. Nevertheless, as it is explained in Annex G, there exists a way to get an estimate of the dispersion of the latent variable as:

$$Var(\hat{y}^*) = \underline{\hat{\theta}}' Var(\underline{x}) \underline{\hat{\theta}}.$$

The syntax of `fitstat` is extremely simple: after the command `svy: logit` it suffices to use `fitstat`, and all the measures of fit are provided, with the restrictions imposed by the sample design.

The following two subsections deal respectively with the dependent variables employed in logit and NBR regressions and with the independent variables employed in the models. In addition, we explain our choice in selecting the variables as a measure of the factors being supposed to have an effect on the propensity to collaborate (for logit analysis) and on the number of collaborative arrangements (for NBR regression).

2.3.1 DEPENDENT VARIABLES

For the logit analysis, we employ several dependent variables, in order to consider the effect of the factors on the propensity to be involved in different types of alliances and with different type of partners. In particular, we have considered knowledge-related and production/commercialization-related variables, with a firm and with a public institution. This choice is justified by the specific pattern of alliances at different stages of product development in the biotechnology industry, as described in the first chapter of this work. The differences between the four questionnaires are reflected in the non-availability of all the dependent variables in the four surveys.. Table 2.3 shows the dummies we employed as dependent variables for logit analysis in the four surveys.

Table 2.3 : Dependent variables for logit analysis, 1999 to 2005 surveys.

Variable	Descprition	1999	2001	2003	2005
ec	The firm is involved in at least one collaborative arrangement	X	X	X	X
ecc	The firm has at least one partnership concerning knowledge	X	X	X	X
ecpc	The firm has at least one partnership concerning production/commercialisation	X	X	X	X
ecepri	The firm has at least one partnership with another firm	X	X	X	X
ecipub	The firm has at least one partnership with a public institution	X	X	X	X

Table 2.4 shows the counts used as dependent variables for Negative Binomial Regression analysis. With the exception of nec, measuring the total number of collaborative arrangements a firm participates in, no other variable is available throughout the four surveys. The variables necc, necepri, necipub and necpc are not available for the 1999 survey.

Table 2.4 : Dependent variables used in Negative Binomial analysis, 1999 to 2005 surveys.

Variable	Descprition	1999	2001	2003	2005
nec	Number of collaborative arrangements	X	X	X	X
necc	Number of partnerships concerning knowledge		X	X	X
necpc	Number of partnerships concerning production/commercialisation		X	X	X
necepri	Number of partnerships with another firm		X	X	X
necipub	Number of partnerships with a public institution		X	X	X

2.3.2 INDEPENDENT VARIABLES

As it has been said, we aim at testing the influence of the following factors both on the propensity to be involved in at least one collaborative arrangement with another firm/public institution and on the number of partnerships that a firm forms:

- Size
- Biotechnology orientation
- Spin-offs
- Level of innovativeness
- Biotechnology products/processes
- IP protection
- Export orientation
- Public support
- Capital financing
- Control for the environment

We now provide a detail description of the independent variables used to measure the characteristics that we hypothesize to have an effect on a firm's propensity to collaborate. The main references inspiring the choice of the independent variables are provided by Dachs et al. (2004) and by Schmidt (2007).

2.3.2.1 SIZE AND ORIENTATION TOWARDS BIOTECHNOLOGY

We measure firm size in two different ways: through the total number of employees of the firm (variable `e`) and through the total revenues (variable `rt_0`). To maintain coherence in the measures within each model, where the variable `e` is chosen to determine the firm size, the biotechnology orientation is determined using the variable `ebper` (percentage of the firm's employees having biotechnology-related responsibilities). Conversely, when firm size is measured through its revenues (variable `rt_0`), biotechnol-

ogy orientation is determined by the variable rbper0 (percentage of revenues from biotechnology).

As we expect the values of rt_0 to be much higher than the other variables (and therefore the associated coefficients of the logit and negative binomial regression to be much smaller than the others), we take its log, creating the variable lrt_0.

Examining the effect of biotechnology orientation on the propensity to collaborate we aim at understanding whether biotechnology-dedicated firms need to form alliances more than other inventive firms that have a significant portion of their operations in other industrial sectors. The adoption of a relative measure (a percentage) eliminates the correlation between the variables related to firm size.

2.3.2.2 BIOTECHNOLOGY R&D ORIENTATION

Orientation towards R&D is measured trough the variable lrdt_0 (the log of rdt_0, as we did in the case of the variable rt_0). This variable represents the total R&D expenditures, and is employed together with rdbper0, measuring the percentage of R&D expenditures for biotechnology. This choice poses a problem, as there could reasonably be a strong correlation between rdbper0 and ebper: in proportion, the more a firm spends for R&D, the larger the R&D department. Surprisingly, the highest value of the rough coefficient of correlation is 46.9% in 1999 (45.9% in 2005), which is still acceptable. In 2001 and 2003, R^2 is equal to 7.5% and 10.0%, respectively. Although we take 40% as a threshold to distinguish between correlated and uncorrelated variables, values between 40% and 50% could still be considered acceptable. This fact opens a major issue on the use of this proxy's as measures of a firm's level of innovativeness. It could be that the variables rdbper0 and ebper may be employed in the same model in order to investigate on the different effect they have on collaborative behavior. This way, if we detect a significant difference, it could be that these two proxy's do not measure the same characteristic, and further work would be required. In this case, an in-the-field research would be required to assess the meaningfulness of revenues from biotechnology

and personnel having biotechnology-related responsibility in explaining the orientation towards collaboration.

2.3.2.3 SPIN-OFFS AND SUBSIDIARIES

We also consider the impact on collaborative behavior of whether the firm was created through a spin-off or not. The effects of being generated as a spin-off are not easy to assess, as they fade away over time, in particular when we consider the type of partnership the firm is involved in. It is reasonable to think that a spin-off, created for example to turn a new idea into a marketable product, needs to form alliances to raise capital in the first stages of its life, and to gain access to complementary assets (regulatory and clinical affairs, marketing, production and manufacturing) later on in its activities. However, the differences in collaborative behavior between a spin-off and a non-spin-off remain to be investigated. Including the variable `dspin` in the models that we tested, we intend to get an understanding of this particular mechanism of generation and its effect on the rate and type of collaboration. Finally, and only for the 2005 survey, we will investigate the effect on the probability to collaborate of being a subsidiary of a larger company through the variable `dsubs`.

2.3.2.4 BIOTECHNOLOGY PRODUCTS

In most models, we investigate whether there is a difference in the collaborative behavior among firms developing products rather than processes that require the use of biotechnology. This way, we employ the dummy variables `dprod` (which assume the value one if the firm is currently developing products requiring the use of biotechnology) and `dproc` (which assume the value one if the firm is currently developing processes requiring the use of biotechnology).

A specific model is tested to investigate the effect that biotechnology products have at different stages of their development on collaborative behavior. The variables `nprodrd` (Number of products/processes at R&D stage) `nprodpc` (Number of prod-

ucts/processes at pre-clinical trials/confined field trials/pre-market stage) nprodrc (Number of products/processes at regulatory phase, unconfined release assessment or final pre-market assessment stage) nprodpm (Number of products/processes at approved, on market and at production stage) are used in substitution of dprod and dproc.

2.3.2.5 INTELLECTUAL PROPERTY PROTECTION AND CONTRACTS

Two measures of IP protection strategies are used, at two different levels of generalization. The first one consider the number of existing patents (variable nbe) and of pending patents (variable nba), while the second employs the total number of rights assigned/licensed to another Canadian firm (variable ndroitaec) and the total number of rights obtained from another Canadian firm (variable ndroitoec). The former method is adherent to the argument exposed in the first chapter, while the latter considers IP protection in a broader perspective. In addition to considering patents and, more in general, IP rights, we estimate a model employing the variables ncont (total number of contracts) and nacont (total number of provided contracts). This choice is justified considering that the number of contracts is related to the formal protection of Intellectual Property rights: in general, a firm willing to contract out some of its activities needs first to protect new knowledge from the competitors. For this reason, we deem reasonable not to employ contract-related variables and IP-related variables in the same model.

2.3.2.6 EXPORT ORIENTATION

In each model, orientation towards exportation is measured through the variable rexpperb_0 (percentage of revenues from exportation of biotechnology products). By taking the percentage, we aim at obtaining a measure that is not correlated with other variables that depend on the size of the firm.

2.3.2.7 PUBLIC SUPPORT

Public support is measured through the variables `sourcekgov` (percentage of funding from government sources) and `lsourcekgovtot` (capital raised from government sources). As in the case of the variable `rt_0`, we used the log of `sourcekgovtot` to avoid large-scale discrepancy between the coefficients.

2.3.2.8 CAPITAL FINANCING

A model taking into account whether firms were successful in raising capital (variable `reusk`) and the total capital raised (variable `fkreun`) is tested. The use of this model suggests a preliminary question on the direction of the relationship collaboration-attempt to raise capital. For example, if we find a significant correlation between being successful in raising capital and propensity of being involved in at least one alliance, it would be hard to define the direction of the cause-effect link. In other words, is it the failure in raising capital that causes a lower propensity to collaborate or vice-versa? The answer is not straightforward in this case, and a precise understanding of this two-way relationship requires further investigation and research. Finally, we to avoid scale discrepancy among coefficients, we employed the log of the variable `fkreun`, creating the new variable `lfkreun`.

2.3.2.9 CONTROL FOR THE ENVIRONMENT

An external variable needs to be included in each model to capture the effect of the environment on the collaborative behavior. After the first tests for correlation among the variables, and after checking the significance of different "environment variables", we chose to employ the variable `nbppopprov`, defined as the number of biotechnology-related patents issued within a certain province during the considered time lag, divided by the population of that province.

CHAPTER 3

RESULTS - SUMMARY STATISTICS

In this chapter, we provide the summary statistics describing collaborative behavior within the Canadian biotechnology industry, distinguishing firms by their size and, where particular interest exists, by their location.

3.1 GENERALITIES

The number of biotechnology firms in Canada has been constantly growing over the last years: they were 358 in 1999, 375 in 2001, 490 in 2003 and finally 532 in 2005, as Table 3.1 shows. Disaggregate by-size data reveal that this industry is mainly composed by small firms, confirming what has been said in Chapter 1: biotechnology is a knowledge-intensive sector, reposing on scientific research, and small firms have the necessary dynamism, flexibility and nimbleness to perform leading edge R&D projects.

Table 3.1 : Changes in biotechnology firms in Canada by size, 1999 to 2005.

Firms	1999	2001	2003	2005
Small firms	259	267	352	397
% of small firms	72.3%	71.2%	71.8%	74.6%
Medium firms	60	62	77	83
% of medium firms	16.8%	16.5%	15.7%	15.6%
Large firms	38	47	61	52
% of large firms	10.6%	12.5%	12.4%	9.8%
Total biotech firms	358[20]	375	490	532

[20] The total is slightly different than the sum of the small, medium and large firms due to rounding effects.

It is interesting to note that from 1999 to 2005, firms located in Québec surpassed in number firm located in Ontario. Table 3.1 also reports the proportions of the firms of each size with respect to the total number of biotechnology firms for every survey. For instance, we see that small firms accounted for 72.3% of the total biotechnology firms in 1999, for 71.2% in 2001, for 71.8% in 2003 and for 74.6% in 2005. Small firms always account for more than the two thirds of the total number of biotechnology-oriented firms in Canada. About one sixth of the total biotechnology firms in Canada are composed of medium firms, and the rest are large firms. It is worth noting the drop both in the proportion and in the number of large biotechnology firms between 2003 and 2005. Further details on the geography and the division by size of biotechnology firms in Canada are provided in Table C.1 in Annex C. What is interesting to note is the constant growth in the number of biotechnology small firms located in Québec: there were 66 in 1999, 88 in 2001, 104 in 2003 and 133 in 2005. Ontario, by contrasts, registered a fluctuating trend: small biotechnology firms were 83 in 1999, and they dropped to 71 in 2001[21]; then, they grew to 92 in 2003 and to 99 in 2005. Maybe the most interesting insight concerns British Columbia. After a drop from 63 small biotechnology firms in 1999 to 53 in 2001, the recovery was fast. In 2001, there were 65 small biotechnology firms in this province and 82 in 2005.

While the total number of biotechnology firms has been growing over time, the percentage of those involved in at least one collaborative arrangement has been lowering: from 62.5 % in 1999 to 52.6% in 2005, as shown in Table 3.2. Actually, the drop in the proportion of collaborative firms took place between 2001 and 2003, while between the other surveys it maintained approximately constant. Disaggregate data reveals that even though small firms are at the heart of biotechnology innovation, the proportion of small collaborative firms is significantly lower than that of larger firms. It is important to re-

[21] It is natural to ask the reason of this drop: small firm may have grown and become medium-sized or even large-sized; or they may have exit. To further investigate on this aspect, a "linked analysis" needs to be performed using the Biotechnology Use and Development Survey, in order to follow the evolution of the firms.

mark the drastic drop in the proportion of large collaborative firms: in 1999, large collaborative firms accounted for 81.6% of the total large biotechnology firms and just for 59.6% in 2005. The same overall trend, although less pronounced, is observed for medium firms: 80.0% of them were involved in at least one collaborative arrangement in 1999, 51.6% in 2001, 49.4% in 2003 and 56.6% in 2005. For small firms we recognize the same overall trend: 56.0% of them were involved in at least one collaborative arrangement in 1999, 58.8% in 2001, 50.3% in 2003 and 50.9% in 2005.

Table 3.2 : Changes in collaborative biotechnology firms in Canada by size, 1999 to 2005 and by-size percentages of collaborative firms.

Firms	**1999**	**2001**	**2003**	**2005**
Small collaborative firms	145	157	177	202
% of small collaborative firms	56.0%	58.8%	50.3%	50.9%
Medium collaborative firms	48	32	38	47
% of medium collaborative firms	80.0%	51.6%	49.4%	56.6%
Large collaborative firms	31	38	36	31
% of large collaborative firms	81.6%	80.9%	59.0%	59.6%
Total collaborative firms	224	226	251	280
% of collaborative firms	62.5%	60.3%	51.2%	52.6%
Contingency table: collaborative firms				
Small firms	64.7%	69.5%	70.5%	72.1%
Medium firms	21.4%	14.2%	15.1%	16.8%
Large firms	13.8%	16.8%	14.3%	11.1%

By conveniently rearranging data, we can calculate the proportions of collaborative firms of each size with respect to all collaborative firms of a certain year (contingency table sorted by column). For example, in 1999, among the 224 collaborative firms, 64.7% was composed by small firms, 21.4% by medium firms and the remaining 13.8%

by large firms. It is interesting to note that the proportion of collaborative small firms is constantly increasing over time. The trend for medium and large firms, instead, is not homogeneous over time: the proportion of medium collaborative firms with respect to the total number of collaborative firms underwent a drop between 1999 and 2001 (from 21.4% to 14.2%), while in the subsequent years the proportion has been slightly growing up (15.1% in 2003 and 16.8%). Considering that in the biotechnology industry patterns of new knowledge creation are now more established and understood, this result provides strong evidence that collaborative behavior is a necessary aspect of the business model in this industry. Further data are provided in Table C.2 and Table C.3 in Annex C, which present respectively the number of collaborative biotechnology firms and the percentage of collaborative ones, divided by size and by province. We report here a summary table, disaggregated by province, in order to investigate on the different rates of collaboration between provinces. Table 3.3 provides evidence on the geography of small collaborative firms across the Canadian provinces, and should be interpreted as follows. For example, in 2005, of the total collaborative firms in British Columbia, 89.4% had less than 50 employees, while this proportion was 77.8% in 2003, to 74.6% in 2001 and to 87.7% in 1999. We notice that, on average, in this province small enterprises represent the highest proportion of collaborative firms in all the four surveys. Moreover, it is interesting to note that the proportion of small collaborative firms has gone up from 1999 to 2003 in Québec and Ontario, undergoing then a slight decrease from 2003 to 2005.

Table 3.3 : Proportion of small collaborative firms by province, 1999 to 2005.

Province	1999	2001	2003	2005
Atlantic provinces	n/a	75.9%	77.0%	n/a
Québec	56,2%	63.8%	72.1%	69.7%
Ontario	55.9%	73.3%	76.1%	71.8%
Prairie provinces	51.8%	64.1%	56.0%	50.0%
British Columbia	87.7%	74.6%	77.8%	89.4%

Along with the drop in the proportion of collaborative firms, also the total number of collaborative arrangements decreased from 1148 in 2001 to 1031 in 2003. However, the overall trend of the total number of collaborative arrangements set by biotechnology firms increased over time (from 694 in 1999 to 1427 in 2005), as Table 3.4 shows. It is interesting to notice that the most important change in the number of alliances comes from the collaborative behavior of large firms: in 2003 large firms formed only 180 partnerships, less than the half compared to 2001. Comparing data gathered in Table 3.2 and in Table 3.4 reveals something interesting: the proportion of the number of alliances formed by small firms is always smaller than the proportion of small collaborative firms, meaning that, among collaborative firms, large and medium firms are involved, on average, in a higher number of alliances than small firms do. For example, in 2003, 70.5% of the firms involved in at least one collaborative arrangement were small-sized, but the number of the alliances formed by them accounted only for 55.0% of the total number of partnerships of that year.

Table 3.4 : Change in the number of alliances by firm size, 1999 to 2005.

Firm size	1999	2001	2003	2005
Alliances formed by small firms	401	621	567	954
% of alliances formed by small firms	57.8%	54.1%	55.0%	66.9%
Per-firm alliances: small firms	2.8	4.0	3.2	4.7
Alliances formed by medium firms	131	156	284	328
% of alliances formed by medium firms	18.9%	13.6%	27.5%	23.0%
Per-firm alliances: medium firms	2.7	4.9	7.5	4.7
Alliances formed by large firms	162	372	180	145
% of alliances formed by large firms	23.3%	32.4%	17.4%	10.2%
Per-firm alliances: large firms	5.2	9.8	5.0	4.7
Total number of alliances	694	1148	1031	1427

Table 3.4 also provides insight on the average number of per-collaborative-firm alliances, distinguishing between small, medium and large firms. While in 1999 and 2001 small firms were involved, on average, in much less collaborative arrangements than medium and large firms, this trend seems to fade away over time: in 2005, both small collaborative firms and large collaborative firms were involved in, on average, 4.7 partnerships each. In addition, the average number of per-collaborative firm partnerships underwent a drop between 2001 and 2003 (from 4.1 to 5.1), mostly due to changes in the intensity of collaboration of large firms. The usual anomaly for large firms between 2001 and 2003 is registered also by this index: the average number of collaborative arrangements dropped from 9.8 in 2001 to 5.0 in 2003, and continued to slightly lower in 2005, reaching the value of 4.7.

Table C.4 in Annex C shows the total number of collaborative arrangements distinguishing firms by their size and province, while Table C.5 in Annex C shows the average number of partnerships per collaborative firm by province and firm size. Moreover, Tables C.6 to C.9 in Annex C provide complete contingency tables for the variable `nec` (number of collaborative arrangements) by firm size and province for each survey. Small firms located in British Columbia appear to be the most collaborative, especially from 1999 to 2003. In 2005 small firms located in Québec and Ontario, show a more intense collaborative behavior, with 5.12 and 5.34 alliances per collaborative firm respectively. It is worth noting that the increase in the number of partnership per collaborative firms from 1999 to 2005 was 39.8% in Québec and 97.5% in Ontario[22]. It is worth noting the fluctuation in the number of collaborative arrangements per collaborative medium and large biotech firm, especially for Québec and Ontario.

[22] These percentages are calculated with respect to the average of the extreme values.

3.2 PARTNERS

We now tackle the issue of understanding collaborative behavior by considering the partners of collaborative arrangements. As Table 3.5 shows, the number of firms collaborative with another firm has, on average, grown up consistently from 1999 to 2005, against the overall trend of collaborative behavior. In addition, the number of firms involved in at least one interfirm collaborative arrangement has grown up: form 107 in 1999 to 194 in 2005. Concerning the proportions, we observe a fluctuating behavior in interfirm collaboration: in 1999, among all collaborative biotechnology companies, 47.8% was involved in at least one interfirm collaborative arrangement. This proportion was equal to 61.5% in 2001, dropped down to 47.7% in 2003 and grew to 62.3% in 2005. The overall trend marks a growth in interfirm alliances among collaborative firms, with the usual anomaly between 2001 and 2003. Among firms having at least one partnership with another firm, the vast majority collaborates within the biotechnology sector. In 2005, for example, 128 firms had partnerships with another biotechnology firm (66.0% of the firms having alliances with another firm), 92 (77.3% of the firms having alliances with another firm) in 2003 and 118 in 2001 (84.9% of the firms having alliances with another firm). This fact reveals how inter-sectoral collaboration is more and more diffuse over time. The reason of this behavior remains to be explained. One possible suggestion could be the following: as the development of a product/process that requires the use of biotechnology is very complex and needs a variety of competences/skills, and as each firm needs to focus on its core competences, partnerships with different agents belonging to different sectors are more and more necessary.

Table 3.5 : Collaborative firms by type of partner, 1999 to 2005.

Counts and percentages	1999	2001	2003	2005
Collaborative firms	224	226	251	280
Percentage of collaborative firms	62.5%	60.3%	51.2%	52.6%
Firms collaborating with another biotechnology firm	-[23]	118	92	128
Percentage of collaborative firms having a partnership with another biotechnology firm	-	52.2%	36.5%	45.7%
Firms collaborating with a pharmaceutical firm	-	-	-	54
Percentage of collaborative firms having a partnership with a pharmaceutical firm	-	-	-	19.3%
Firms collaborating with a non-biotech and non-pharmaceutical firm	-	-	-	67
Percentage of collaborative firms having a partnership with a non-biotech and non- pharmaceutical firm	-	-	-	23.8%
Firms collaborating with a non-biotech firm	-	21	27	-
Percentage of collaborative firms having a partnership with a non-biotech firm	-	9.3%	10.8%	-
Firms collaborating with another firm	107	139	119	194[24]
Percentage of firms collaborating with another firm	47.8%	61.5%	47.4%	62.3%
Firms collaborating with a hospital/university	122	83	79	138
Percentage of collaborative firms having a partnership with a hospital/university	54.5%	36.7%	31.5%	49.3%
Firms collaborating with a government lab	89	63	39	96
Percentage of collaborative firms having a partnership with a government lab	39.8%	27.9%	15.5%	34.4%
Firms collaborating with a public institution	165	109	100	192
Percentage of collaborative firms having a partnership with a public institution	73.7%	48.2%	39.8%	68.6%

[23] The symbol "-" means that data is not available due to differences in the questions in each survey.
[24] Clearly, as the categories are not mutually exclusive, the total of biotechnology firms collaborating with another firm is less than the sum to the different components.

The rate of collaboration with a public institution undergoes a singular trend: 73.3% of the collaborative firms in 1999 had at least one partnership with a public institution; the proportion dropped during the subsequent four years. In 2001, 109 biotechnology firms (48.2% of the collaborative firms) had alliances with a public institution, 100 in 2003 (39.8% of the collaborative firms) and 192 in 2005 (68.6% of the collaborative firms). Between 2003 and 2005, the rate of collaboration with a public institution (both universities/hospital and governments laboratories) increased substantially. Table 3.6 shows the number of small, medium and large firms that collaborate with another firm or with a private institution. While in 1999 and in 2001, the proportion of small firms (among those that collaborated with another firm) was lower than the overall proportion of small biotech firms, this characteristic fades away over time. Recall that small firms accounted for 72.3% of the total biotechnology firms in 1999, for 71.2% in 2001, for 71.8% in 2003 and for 74.6% in 2005. In 1999, the proportion of small firms that had at least one partnership with another firm was equal to 61.6%, to 68.0% in 2001, to 72.9% in 2003 and to 74.2% in 2005. The same overall trend can be recognized concerning alliances with public institutions. In 1999, 59.7% of the firms having at least one partnership with a public institution was small-sized; in 2001, this proportion was equal to 70.5%, in 2003 to 67.4% and in 2005 to 68.8%.

Table 3.6 : Collaborative firms and proportions by size and type of partner, 1999 to 2005.

Counts and percentages	1999	2001	2003	2005
Small firms collaborating with another firm	66	94	87	144
% of small collaborative firms collaborating with another firm	61.6%	68.0%	72.9%	74.2%
Medium firms collaborating with another firm	32	22	22	30
% of medium collaborative firms collaborating with another firm	29.6%	15.7%	18.8%	15.2%
Large firms collaborating with another firm	9	23	10	20
% of large collaborative firms collaborating with another firm	8.8%	16.3%	8.3%	10.5%
Small firms collaborating with a public institution	99	77	67	132
% of small collaborative firms collaborating with a public institution	59.7%	70.5%	67.4%	68.8%
Medium firms collaborating with a public institution	39	11	13	35
% of medium collaborative firms collaborating with a public institution	23.8%	10.3%	13.3%	18.3%
Large firms collaborating with a public institution	27	21	19	25
% of large collaborative firms collaborating with a public institution	16.5%	19.3%	19.4%	12.9%

Further details on the geography of collaboration by type of partner and firm size are reported in Tables C.24 to C.49 in annex C.

3.3 REASONS TO COLLABORATE

We want now to investigate on the reasons leading to collaboration. Table 3.7 shows that almost the totality of collaborative firms (96.0% in 1999, 91.6% in 2001, 65.7% in 2003 and 95.0% in 2005) has at least one knowledge-related partnership. The proportion of collaborative firms involved in at least one production/manufacturing alliance is much lower: 57.6% in 1999, 33.0% in 2001, 42.2% in 2003 and 35.7% in 2005. This result

emphasizes the fact that biotechnology industry is a knowledge-intensive sector, and that innovation stems from collaboration, corroborating the findings provided in Chapter 1. It is worth noting that the number of firms that collaborate for production and commercialization-related reasons follows a singular trend: they were 129 in 1999, then this value dropped to only 75 in 2001, it increased to 106 in 2003 and then underwent a slight drop in 2005, being equal to 100.

Table 3.7 : Collaborative firms by reason, 1999 to 2005.

Counts and percentages	1999	2001	2003	2005
Collaborative firms	224	226	251	280
Percentage of collaborative firms	62.5%	60.3%	51.2%	52.6%
Firms collaborating for knowledge-related purposes	215	207	165	266
Percentage of firms collaborating for knowledge-related purposes	96.0%	91.6%	65.7%	95.0%
Firms collaborating for production/commercialization	129	75	106	100
Percentage of firms collaborating for production/commercialization	57.6%	33.0%	42.2%	35.7%

Table 3.8 presents the proportions of small, medium and large firms involved in at least one knowledge-related partnership with respect of all firms having at least one alliance of this type. For example, in 2005, of the 280 firms collaborating for knowledge-related reasons, 71.3% were small-sized, 17.7% were medium-sized and 11.0% were large-sized. The comparison between this result and data gathered in Table 3.1 does not show any remarkable disproportion between the by-size proportions of firms collaborating for knowledge-related reasons and the overall by-size proportion of collaborative firms. Knowledge-related collaborative behavior can be considered quite homogeneous among firms of different size, as the proportion of small, medium and large firms collaborating for knowledge-related purposes is quite similar to the overall proportions of these types of firms with respect to the whole biotechnology sector. Table 3.8 provides a useful in-

sight on the nature of the drop between 2001 and 2003 in the number of firms that collaborated for knowledge-related reasons. The change in the number of small firms having at least one knowledge-related partnership was less important than the same change for small and medium firms. In 2003, 118 small firms collaborated for knowledge-related reasons, 16.4% less than 2001. This drop was equal to 18.3% for medium firms (form 31 in 2001 to 25 in 2003) and to 37.6% for large firms (from 36 in 2001 to 22 in 2003). This fact means that the drop in knowledge-related collaborative behavior comes mainly from large companies, which is confirmed by the changes in the number of per-firm knowledge-related partnerships. In 2001, large firms which collaborated for knowledge-related purposes had on average 8.8 partnerships each; this value consistently dropped in 2003 (3.4 partnerships per large firm) and didn't increase significantly in 2005, maintaining the value of 3.6 collaborations per large firm, less than the number of per-small firm knowledge-related alliances, which is equal to 3.8. Concerning medium-sized firms, the average number of collaborative arrangements constantly increased over time, from 4.2 in 2001, to 5.5 in 2003 to 6.3 in 2005.

Table 3.8 : Proportion of firms having at least one knowledge-related partnership, by size, 1999 to 2005.

Counts and percentages	1999	2001	2003	2005
Small firms collaborating for knowledge-related purposes	136	141	118	190
% of small collaborative firms collaborating for knowledge-related purposes	63.2%	67.9%	71.3%	71.3%
Per-small collaborative firm knowledge-related partnerships	-	3.3	2.6	3.8
Medium firms collaborating for knowledge-related purposes	48	31	25	47
% of medium collaborative firms collaborating for knowledge-related purposes	22.5%	14.8%	15.1%	17.7%
Per-medium collaborative firm knowledge-related partnerships	-	4.2	5.5	6.3
Large firms collaborating for knowledge-related purposes	31	36	22	29
% of large collaborative firms collaborating for knowledge-related purposes	14.3%	17.3%	13.6%	11.0%
Per-large collaborative firm knowledge-related partnerships	-	8.8	3.4	3.6
Small firms collaborating for production/commercialization-related purposes	88	60	80	72
% of small collaborative firms collaborating for production/commercialization-related purposes	68.1%	79.9%	76.1%	72.4%
Per-small collaborative firm production/commercialization partnerships	-	2.4	2.0	2.4
Medium firms collaborating for production/commercialization-related purposes	29	8	12	13
% of medium collaborative firms collaborating for production/commercialization-related purposes	22.8%	10.0%	11.7%	12.5%
Per-medium collaborative firm production/commercialization partnerships	-	3.6	8.4	2.3
Large firms collaborating for production/commercialization-related purposes	12	8	13	15
% of large collaborative firms collaborating for production/commercialization-related purposes	9.0%	10.1%	12.2%	15.0%
Per-large collaborative firm production/commercialization partnerships	-	6.5	2.6	2.6

Table 3.8 also reports disaggregate by-firm-size data concerning production or commercialization-related collaborative arrangements. In this case, the number of small firms collaborating for this reason dropped between 1999 and 2001 form 87 to 59; nevertheless, the proportion of small firms that are involved in at least one production/commercialization-related alliance increased form 68.1% to 79.9%. This is mainly due to the change in medium firms collaborating for production and commercialization, which were 29 in 1999 and just 8 in 2001. Besides these details, however, we can draw the same conclusions as above: no disproportion exists between this class of alliances and the overall collaboration proportions among firms of different size. Further insight comes from analyzing the average number of per-firm collaborative arrangements related to production/commercialization. Concerning small firms, we observe that they are involved, on average, in fewer partnerships for production or commercialization than for knowledge, and these values (2.4 in 2001, 2.0 in 2003 and 2.4 in 2005) are quite homogeneous over time. A consistent drop in the average number of alliances is observed again for large firms: they were involved in 6.5 production/commercialization-related alliances each in 2001, and 2.6 in 2003 and 2005. However, the most interesting result concerns medium firms: in 2001 they set on average 3.6 production/commercialization-related partnerships, 8.4 in 2003 and 2.3 in 2005. Recall how the average number of knowledge-related partnerships for this class of firms has been constantly growing over time, while for manufacturing and commercialization the drop between 2003 and 2005 is important. Further work would be required to deepen the understanding on this issue: maybe medium firms are, on average, internalizing production capacity, and do not need anymore to form this type of partnerships.

We want now to further focus on knowledge-related and manufacturing or commercialization-related alliances, further disaggregating by reason. Alliances formed to perform R&D, for regulatory affairs, to gain access to patents (and, more in general, to IP rights) and to gain access to external knowledge and skill are analyzed, along with partnerships for production and manufacturing, to gain access to markets and distribution channels, to

gain access to capital and to lower costs. Table 3.9 shows that the vast majority of knowledge-related partnerships are formed to perform R&D: 83.7% (763) in 2001 and 74.2% (835) in 2005. Moreover, the number of knowledge-related partnerships accounts for 79.7% of the total number of alliances formed in 2001, 73.4% in 2003 and 78.9% in 2005. This confirms the general understanding of the biotechnology sector, and shows that in this sector innovation comes from a collective dimension. Firms do not have the necessary skills and competences to invent, develop, test, produce and commercialize a new product of process on their own, and need to form alliances with other agents in order to focus on their core competences. Collaboration is therefore a win-win strategy, and each partner contributes with its expertise to a certain stage of the innovation process.

Table 3.9 : Distribution of alliances by reason, 2001 to 2005.

Reason	2001	2003	2005
R&D	763	-	835
Regulatory affairs	36	-	68
Access to patents	54	-	37
Access to IP rights	59	-	55
Access to knowledge/skill	-	-	132
Total knowledge-related alliances	912	757	1126
Production/manufacturing	91	-	70
Access to market/distribution channels	93	-	86
Access to capital	35	-	32
Lower costs	-	-	53
Production/commercialization-related	-	301	-
Total production/commercialization – related alliances	219	301	241

The most interesting result concerns the comparison between knowledge-related and production/manufacturing-related partnerships, and reinforces the results we found for the number of firms involved in this type of alliances. Knowledge-related collaboration is more important within the biotechnology sector. In 2001, 219 alliances were formed for production/commercialization, which represents 19.1% of the total number of partnerships of that year; in 2003 this proportion was 29.2%, and in 2005 16.9%.

Tables C.10 to C.23 in Annex C provide further insight on the geography of knowledge-related and production/commercialization collaboration, distinguishing by firm size. In general, small firms are more likely to collaborate in order to gain access to new external knowledge and skill: in every survey, the percentage of small collaborative firms that set at least one knowledge-related partnership is around 70%. These proportions are quite homogeneous among provinces (especially between Québec, Ontario and British Columbia), with an exception for 2005, when 92% of collaborative firms having at least one knowledge-related partnership in British Columbia were small. It is worth noting that this fact does not necessarily mean that in this province small firms collaborate more than in the other provinces; it rather means that among all collaborating-for-knowledge firms, almost the totality were small.

3.4 SUMMARY

In this chapter, we have provided a description of the collaboration among the different agents of the biotechnology system of innovation, distinguishing between partnerships formed with a firm or with a public institution, and between partnerships aiming at gaining access to external knowledge or for production and commercialization. This way, we have provided the basis for the interpretation of the results of the regression models, discussed in the following chapter.

The analysis began considering the constant growth in the number of firms involved in biotechnology from 1999 to 2005; most importantly, the proportion of firms with less than 50 employees (small firms) has been growing from 2001 to 2005. However, among firms operating within the biotechnology sector, those involved in at least one collaborative arrangement has been lowering: from 62.5 % in 1999 to 52.6% in 2005. Although this negative trend, we detected that the number of firms collaborating with another firms is constantly growing, reinforcing the findings we presented in chapter 1. In particular, the vast majority of inter-firm collaboration exists between biotechnology firms. From the point of view of the reasons leading firms to participate in a partnership, access to external knowledge is the most important, further stressing the knowledge-intensive nature of the biotechnology industry, and the collective dimension charactering the pattern leading to innovation. Moreover, and especially for small firms, collaboration aimed at gaining access to external knowledge is characterized by a higher intensity compared to collaboration for production and commercialization. This fact sheds light, once again, on the prominent role that the necessity of knowledge exchange and transfer play in the biotechnology industry.

CHAPTER 4

RESULTS - REGRESSION MODELS

In this chapter, we examine the effects that the factors introduced in Chapter 2 have on the firm's collaborative behavior. Distinguishing between small, medium and large firms, we test different models using different dependent variables, in order to get a deeper understanding on the collaboration with private firms or public institutions aimed at performing R&D, production or commercialization. In the following, we introduce the results of the logit analysis; then, results for the dependent variable `ec` (which indicates if a firm is involved in at least one collaborative arrangement) are presented, and the effect of each factor is discussed. Finally, we analyze the results of the logit analysis using different dependent variables: `ecc` (knowledge-related collaboration) and `ecpc` (production/commercialization collaboration), `ecepri` (collaboration with another firm) and `ecipub` (collaboration with a private institution).

4.1 LOGIT ANALYSIS

We examine the differences among the surveys of the propensity to collaborate by comparing the coefficients of the split models for (1) small firms, (2) small and medium firms and (3) small, medium and large firms. This choice is justified by considering that the subsamples including medium firms only and large firms only have too few observations, and it would not be possible to perform analysis on them. Considerations on the differences between the significance and the quality of fit of the models tested on the subsamples lead us to the decision to exclude from our study the models fitted using the whole sample. In the following, we justify this decision by explaining that large firms can be considered as outliers, introducing heavy distortion in the estimates. This fact raises an issue: if large firms are to be excluded from part of the analysis, how can we

get insight on the collaborative behavior of them? Actually, through an econometric analysis it is not possible to achieve a satisfying level of detail, as a quite large sample is required if a satisfactory precision in the estimates is to be obtained. This is the main limitation of statistics-based methods: while they provide quantitative and robust results, they require some conditions to hold and some hypotheses not to be violated. The alternative is an in-the-field study through interviews and qualitative analyses, whose results do not have the same quantitative significance, but have the advantage of taking into account the cause-effect relationships of a larger number of factors, without being limited by an exiguous number of observations.

Before we give the details of the results, we provide a bird's eye view on the overall significance of the models. We do not report all the tables in the text: a complete list of the most interesting coefficients, standard errors, goodness of fit measures and overall significance are gathered in Annex D. The coefficients of correlation between all the independent variables are reported in Annex F. At a first glance, we notice that the models fitted using data of the 2003 and the 2005 surveys are characterized by a higher overall significance, measured through the F-test. In addition, for these two surveys we notice a higher level of homogeneity among the single significant factors, which makes us more confident about the quality of the results. Conversely, estimates on the 1999 and 2001 surveys do not provide strong evidence on the collaborative behavior within the biotechnology industry; one possible hypothesis to explain this fact could be that only after 2001 the business model of the biotechnology industry began to be understood and internalized by the firms, as suggested by the aforementioned Ernst&Young (2007) biotechnology report.

The measure of fit (McKelvey and Zavoina R^2) is, in general, high when the models are fitted on subsamples including small firms and small and medium firms; conversely, it is quite low when the analysis is performed on the whole sample.

However, we must be careful in considering this scalar measure of fit as a complete and exhaustive indicator of the quality of the model. Recalling the definition of this particu-

lar pseudo-R^2, its value tends towards 1 when the variance of the error is negligible compared to the estimated variance of the latent variable. In other words, a high value of MZ-R^2 can be caused by both a decrease in the variance of the error or by an increase in the variance of the latent variable. However, during the stages of model improving, we witnessed an increase in MZ-R^2 after every step, which makes us more confident about the good quality of the estimates. Finally, it is necessary to remark that, unfortunately, MZ-R^2 represents the only measure of fit we have to test the models. However, our main concern is to assess the effect of each parameter on the propensity of being involved in at least one collaborative arrangement, rather than in the prediction of the probability to collaborate, given a certain set of values for the parameters. This approach will allow us to partially disregard the measure of fit, focusing rather on the overall significance of the model and on the significance of each parameter.

4.1.1 OVERALL PROPENSITY TO COLLABORATE

We first consider the dependent variable `ec`, which takes the value 1 if the firm is involved in at least one collaborative arrangement, 0 otherwise. As anticipated above, when the models are fitted using the subsample including small firms only and small and medium firms, they provide the most interesting, consistent and coherent results, especially for the 2003 and 2005 surveys. We attribute this fact to the higher level of homogeneity among small and medium firms: in this respect, large firms can be indeed considered as outliers. Even though the number of large firms is much smaller in number, their effect on the estimates is detrimental. Therefore, analyzing the collaborative behavior of large firms through an econometric analysis is not possible, and qualitative methods need to be employed.

4.1.1.1 SIZE

Amongst small biotechnology-dedicated firms, size (measured through the total number of employees in Canada) appears to be the most relevant characteristic in determining

the propensity to collaborate. Its effect is positive and always significant, at least at a 5% level in almost all cases. When size is measured in terms of total revenues, it is never significant (second model), even though variables `e` and `rt_0` are correlated (in 2005, the coefficient of correlation was higher than 70%, while it attains the value of 96% in 2003). In addition, when size is measured through total revenues, the scalar measure of fit consistently decreases, particularly using the subsample including both small and medium firms (15.3% in 2005 and 24.3% in 2003). However, due to the effect of other factors, the overall significance of these models remains high (at least at 5%). Table 4.1 shows that the effect of the number of the employees among small and medium firms is always significant and its effect is positive. In particular, if we take account of the 2003 and 2005 survey only, it is possible to draw a preliminary conclusion on the relationship between size and propensity to collaborate: a small biotechnology firm is more likely to be involved in at least one partnership as its size grows. Recall here that a firms is defined as "small" when the total number of employees is smaller than 50.

Although in the second model we take the log of the total revenues (variable `lrt_0`) in order to smooth out the discrepancies that a size effect may cause, the measure of fit in this case remains very low. Therefore, we conclude that this model must be excluded from our analysis, as the variable `lrt_0` is not significant in determining a firm's orientation towards collaboration.

4.1.1.2 ORIENTATION TOWARDS BIOTECHNOLOGY

The propensity to collaborate is positively affected also by the orientation towards biotechnology, particularly when it is measured through the percentage of employees having biotechnology-related responsibilities. In addition, the significance of the coefficient of this factor increases when the sample including both small and medium firms is considered. We can get an understating of this fact considering the average percentage of employees having biotechnology-related responsibilities across firms of different size. The 2003 survey reveals that 69.8% of the personnel employed in a small biotechnology firm had biotechnology-related responsibility, while for medium firms this proportion

was 58.4% (in 2005 the percentages were 76.8% and 59.2% respectively). This reflects the fact that small firms are biotechnology-dedicated, and all their resources are consecrated to their core competence. The homogeneity of this characteristic among small firms has the effects of decreasing the significance of the coefficient of the variable `ebper`. Conversely, medium firms can choose to be more or less biotechnology-oriented, as they have the necessary resources to perform complementary activities; in this case, the role played by the orientation towards biotechnology is decisive in determining the collaborative behavior. The analysis of the mean value assumed by the variable `ebper` considering the subpopulation of large firms (7.0% in 2003 and 7.2%) sheds light on the lower quality of fit and significance of the models fitted considering the whole population. Consider, for example, the 2005 survey: the percentages of small, medium and large collaborative firms were equal to, respectively, 50.9%, 56.6% and 59.6%. Clearly, when large firms are included in the sample, a determinant discontinuity is introduced, leading to poor quality of model fit. The same considerations hold considering the variable `e` (total employees): small firms had, in 2005, 14.6 employees each, medium firms 74.2 and large firms had 1440 employees each. This disproportion also explains for the poor quality of the models fitted using the whole sample.

4.1.1.3 SPIN-OFFS

In all models, the fact of being generated through a spin-off has a strong and positive effect on the likelihood of being involved in at least one collaborative arrangement. In the fourth model, for example (as we will see, this can be considered as the most meaningful model), the coefficient of this variable is always significant, at least at the 5% level. This effect is to be expected, as a spin-off usually maintains close links and relationships with the originating institution (enterprise, university, hospital, laboratory, etc...). Through a spin-off, a group of researchers or a department (or a portion of it) of a larger firm/institution formally acquires total independence, but remains closely linked to the originating agent. In Chapter 3, we noticed that firms created through a spin off are mostly generated from a public institution (particularly hospitals and universities). In the

following, we further investigate on this issue, by evaluating the significance of this characteristic on the propensity to collaborate distinguishing on the type of partner (public institution or another firm).

4.1.1.4 CAPITAL FINANCING

In the 2005 survey, when considering the subsample of small firms only, variables `reusk` and `lfkreun` are significant at the 10% level in explaining the probability of being involved in at least one collaborative arrangement. Conversely, when the seventh model is fitted on the subsample including both small and medium firms, the coefficients are not significant. This fact provides insight on the dynamics of capital financing for small biotechnology firms. More and more, they need to resort to alliances in order to obtain capital to carry on R&D projects: as many authors note, partnerships are an important source of capital for small firms. However, correlation between these two variables is very strong (99% for both 2003 and 2005) and we must be careful in drawing any result. If we do not take the log of the variable `fkreun`, the value of its coefficient would be extremely low, compromising the results as well[25]. Nevertheless, the correlation between `lfkreun` and `reusk` is mainly due to the zeros. When `reusk` is zero, `fkreun` is zero as well, while if `fkreun` is greater than zero, `reusk` is one. This particular structure of the variables makes the results less unreliable.

4.1.1.5 BIOTECHNOLOGY PRODUCTS

Results show, in particular for the 2003 survey, that firms developing products (rather than processes) requiring the use of biotechnology are more likely to be involved in at least one collaborative arrangement. In the fourth model, the coefficient of the variable `dprod` is significant at the 5% level for small firms and at the 10% level for small and medium firms. Conversely, the coefficient of the variable `dproc` is not significant in

[25] We take the log of capital-related variables in order to take into account the scale effect. This way, as we expect, when we use the raw value of the variable, the quality of fit drops significantly.

2003. In 2005, neither the variable dprod nor dproc are significant: developing biotechnology products rather than processes does no longer shape the likelihood to collaborate. The 2005 survey reveals instead an influence at the 10% level of the number of biotechnology products at the pre-market stage (variable nprodrc) on the probability to collaborate when the subsample including small biotechnology firms only is considered. Conversely, when we employ the subsample of small and medium firms, the variable nprodpm (number of biotechnology products at production stage) is significant at the 10% level. Globally, these results do not provide meaningful insight in explaining collaborative behavior, as the level of significance is lower than the other determinant variables.

4.1.1.6 OTHER CHARACTERISTICS

The models reveal no significant effect of the other characteristics we tested on the probability of being involved in at least one collaborative arrangement. Orientation towards exporting, along with public support, is not significant in each model, considering both small firms only and small and medium firms. In addition, we see that it makes no difference to develop processes rather than requiring the use of biotechnology. Comparing these results with the influence of the orientation towards biotechnology emphasizes how collaboration is a characteristic that depends on whether a firm employs biotechnology, and not on how it employs it. In addition, the other measures of biotechnology orientation do not show any relevant effect on the propensity to collaborate.

Means for Intellectual Property protection, measured through the number of obtained and licensed patents (variables nbe and nba for the number of obtained and pending patents, respectively) or, more in general, IP rights (variables ndroitoec and ndroitaec respectively) do not seem to have any impact on the propensity of being involved in collaborative arrangements, with an exception. In the 2005 survey, the coefficient of the variable ndroitoec is significant at the 10% level for small firms and at the 5% level for small and medium firms, while variables nbe and nba are never sig-

nificant. This means that access to external IP and collaboration are correlated, but this consideration is not limited to the case of gaining access to others' patents. Further research is required to assess the cause-effect relationships and the various facets of collaborative behavior aimed at obtaining access to external intellectual property, and to understand the overall benefits on a firm's productivity.

It is interesting to note that, in contrast with what has been found by other authors (see Chapter 1), the percentage of revenues coming from exportation of biotechnology products is not determinant in explaining the probability to collaborate within the Canadian biotechnology system of innovation.

Finally, the influence of the control for the environment is significant at the 1% level in almost all the models tested using data from the 2005 survey, but it is never significant in 2003.

4.1.2 MODELS

Among all the models we tested, we exclude from the analysis the second (due to the non-meaningfulness of the variables `lrt_0` and `rbper0` in explaining collaborative behavior). In fact, by replacing the variables `e` and `ebper` with `lrt_0` and `rbper0`, we would exclude from the analysis the most important and significant factors explaining the differences in the propensity to collaborate among small and medium biotechnology firms. The sixth model could potentially provide interesting results, as the correlation between the variables `lrdt_0` and `rdbper0` is equal to 18% in the 2003 survey and to 46% in the 2005 survey. Although this value is higher than the threshold of 40%, can still be considered acceptable. However, these variables are not found to be significant (with an exception for 2003 when employing the subsample including small and medium firms, where `rdbper0` is significant at the 10% level, but the value of the coefficient is extremely low).

The above considerations on the effect of IP protection lead us to consider the fourth model as the most significant. No critical correlation between the independent variables exists, both for the 2003 and the 2005 surveys; the overall significance of the model is

very high (at the 1% level) for both the survey and for both the subsamples we considered. These conclusions hold just for the 2003 and 2005 surveys. As we mentioned before, results from the 1999 and 2001 surveys reflect the evolution of the biotechnology industry business model. In those years, collaboration among biotechnology firms was in fact perceived just as a strategy among others, while nowadays its value and its role in the process leading to innovation is clearer. For this reason, we argue that in 1999 and 2001 the characteristics of the collaborative firms were more heterogeneous, explaining for the results of these two surveys.

Table 4.1 : Fourth logit model, dependent variable ec, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0691***	0.0145	0.0314**	0.0136	0.0119***	0.0045	0.0168***	0.0053
ebper	0.0148**	0.0058	0.0070	0.0050	0.0112**	0.0046	0.0087**	0.0042
dspin	0.8426***	0.3069	1.0797***	0.2941	0.7245***	0.2737	0.8842***	0.2687
dprod	0.9151**	0.4478	0.4483	0.5001	0.6803*	0.4117	0.3235	0.4448
dproc	0.4525	0.3656	0.2978	0.3126	0.4627	0.3223	0.1773	0.3015
ndroitaec	0.2357	0.5371	0.0175	0.0654	0.3625	0.3222	0.0181	0.0669
ndroitoec	0.1619	0.1932	0.9071*	0.4934	0.2694	0.2024	0.8571**	0.4100
dsubs	-	-	0.3445	0.4216	-	-	0.1708	0.3776
rexpperb_0	-0.0017	0.0039	0.0000	0.0038	0.0013	0.0037	-0.0010	0.0034
sourcekgov	0.0106	0.0068	0.0160	0.0141	0.0097	0.0064	0.0178	0.0114
lsourcekgovtot	-0.0535	0.0415	-0.0366	0.0714	-0.0390	0.0361	-0.0501	0.0524
nbppopprov	-3.8075	7.2901	-17.4511***	6.7215	3.4694	6.5475	-16.5401***	6.1529
Intercept	-3.4480***	0.7976	-1.2828	0.8973	-2.7024***	0.6823	-0.8365	0.7830
Observations	227		266		271		318	
Weighted obs.	352		397		428		480	
F-statstic	3.48***		2.59***		2.72***		2.82***	
R^2_{MZ}	99.9%		99.6%		97.5%		98.8%	

Table 4.1 in the preceding page shows in detail the results of fitting the fourth logit model for the 2003 and 2005 surveys, for subsamples including small firms only and small and medium firms. Firm size, orientation towards biotechnology and being a spin-off play a determinant role in influencing collaborative behavior. In particular, in 2005, the effect of biotechnology orientation is significant just when the subsample including both small and medium firms is considered, emphasizing how small firms are, in general, highly biotechnology-dedicated, as it has been explained above. The number of obtained IP rights is significant at the 10% level for small firms and at the 5% level for small and medium firms in 2005. As we anticipated above, the econometric model we employ assumes that the number of licensed or obtained intellectual property rights has an influence on the propensity to collaborate. In other words, the cause-effect relationship goes from the independent variables to the dependent variable. Yet, this is just a conceptual model, a blurry picture of the reality, and we must be careful in drawing any conclusion involving the dynamics of the process characterizing the pattern of innovation in the biotechnology industry. We can nevertheless assert that the propensity towards collaboration and the number of obtained intellectual property rights are related, but we cannot tell whether the former influences the latter, whether the latter has an impact on the former or whether the two aspects are linked in a more complex way and integrated in intricate system. Table 4.2 reports the details of the estimation of the seventh model. For both 2003 and 2005 surveys, and for both subsamples, the overall significance of the models is at the 1% level, and the measure of fit is high. As anticipated, the importance of the variables reusk and lfkreun is negligible for the 2003 survey, while in 2005 it is significant at the 10% level only for small firms. This can be interpreted as a change in the strategy of small firms in trying to gain access to capital. However, as we anticipated, we must be particularly careful in assessing the importance of these characteristics, as the two variables are highly correlated, due to they particular structure. As it has been said, the biotechnology industry is changing, and the business models are emerging and settling, improving and optimizing the processes leading to innovation. In

this perspective, collaboration is perceived as a powerful means for small biotechnology-dedicated firms to raise capital in order to survive.

Table 4.2 : Seventh logit model, dependent variable ec, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0664***	0.0152	0.0355**	0.0140	0.0109**	0.0047	0.0172***	0.0055
ebper	0.0150**	0.0054	0.0095*	0.0050	0.0113**	0.0044	0.0101**	0.0042
dspin	0.8072**	0.3111	1.0289***	0.2897	0.6670**	0.2739	0.8421***	0.2662
reusk	0.3706	2.2869	4.6285*	2.5188	0.0954	1.9102	3.1231	2.1188
lfkreun	0.0357	0.1607	-0.3074*	0.1723	-	-	-0.1991	0.1418
dsubs	-	-	0.2979	0.4335	0.0513	0.1315	0.1159	0.3733
rexpperb_0	-0.0005	0.0040	0.0006	0.0036	0.0023	0.0036	-0.0005	0.0032
sourcekgov	0.0047	0.0062	0.0026	0.0068	0.0040	0.0055	0.0023	0.0064
nbppopprov	-3.1157	7.2240	-17.1758***	6.5633	3.3268	6.5538	-16.5888***	6.0136
Intercept	-2.6202***	0.6635	-0.8377	0.7020	-1.9769***	0.5481	-0.4536	0.5921
Observations	227		266		271		318	
Weighted obs.	352		397		428		480	
F-statstic	4.78***		3.36***		3.58***		3.49***	
R^2_{MZ}	99.9%		99.7%		97.0%		98.8%	

4.1.3 COLLABORATION AND PARTNERS

In this subsection, we investigate on firm-specific characteristics affecting the propensity to collaborate with another firm and with a public institution. The comparison between the models with respect to the dependent variable employed (`ecipub` and `ecepri`) sheds light on the different firm-specific characteristics influencing the propensity to collaborate with a different partner. In particular, we take into account the fourth model, which we deem to be the most meaningful.

When we consider collaborative behavior with a public partner, firm size is determinant at the 5% level only for the 2005 survey, while in 2003 is never significant, as Table 4.4 shows. Conversely, size is important in influencing the propensity to collaborate with another firm for both subsamples and for both 2003 and 2005 surveys, as Table 4.3 shows. In addition, an important effect of the orientation towards biotechnology is found in 2003 only and for both subsamples, while being a spin-off affects positively the probability of collaborating only in 2005. This finding suggests that an evolution has taken place in collaborative behavior with a public partner.

Biotechnology orientation is significant at the 5% level in 2005 for small and medium firms in describing the probability to collaborate with another firm, along with biotechnology orientation and with being a spin-off. As in the case of the overall propensity to collaborate (variable `ec`), orientation towards biotechnology is not significant when the subsample including small firms only is taken into account. It is worth noting that, in contrast with the case of collaboration with a public institution, the density of biotechnology-related patents in each province (control for the environment, variable `nbppop-prov`) has a positive effect on the propensity to form alliances with another firm. However, we must be cautious in drawing conclusion in this last case, as the overall significance of the models is quite low, except for small and medium firms in 2005; the scalar measure of fit is nevertheless always high. The conclusions we can draw from fitting the model using these two dependent variables outline the differences in the characteristics affecting collaborative behavior with private or public partners. The most insightful result concerns orientation towards biotechnology, which in the most recent survey posi-

tively affects the propensity to collaborate with another firm, but not with a public institution, when we consider the subsample including both small and medium firms. We could explain this fact by considering that alliances with public institutions (universities, hospitals and governmental labs) are usually related to knowledge and research project. This way, a firm can "outsource" R&D in order to increase its orientation towards biotechnology. Conversely, partnerships with other firms can take various forms. For example, they may aim at increasing production capacity, at testing or commercializing a new biotechnology product/process or at improving R&D performance.

Table 4.3 : Fourth logit model, dependent variable `ecepri`, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0340**	0.0136	0.0232*	0.0126	0.0110**	0.0044	0.0116**	0.0046
ebper	0.0048	0.0060	0.0079	0.0053	0.0065	0.0055	0.0115**	0.0047
dspin	0.3006	0.3403	0.6067**	0.2927	0.5317*	0.3097	0.6190**	0.2717
dprod	0.1315	0.5154	0.2056	0.4762	0.3666	0.5126	0.0671	0.4334
dproc	0.5161	0.4551	-0.0222	0.3078	0.4610	0.3962	-0.0641	0.3048
ndroitaec	-0.6380	0.7554	0.0505	0.0744	-0.1842	0.3214	0.0577	0.0772
ndroitoec	0.0240	0.2101	0.2134	0.3390	0.0391	0.2048	0.2381	0.3265
dsubs	-	-	-0.0692	0.4191	-	-	0.0696	0.3939
rexpperb_0	-0.0058	0.0050	-0.0010	0.0039	-0.0028	0.0042	-0.0034	0.0035
sourcekgov	0.0003	0.0080	0.0103	0.0127	0.0030	0.0073	0.0064	0.0102
lsourcekgovtot	-0.0067	0.0462	-0.0476	0.0646	0.0117	0.0376	-0.0243	0.0464
nbppopprov	3.5942	8.5158	-14.7389**	6.5993	6.2446	7.8462	-15.1533**	6.0479
Intercept	-2.7926***	0.8027	-0.9313	0.8420	-3.1657***	0.8229	-1.0006	0.7820
Observations	227		266		271		318	
Weighted obs.	352		397		428		480	
F-statstic	1.20		1.29		1.34		1.87**	
R^2_{MZ}	99.7%		99.4%		97.1%		97.4%	

Table 4.4 : Fourth logit model, dependent variable `ecipub`, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	-0.0047	0.0148	0.0265**	0.0134	0.0069	0.0059	0.0124**	0.0054
ebper	0.0180**	0.0076	0.0042	0.0059	0.0267***	0.0077	0.0053	0.0045
dspin	0.2458	0.3649	1.1555***	0.3041	0.0940	0.3343	0.7479***	0.2713
dprod	0.9027	0.6070	0.3949	0.5566	1.0039*	0.5845	0.5057	0.4756
dproc	1.1161*	0.5744	0.1948	0.3360	1.0871**	0.5291	0.0618	0.3100
ndroitaec	-	-	0.0135	0.0750	-0.0341	0.3399	-0.0156	0.0802
ndroitoec	0.2721	0.2500	0.8261*	0.4712	0.2766	0.2363	0.6620*	0.3722
dsubs	-	-	0.2634	0.4512	-	-	0.0838	0.3670
rexpperb_0	-0.0062	0.0057	0.0011	0.0039	-0.0029	0.0049	0.0013	0.0034
sourcekgov	0.0105	0.0079	0.0073	0.0115	0.0125	0.0077	0.0056	0.0105
lsourcekgovtot	-	-	0.0084	0.0538	-0.0805*	0.0454	0.0071	0.0469
nbppopprov	8.0277	10.2731	-11.7135	7.3286	8.3531	9.2337	-9.2257	6.4044
Intercept	-5.1311***	1.1761	-2.0649	1.0078	-6.2093***	1.2311	-1.7269**	0.8476
Observations	207		266		271		318	
Weighted obs.	321		397		428		480	
F-statstic	1.95**		2.45***		2.45***		1.61*	
R^2_{MZ}	88.6%		99.5%		92.9%		97.8%	

4.1.4 TYPE OF COLLABORATION: KNOWLEDGE AND PRODUCTION/COMMERCIALIZATION

We now analyze the factors affecting the propensity to collaborate for knowledge-related reasons (dependent variable `ecc`) and for production/commercialization reasons (dependent variable `ecpc`). As Table 4.6 reports, there is no strong correlation between firm-specific characteristics and the propensity to collaborate for production or commercialization of biotechnology products. An exception exists for the subsample including small and medium firms in the 2003 survey, where the fact of being created through a spin-off is significant at the 1% level, and the model is significant at the 5% level. No other model detects a strong correlation between firm characteristics and collaboration for production and commercialization. Insight that is more interesting is instead provided by analyzing the effect of the variables on the propensity of being involved in collaboration related to knowledge, as Table 4.5 shows. In this case, firm size plays an important role, positively affecting collaborative behavior aimed at acquiring external knowledge. Recalling the results from summary statistics, the 96% of collaborative firms had at least one partnership related to knowledge; among these, 63.2% had less than 50 employees, and 17.7% had between 50 and 150 employees. The percentage of small and medium firms that collaborate for knowledge is therefore high; among small firms, those with a higher number of employees are more likely to be involved in partnerships aimed at acquiring and at transferring knowledge.

Table 4.5 : Fourth logit model, dependent variable ecc, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0520***	0.0144	0.0354***	0.0137	0.0085	0.0042	0.0195***	0.0055
ebper	0.0058	0.0059	0.0073	0.0050	0.0086	0.0049	0.0089**	0.0042
dspin	0.3571	0.3323	1.0206***	0.2924	0.3221	0.2871	0.8439***	0.2669
dprod	1.1512**	0.6095	0.3621	0.4876	0.7640	0.5144	0.2759	0.4433
dproc	0.4568	0.4022	0.2931	0.3145	0.6501	0.3701	0.1918	0.3026
ndroitaec	-1.0448	1.2991	0.0211	0.0670	0.1726	0.2563	0.0219	0.0686
ndroitoec	0.0270	0.2130	0.9771*	0.5203	0.1398	0.1963	0.9215**	0.4242
dsubs	-	-	0.7277*	0.4390	-	-	0.4353	0.3775
rexpperb_0	-0.0064	0.0044	-0.0029	0.0038	-0.0025	0.0038	-0.0033	0.0034
sourcekgov	-0.0054	0.0071	0.0162	0.0141	-0.0033	0.0066	0.0173	0.0112
lsourcekgovtot	-0.0226	0.0432	-0.0374	0.0719	0.0030	0.0348	-0.0483	0.0517
nbppopprov	16.2609*	8.5649	-14.0542**	6.9347	19.8739	7.5921	-13.7726**	6.2635
Intercept	-4.2770***	1.0097	-1.9192**	0.9312	-4.0696	0.9018	-1.3582*	0.8078
Observations	227		266		271		318	
Weighted obs.	352		397		428		480	
F-statstic	2.60***		2.62***		1.94**		2.88***	
R^2_{MZ}	99.9%		99.7%		95.3%		99.1%	

Table 4.6 : Fourth logit model, dependent variable ecpc, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff	Std. Err.
e	0.0205	0.0142	0.0030	0.0136	0.0050	0.0054	0.0020	0.0062
ebper	0.0083	0.0065	-0.0021	0.0064	0.0117*	0.0063	-0.0034	0.0052
dspin	0.7336**	0.3368	0.5843	0.3587	0.8696***	0.3220	0.5324	0.3350
dprod	0.0494	0.4795	0.4228	0.6369	0.1266	0.4684	0.3264	0.5655
dproc	0.2958	0.4293	0.1083	0.3732	0.2320	0.4031	0.0790	0.3654
ndroitaec	0.4919	0.4928	0.0122	0.0694	0.0404	0.3388	0.0144	0.0717
ndroitoec	0.1001	0.2208	0.6782**	0.3325	-	-	0.6937**	0.3406
dsubs	-	-	-0.2494	0.4908	0.1285	0.2124	-0.5600	0.4124
rexpperb_0	0.0024	0.0045	0.0032	0.0045	0.0027	0.0042	0.0029	0.0040
sourcekgov	0.0081	0.0074	-0.0067	0.0156	0.0094	0.0069	0.0076	0.0114
lsourcekgovtot	-0.0009	0.0393	0.0818	0.0731	0.0229	0.0352	0.0075	0.0509
nbppopprov	-3.7202	7.5185	-10.4622	7.9988	-1.3422	7.5348	-11.7539	7.3223
Intercept	-2.7613***	0.7681	-1.4352	1.1163	-3.1661***	0.7911	-0.8594	0.9824
Observations	227		266		271		318	
Weighted obs.	352		397		428		480	
F-statstic	1.12		1.05		2.07**		1.01	
R^2_{MZ}	99.1%		74.9%		87.0%		60.6%	

4.2 NEGATIVE BINOMIAL ANALYSIS

This part of the analysis aims at providing an understanding of the effects that firm-specific characteristics have on the number of collaborative arrangements a biotechnology enterprise is involved in. We described all the models in Chapter 2; they use the same independent variables as logit models. Unfortunately, no scalar measure of fit is provided for negative binomial models, ad the only way to test the goodness of fit is the analysis of the residuals, which is not performed. However, as mentioned in Annex G, a scalar measure of fit is rarely significant in the case of NBR. The analysis of correlation among independent variables, as we have seen for logit models, will allow us to exclude some of the models from the analysis. Moreover, it is interesting to compare in particular the fourth and the seventh model, to detect whether or not the same factors affecting the propensity to collaborate affect also the number of collaborative arrangements in which a firm is involved. In the text, we report only the essential results; more details are provided in Annex E.

4.2.1 OVERALL INTENSITY OF COLLABORATION

Table 4.7 shows that firm size is determinant in predicting the number of collaborative arrangements in which an enterprise is involved. The effect of orientation towards biotechnology shows a singular trend in affecting the intensity of collaboration: The 2003 survey reveals a strong link between this characteristic and the number of partnerships, while in 2005, this relationship is not significant for small firms, and it is significant only at the 10% level for small and medium firms.

We note that, while the variable `dspin` positively influences the propensity to collaborate, it is not significant in explaining for the number of alliances. Conversely, significance at the 5% level is detected for the variable `dprod` and not for the variable `dproc` in 2005: developing products requiring the use of biotechnologies has a positive effect on the intensity of collaboration. A new product requires new knowledge creation, and synergies of a variety of agents within the system of innovation are necessary. However,

the most intriguing result comes from the relationship between IP protection and the number of alliances, which opens up for further investigation. In 2003, a negative effect of the number of acquired IP rights (at the 5% level for small firms and at the 1% level for small and medium firms) existed, while the same variable is not significant at all in 2005. Conversely, a positive effect (significant at the 5% level for both subsamples) is detected in 2005 for the number of obtained IP rights; in 2003, this characteristic is not significant. Explaining this singular trend requires in-the-field research on the cause-effect relationships of IP protection strategies, as we noticed explaining the results of fitting the logit models.

Table E.5 in Annex E deals with the effect of firms' behavior aimed at raising capital on the number of alliances in which they are involved; neither success in gaining access to capital, nor the amount of raised capital has a significant effect in predicting the intensity of collaboration. This fact, integrated with the findings presented in the preceding subsection sheds more light on this issue, showing that successful attempts to raise capital, and the amount of capital raised do not have significant effect neither on the propensity to collaborate nor on the number of partnerships. We recall, however, that we must be cautious in drawing conclusions from this model, as variables `reusk` and `lfkreun` suffer from severe correlation.

Table 4.7 : Fourth negative binomial model, dependent variable nec, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0304***	0.0105	0.0201**	0.0089	0.0152***	0.0056	0.0134***	0.0040
ebper	0.0131***	0.0048	0.0069	0.0046	0.0172***	0.0046	0.0070*	0.0039
dspin	0.1037	0.2424	0.3992	0.2594	0.0637	0.2379	0.4092	0.2513
dprod	-0.0104	0.3878	0.8337**	0.3886	0.0905	0.4011	0.8403**	0.3365
dproc	0.5078	0.3259	0.3061	0.2758	0.7934	0.3170	0.2064	0.2414
ndroitaec	-1.2392**	0.5376	-0.0192	0.1189	-0.7980***	0.2564	-0.0245	0.1161
ndroitoec	0.1392	0.1618	0.5160**	0.2480	0.1524	0.1659	0.4813**	0.2372
dsubs	-	-	0.2608	0.3027	-	-	0.1257	0.2762
rexpperb_0	-0.0056	0.0035	-0.0050	0.0035	-0.0022	0.0038	-0.0027	0.0031
sourcekgov	0.0018	0.0051	0.0003	0.0097	0.0040	0.0052	0.0047	0.0076
lsourcekgovtot	-0.0447	0.0330	0.0289	0.0494	-0.0073	0.0295	-0.0030	0.0350
nbppopprov	1.5424	5.8702	-0.5440	4.8197	6.1652	6.0865	3.6103	4.2604
Intercept	-1.9038***	0.6590	-1.4238	0.7008	-2.7285***	0.6574	-1.4448**	0.6149
Alpha	2.6269		2.4346		2.9910		2.4838	
Observations	227		266		271		318	
Weighted obs.	352		397		428		480	
F-statstic	2.95***		2.69***		3.03***		2.93***	

4.2.2 INTENSITY OF COLLABORATION AND PARTNERS

With the exception of the subsample including small firms only for the 2005 survey, size plays a determinant role, positively affecting the number of collaborative agreements with another firm. It is worth noting that, while size is in general significant, the values of the associated coefficients are not very high, especially compared to the ones associated to IP protection behavior. In this respect, it is surprising to see how in 2003 the number of licensed patents negatively affects the number of partnerships with another firm, for both subsamples. The coefficients are significant at the 5% level (small firms) and at the 1% level (small and medium firms), and their values are quite large. In addition, whether a firm develops biotechnology products rather than processes has a positive impact on the number of collaborations, as

Table 4.8 shows. However, we must be careful drawing strong conclusions form these estimates, as the overall significance of the model is not high, with an exception for the 2003 survey, when considering the subsample including small and medium firms.

Table 4.9 shows that the model is much more significant when the variable `necipub` (number of collaborative arrangements with a public institution) is employed, and provides further insight on the role that orientation towards biotechnology plays in describing the collaborative behavior. This characteristic is significant at the 10% level when small and medium firms are considered, but remains non-significant for small firms. It is interesting to note the effect of being a spin-off, which is positive and significant (at the 1% level for small firms, and at the 5% level for small and medium firms) in 2005, is never significant in predicting the number of partnerships with another firm. Recall that this characteristic positively affects the propensity to collaborate with both private and public institutions, but affects the number of alliances with a public institution only.

The most interesting result coming from the comparison between Table 4.8 and Table 4.9 concerns the impact of whether the firms develops biotechnology products or processes (variables `dprod` and `dproc`). In 2005, firms developing biotechnology products were more likely to be involved in a greater number of partnerships with another firm.

Conversely, in 2003 firms developing processes requiring the use of biotechnology showed a greater collaborative intensity with a public institution.

Focusing on small firms in 2005, the differences between collaborative behavior with a public institution and with another firm can be outlined as follows. The size of a small firm positively affects the likelihood of being involved in a higher number of agreements with a public institution, while its effect is negligible in describing the intensity of collaboration with other firms. Small firms created through a spin-off are more likely to form a larger number of alliances with a university, hospital or government lab than non-spin-off do. The same consideration does not hold for collaboration with a private firm. Concerning the influence of the intellectual property protection means, a strong positive effect is determined by the number of obtained IP rights in the number of partnerships with a public institution, but not with another firm. Finally, a negative and significant at the 5% level relationship exists between the percentage of the revenues from exporting biotechnology products and the intensity of collaboration with another firm. This result is quite surprising; however, the value of this coefficient in extremely low.

Table 4.8 : Fourth negative binomial model, dependent variable necepri, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0313***	0.0119	0.0092	0.0105	0.0169***	0.0065	0.0112**	0.0056
ebper	0.0077	0.0056	0.0064	0.0050	0.0119**	0.0052	0.0066	0.0046
dspin	0.1909	0.3123	0.1151	0.2876	0.1794	0.2899	0.3492	0.3064
dprod	0.0121	0.4592	1.2874***	0.4940	0.1041	0.4830	1.0268**	0.4387
dproc	0.1709	0.3794	0.5060	0.3274	0.4617	0.3687	0.3683	0.3077
ndroitaec	-1.0391**	0.5744	-0.0771	0.0810	-0.7805***	0.2743	-0.0657	0.0851
ndroitoec	0.1467	0.2717	0.3053	0.3122	0.1471	0.2618	0.2869	0.3107
dsubs	-	-	-0.1457	0.3291	-	-	-0.1900	0.3518
rexpperb_0	-0.0049	0.0043	-0.0079**	0.0040	-0.0024	0.0044	-0.0037	0.0040
sourcekgov	-0.0053	0.0061	-0.0058	0.0120	-0.0043	0.0057	0.0003	0.0102
lsourcekgovtot	-0.0087	0.0331	0.0046	0.0564	0.0133	0.0303	-0.0297	0.0450
nbppopprov	2.1108	7.4365	-0.5162	5.5929	10.0289	7.4148	3.5468	5.1271
Intercept	-1.8125**	0.8395	-1.7411**	0.8217	-2.8102***	0.7940	-1.8273**	0.7606
Alpha	4.3346		3.7605		4.8113		4.1400	
Observations	227		266		271		318	
Weighted obs.	352		397		428		480	
F-statstic	1.36		1.45		1.86**		1.54	

Table 4.9 : Fourth negative binomial model, dependent variable necipub, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0195	0.0146	0.0270***	0.0104	0.0146***	0.0049	0.0179***	0.0049
ebper	0.0218***	0.0059	0.0071	0.0052	0.0293***	0.0057	0.0076*	0.0042
dspin	-0.0881	0.3118	0.7799***	0.2930	-0.0609	0.3096	0.5416**	0.2698
dprod	0.7225	0.5623	0.3842	0.4818	0.8214	0.5854	0.7420*	0.4420
dproc	1.6911***	0.5242	0.0550	0.2986	1.7189***	0.4946	0.1064	0.2782
ndroitaec	-13.1751***	0.9330	0.0381	0.1312	-0.9231***	0.3403	0.0181	0.1320
ndroitoec	0.1695	0.1499	0.7992***	0.2967	0.1673	0.1467	0.7093***	0.2568
dsubs	-	-	0.6813	0.3795	-	-	0.4493	0.3312
rexpperb_0	-0.0119**	0.0053	-0.0011	0.0037	-0.0046	0.0045	-0.0018	0.0034
sourcekgov	0.0109	0.0076	-0.0009	0.0109	0.0125	0.0076	0.0059	0.0089
lsourcekgovtot	-0.3081***	0.0704	0.0608	0.0587	-0.0291	0.0415	0.0207	0.0397
nbppopprov	-0.5146	8.4086	-0.8243	5.4139	0.5722	8.4985	3.2929	5.3562
Intercept	-4.8983***	0.9807	-2.5256***	0.8594	-5.8514***	0.9205	-2.7006***	0.8067
Alpha	3.439022		2.9316		4.3567		3.2187	
Observations	227		266		271		318	
Weighted obs.	352		397		428		480	
F-statstic	24.60***		3.44***		5.19***		2.99***	

4.2.3 INTENSITY OF COLLABORATION AND REASONS

In this subsection, we investigate the effect of firm-specific characteristics on the number of collaborative arrangements distinguishing between knowledge-related and production/commercialization-related partnerships. We notice an important difference in the significance of the models when fitted using `necc` (number of alliances related to knowledge) and `necpc` (number of alliances related to production or commercialization). In the first case, the overall significance is always very strong (at the 1% level), while in the latter it is not homogeneous, and for the 2005 survey, only one variable seems to affect the number of collaborative arrangements a firm is involved in. These results are reported in Table 4.10. Let us first consider the subsample including small firms only in the 2005 survey. Here size, orientation towards biotechnology, product development and number of obtained IP rights positively affect the intensity of collaboration; this trend is even more evident in the subsample including both small and medium firms. While a spin-off has a higher propensity to collaborate for knowledge (see Table 4.5), it is not likely to be involved in a higher number of collaborative arrangements of this type. Conversely, the number of obtained IP rights positively affects both the propensity towards collaboration and its intensity. Moreover, orientation towards exportation negatively affects the intensity of collaboration (significant at the 5% level); however, the value of the associated coefficient is extremely low. In 2003, size is significant at the 1% level for both subsamples, while orientation towards biotechnology is significant (at the 5% level) only for the subsample including small and medium firms. Surprisingly, results from this survey reveal the effect of the variables `dprod` and `dproc` has a contrasting effect compared to the 2005 survey. Here, firms developing processes requiring the use of biotechnology are more likely to be involved in a higher number of collaborative arrangements related to knowledge. This interesting result requires further research in order to assess the cause-effect relationships between level of innovativeness and intensity of collaboration, and on the various facets of the evolution of the biotechnology industry. Understanding this change in the collaborative behavior will shed light

on the dynamics and the improvements of the business models in this fast-evolving industry.

Concerning the intensity of collaboration related to production and commercialization, Table 4.11 shows that size has not the same important as for knowledge-related collaboration. Singularly, in 2005 no factor but the number of obtained IP rights seems to have a determinant effect on the intensity of collaboration, while in 2003 the overall significance of the model was at the 5% level for small firms and at the 1% level for small and medium firms. We also remark that the effect of developing processes requiring the use of biotechnology has a stronger effect than developing biotechnology products (for the 2003 survey only).

Table 4.10 : Fourth negative binomial model, dependent variable necc, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0309***	0.0095	0.0189**	0.0088	0.0125***	0.0047	0.0160***	0.0044
ebper	0.0079	0.0049	0.0101**	0.0048	0.0111**	0.0047	0.0100**	0.0040
dspin	0.2509	0.2404	0.2911	0.2714	0.2581	0.2306	0.3184	0.2615
dprod	0.8096	0.4964	0.7380*	0.3810	0.5393	0.4279	0.7990**	0.3309
dproc	0.5456*	0.3074	0.4040	0.2824	0.7651**	0.3117	0.3458	0.2463
ndroitaec	-1.3080	0.9698	-0.0347	0.0916	0.2228	0.1598	-0.0477	0.0865
ndroitoec	-0.1746	0.1700	0.5335**	0.2400	-0.0944	0.1367	0.4801**	0.2310
dsubs	-	-	0.3429	0.3409	-	-	0.3058	0.2943
rexpperb_0	-0.0094***	0.0034	-0.0070**	0.0034	-0.0049	0.0035	-0.0045	0.0031
sourcekgov	-0.0130**	0.0057	-0.0010	0.0092	-0.0130**	0.0056	0.0026	0.0070
lsourcekgovtot	-0.0060	0.0306	0.0164	0.0485	0.0332	0.0302	-0.0107	0.0320
nbppopprov	15.6791**	6.2421	-0.4958	4.9749	19.2799***	6.7055	3.9583	4.4100
Intercept	-3.2927***	0.7472	-1.9070***	0.7343	-3.5968***	0.7196	-2.1470***	0.6465
Alpha	2.0145		2.2762		2.6051		2.3302	
Observations	227		266		271		318	
Weighted obs.	352		397		428		480	
F-statstic	4.46***		3.10***		3.75***		3.33***	

Table 4.11 : Fourth negative binomial model, dependent variable necpc, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0126	0.0107	-0.0017	0.0156	0.0145**	0.0069	0.0039	0.0052
ebper	0.0114**	0.0055	-0.0010	0.0055	0.0190***	0.0059	0.0004	0.0048
dspin	0.6811**	0.3024	0.4628	0.3127	0.5807*	0.3006	0.5455*	0.3026
dprod	0.5602	0.3750	0.7178	0.6991	0.7493*	0.4020	0.4383	0.5938
dproc	0.6312*	0.3611	-0.1201	0.3775	0.8485**	0.3933	-0.2063	0.3654
ndroitaec	0.2223	0.3323	0.0586	0.1725	-0.1792	0.3016	0.0872	0.2178
ndroitoec	0.0348	0.2070	0.6951**	0.3498	0.0195	0.2142	0.7244**	0.3553
dsubs	-	-	-0.0477	0.5183	-	-	-0.4659	0.4493
rexpperb_0	-0.0006	0.0038	0.0027	0.0046	0.0013	0.0042	0.0055	0.0043
sourcekgov	-0.0049	0.0052	0.0097	0.0186	-0.0014	0.0045	0.0154	0.0174
lsourcekgovtot	0.0166	0.0292	0.0389	0.0740	0.0246	0.0253	0.0054	0.0688
nbppopprov	-1.0766	6.7548	-8.2828	7.4926	8.8318	7.3963	-6.2862	7.0336
Intercept	-3.1907***	0.7646	-1.3487	1.1922	-4.8077***	0.8844	-1.1019	0.9324
Alpha	3.0214		5.2165		4.2224		5.6928	
Observations	227		266		271		318	
Weighted obs.	352		397		428		480	
F-statstic	1.95**		1.34		3.28***		1.42	

4.3 SUMMARY

In this chapter, we provided evidence on the firm-characteristics that influence the propensity to collaborate and the number of partnerships in which a firm is involved. The most significant results are provided by the 2003 and the 2005 surveys. Results form the 1999 and the 2001 surveys are significantly different and not homogeneous[26]. Considering that the four questionnaires are quite similar, we attribute this major difference to the fact that the biotechnology business has been established only in the last years. Before 2003, biotechnology firms were still struggling to find a way to make the innovation process efficient and effective. Our findings prove that the three main factors positively affecting the overall propensity to collaborate are the number of employees, the orientation towards biotechnology and being created through a spin-off. In particular, as size and growth are closely linked, especially in the case of small firms, we suggest that collaboration and growth need to be set into a dynamic framework in order to be taken into account. It is not clear, however, whether encouraging collaboration has an effect on growth or vice versa, as an econometric analysis assumes that the independent variables affect the outcome. However, the direction of the interaction between these two factors should be further investigated. Concerning the effect of being created through a spin-off, it is interesting to note that this characteristic affects the propensity to collaborate, but not the number of collaborative arrangements in which a firm is involved. The effect and the significance of the other variables we took into account (in particular, the use of IP protection means) varies depending on the type of alliance considered. In the preceding chapter, we observed that among small firms, the intensity of knowledge-related collaboration is higher than production and commercialization-related collaboration. The results of the NBR show that firm size plays a determinant role in explaining for the number of knowledge-related alliances, while it is not significant in the case of collaboration for production and commercialization. These results show that the characteristics that have an effect on collaborative behavior differ depending on the reason leading to set a partner and depending on the partner.

[26] See ANNEX D and ANNEX E.

CONCLUSIONS AND FURTHER RESEARCH

We opened this work considering that the lack of collaboration is associated with the exit of small biotechnology firms, which represent the engine of this industry, the heart of innovation and advancement in this sector. Collaboration is now recognized as an intrinsic aspect and a necessary stage of the pattern of new product and process development in biotechnology. As we said, a variety of different skills and competences are required to turn a new idea into a marketable and profitable product, and diversification makes this process more efficient and effective. As every agent needs to focus on its core competence, collaboration is now intrinsically a part of the business model in this industry. New fundamental knowledge stems from universities (or, more generally, from public institutions), that is transferred to other R&D-oriented agents, which provide further research to turn it into a marketable product. Other agents then have the skills to refine the product, to test it, to manufacture it and to commercialize it. Clearly, this chain is based on collaborative arrangements. It is therefore necessary to develop a deep understanding of the mechanisms affecting the propensity and the intensity of collaborative behavior; in this respect, this work is a premise to further research aiming at providing the means to foster collaboration in order to increase the productivity and the innovativeness of the biotechnology industry in Canada.

The results of this work show that the proportion of small collaborative firms underwent a decrease of about 10% from 1999 to 2005, with a major drop between 2001 and 2003; this change concerns both interfirm and private-public collaboration. Recalling the hypotheses we formulated in Chapter 1, the fact of developing biotechnology products rather than processes (measured through the dummy variables `dprod` and `dproc`) is found to be not significant in determining the propensity to collaborate, but has a positive effect on the number of collaborative arrangements with another firm. The relationship can be expressed in these terms: when a small biotechnology firms develops innovative products rather than processes, its intensity of collaboration with other biotechnology firms increases (in the 2005 survey). Conversely, this characteristic does not sig-

nificantly affect the intensity of collaboration with public institutions. Orientation towards biotechnology is found to be a determinant characteristic in describing the number of collaborative arrangements a firm is involved in (in particular knowledge-related alliances), while it is not significant, especially for small firms, in predicting the propensity to collaborate. Small firms naturally devote most of their resources to biotechnology, and this characteristic does not affect the probability of being involved in at least one alliance. Conversely, when the effect of biotechnology orientation is estimated considering both small and medium firms, its significance increases: as a firm grows in size, it acquires and internalizes some of the complementary assets in order to increase its independence. The positive effect of orientation towards biotechnology on the intensity of knowledge-related collaboration (for the 2003 survey, also on production/commercialization-related collaboration) deserves further work to be explained, as the direction of this relationship is not conceptually clear. Concerning Intellectual Property protection means, we found a link between the number of obtained and licensed rights and the propensity/intensity of collaboration. Yet, we cannot say whether this characteristic influences collaborative behavior, whether vice-versa or whether this should be set in the context of a more complex and dynamic model. Concerning firms created through a spin-off, the direction of the link is more evident, as it involves a precise time-dependent process. Spin-offs are more likely to be collaborative. This fact is the natural result of the particular generating process, as we discussed.

The effect of the orientation towards exporting is found to be not significant in describing both the propensity and the intensity of collaboration. Nevertheless, we recognize that this aspect should be examined by a monetary analysis on the relationships between Canada and US: in fact, the economic complexity of this issue requires a much wider framework to be understood.

The model we built to evaluate the effect of capital financing reveals a slight effect of the success in raising capital on the propensity to collaborate (stronger in the case of knowledge-related collaboration, and negligible for production/commercialization col-

laboration). Concerning the impact on the number of collaborative arrangements, a positive and significant effect is found between the variables related to capital raising in 2005 for small and medium firms and the intensity of collaboration related to production and commercialization. The negligible effect of public support on collaborative behavior deserves particular attention. In fact, if a public policy aims at improving the productivity and the competitiveness of the national biotechnology industry through, among others, fostering collaborative behavior, a relationship between public support and intensity of collaboration should be found. According to this framework, an effective funding program should require firms to collaborate in order to benefit from public support. The strategy of the government of Canada, as explained in the document "Mobilizing Science and Technology to Canada's Advantage" (2007) clearly identifies collaboration (in particular between firms and public institutions) as a determinant for success in knowledge-intensive industries, and aims at fostering the formation of tight links between different agents of the innovation system. We find strong evidence that among small firms, the number of employees plays a determinant role in increasing both the propensity towards collaboration and the number of alliances formed. However, as we anticipated at the end of Chapter 1, the size and the age of a firm must be set in the same dynamic framework describing its growth, and their role in describing collaborative behavior cannot be dissociated and isolated from a more complex framework. Although we performed a cross-section analysis, we suggest that, as an increase in size has a strong positive effect on collaborative behavior, a public policy aiming at fostering the formation of partnerships should first affect the growth of small biotechnology firms. Actually, this consideration opens up for a discussion concerning the cause-effect relationships between growth and collaboration; in order to achieve consistent results, an econometric analysis does not seem to be enough, and in-the-field research needs to be performed in order to set these elements in a dynamic and fast-changing environment. In addition, further investigation on public strategies aimed at supporting innovation within the biotechnology industry can shed light on this topic. By being aware of the factors that have an impact on collaborative behavior, it will be possible to tailor public support to the

biotechnology industry characteristics, in order to foster its growth and level of innovativeness. In this respect, understanding the cause-effect relationships between firm-specific characteristics, environmental characteristics and collaborative behavior seems to be the major threat. Yet, it is necessary to solve this issue if we are to find out what the easier characteristics to be controlled are, and to identify those on which a public policy can have a direct impact. In fact, this work provides evidence on the links between firm-specific characteristics and collaborative behavior, but does not assess the direction of these relationships. An econometric analysis provides robust results, but we must always be careful in interpreting them. For example, a statistical analysis requires the independent variables to be uncorrelated; however, this is a mere numerical artifact, which allows obtaining consistent and robust results. The reality is, nevertheless, much more complex than an econometric model, and the image we get from estimation is necessarily blurry. Econometrics is useful to obtain insight on the main characteristics related to collaborative behavior, but a more qualitative analysis must be performed to get a deeper understanding of the forces and equilibriums characterizing the biotechnology industry.

In conclusion, when the results of this work are integrated with a qualitative in-the-field research and with a model describing the effect of the government strategies to foster the productivity, growth and innovativeness of the biotechnology industry, it will be possible to provide a complete and coherent framework to increase the efficacy of the business model of this industry.

In conclusion, the results of this work needs to be integrated with:

- a qualitative in-the-field research, aimed at providing a deeper understanding of the subtle cause-effect relationship in the context of a wide and dynamic framework
- a model aimed at evaluating the effects of public financing strategies aimed at increasing the growth and the effectiveness of the biotechnology industry

Thus, it will be possible to identify the characteristics on which a public strategy can have direct effect in order to foster the productivity, growth and level of innovativeness of the biotechnology industry through collaborative behavior. This way, through additional research, it will be possible to develop public strategies that, financing and fostering biotechnology inventive firms, affect and improve collaborative behavior with the aim of increasing the global competitiveness of this industry.

REFERENCES

ABERNATHY, W. J., and UTTERBACK, J. M. (1978). Patterns of industrial innovation. *Technology Review, 14*(1).

ABRAMOVSKY, L., KREMP, E., LOPEZ, A., SCHMIDT, T., and SIMPSON, H. (2005). Understanding cooperative R&D activity: evidence from the european countries (IFS Working Paper No. 05/23 ed.). London.

AHUJA, G. (2000). The duality of collaboration: inducements and opportunities in the formation of interfirm linkages. *Strategic Management Journal, 21*, 317-343.

AIKEN, M., and HAGE, J. (1968). Organizational interdependence and intraorganizational structure. *American Sociological Review, 33*, 912-930.

ALDRICH, J., and NELSON, F. (1984). *Linear probability, logit and probit models*. Beverly Hills: Sage Publications.

AMEMIYIA, T. (1981). Qualitative response models: a survey. *Journal of Economic Literature, 19*(4), 481-536.

ARIÑO, A., and DE LA TORRE, J. (1998). Learning from failure: towards an evolutionary model of collaborative ventures. *Organization science, 9*(3), 306-325.

ARORA, A., and GAMBARADELLA, A. (1990). Complementarity and external linkages: the strategies of the large firms in biotechnology. *The journal of industrial economics, 38*(4), 361-379.

AUDRETSCH, D. B. (2001). The role of small firms in U.S. biotechnology clusters. *Small business economics, 17*, 3-15.

AUDRETSCH, D. B., and FELDMAN, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review, 86*(3), 630-640.

AUDRETSCH, D. B., and STEPHAN, P. E. (1996). Company-Scientist locational links: the case of biotechnology. *The American Economic Review, 86*(3), 641-652.

BALDWIN, J. R. (1995). Innovation: the key to success in small firms. *Statistics Canada, catalogue no. 11F0019MPE, No. 76.*

BALDWIN, J. R., GELLATLY, G., and GAUDREAULT, V. (2002). Financing innovation in new small firms: new evidence from Canada. *Statistics Canada, Catalogue No. 11F0019MIE - No. 190.*

BARLEY, S. R., FREEMAN, J., and HYBELS, R. C. (1992). Strategic alliances in commercial biotechnology. In N. Nohria & R. G. Eccles (Eds.), *Networks and organizations: structure, form and action.* Boston, Massachusetts: Harvard Business School Press.

BAUM, J. A. C., CALABRESE, T., and SILVERMAN, B. S. (2000). Don't go it alone: alliance network composition and startups' performance in Canadian biotechnology. *Strategic Management Journal, 21*, 267-294.

BAYONA, C., GRACIA-MARCO, T., and HUERTA, E. (2001). Firm's motivation for cooperative R&D: an empirical analysis of Spanish firms. *Research Policy, 22*(8-9), 1289-1307.

BEAUCAGE, J.-S., and BEAUDRY, C. (2006, 26-28 May 2006). *Importance of knowledge networks within Canadian biotechnology clusters.* Paper presented at the Conférence de l'Association canadienne d'économique, Montréal.

BEAUDRY, C., and BRESCHI, S. (2001). Are firms in clusters really more innovative? *Economics of innovation and new technology, 12*(4), 325-342.

BERNDT, E., HALL, B., HALL, R., and HAUSMAN, J. (1974). Estimation and inference in nonlinear models. *Annals of economic and social measurement, 3*(4), 653-665.

BISHOP, Y. M., FEINBERG, S. E., and HOLLAND, P. W. (1975). *Discrete multivariate analysis: theory and practice.* Cambridge, MA: MIT Press.

BLISS, M. (1982). *La découverte de l'insuline.*

BORTKIEWICZ, L. (1898). *Das Gesetz de Kleinen Zahlen.* Leipzig: Teubner.

BRESCHI, S., MALERBA, F., and ORSENIGO, L. (2000). Technological regimes and Schumpeterian patterns of innovation. *The Economic Journal, 110*, 388-410.

CAMERON, C., and TRIVEDI, P. (1986). Econometric models based on count data: comparisons and applications of some estimators and tests. *Journal of Applied Econometrics, 1*, 29-53.

CAMERON, C., and TRIVEDI, P. (1998). *Regression analysis of count data.* New York: Cambridge University Press.

CAMERON, C., and WINDMEIJER, F. (1996). R-Squared measures for count data regression models with applications to health care utilization. *Journal of Business and Economic Statistics, 14*, 209-220.

CASSIMAN, B., and VEUGELERS, R. (2002). R&D cooperation and spillovers: some empirical evidence from Belgium. *American Economic Review, 44*(3), 1169-1184.

CHANDY, R. K., and TELLIS, G. J. (1998). Organizing for radical product innovation: the overlooked role of willingness to cannibalize. *Journal of Marketing Research, 35*, 474-487.

CHILD, J., and FAULKNER, D. (1998). *Strategies of cooperation*: Oxford University Press.

CHRISTENSEN, C. (1997). *The innovator's dilemma*: Harvard Business School Press.

COCKBURN, I. M., and HENDERSON, R. M. (1998). Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *The journal of industrial economics, XLVI*(2), 157-182.

COHEN, W. M. (1995). Empirical studies on innovative activity. In Cambridge (Ed.), *Handbook of the economics of innovation and technological change*: Blackwell.

COHEN, W. M., FLORIDA, R., RANDAZZESE, L., and WALSH, J. (1998). Industry and the academy: uneasy partners in the cause of technological advance. In R. Noll (Ed.), *Challenges to research universities*. Washington, DC: Brookings Institution Press.

COHEN, W. M., NELSON, R. R., and WALSH, J. P. (2002). Links and impacts: the influence of public research on industrial R&D. *Management Science, 48*(1), 1-23.

COLYVAS, J., CROW, M., GELIJNS, A., MAZZOLENI, R., NELSON, R. R., ROSENBERG, N., et al. (2002). How do university inventions get into practice? *Management Science, 48*(1), 61-72.

COWAN, R., and JONARD, N. (2003). The dynamics of collective invention. *Journal of Economic Behavior & Organization, 52*(4), 513-532.

COX, D. R., and SNELL, E. J. (1968). A general definition of residuals [with discussion]. *Journal of the Royal Statistical Society B, 30*, 248-275.

CRAGG, J. G., and UHLER, R. (1970). The demand for automobiles. *Canadian Journal of Economics, 3*, 386-406.

CRAMER, J. S. (1986). *Econometric applications of maximum likelihood methods*. Cambridge: Cambridge University Press.

CROWDER, M. J. (1976). Maximum Likelihood Estimation for dependent observations. *Journal of the Royal Statistical Society B, 38*(1), 45-53.

DACHS, B., EBERSBERGER, B., and PYKA, A. (2004). *Why do firms co-operate for innovation? A comparison of Austrian and Finnish CIS 3 results*. Augsburg.

DAVIDSON, and MACKINNON. (1993). *Estimation and inference in econometrics*. New York: Oxford University Press.

DELERUE, H. (2004). *Performance des alliances: une gestion duale du risque relationnel.* Paper presented at the 13e conférence de l'AIMS, Vallée de la Seine.

DHRYMES, P. (1984). Limited dependent variables. In Z. Griliches & M. Intriligator (Eds.), *Handbook of econometrics* (Vol. 2). Amsterdam: North Holland.

DOGSON, M. (1994). Technological collaboration and innovation. In M. Dogson & R. R. (Eds.), *The handbook of industrial innovation* (pp. 285-292). Cheltenham: Edward Elgar.

DURBIN, J. (1954). Nonresponse and callbacks in surveys. *BISI, 34*(2), 72-86.

EDWARDS, M. G., MURRAY, F., and YU, R. (2003). Value creation and sharing among universities, biotechnology and pharma. *Nature Biotechnology, 21*(6), 618-624.

EFRON, B. (1978). Regression and ANOVA with zero-one data: measures of residual variation. *Journal of the American Statistical Association, 73*(361), 113-121.

ERNST&YOUNG. (2007). *The global biotechnology report 2007: The global perspective.*

FOMBY, T., HILL, C., and JOHNSON, S. (1984). *Advanced econometric methods.* Needham, MA: Springer-Verlag.

FOX, J. (1991). *Regression diagnostics*. Newbury Park: Sage.

GALBRAITH, J. K. (1952). *American Capitalism*. Boston: Houghton Mifflin.

GANS, J. S., and STERN, S. (2002). Managing ideas: commercialization strategies for biotechnology. *Kellogg Biotechnology Review, 1*(1).

GIARRATANA, M. S. (2004). The birth of a new industry: entry by start-ups and the drivers of firm growth - The case of encryption software. *Research Policy, 33*, 787-806.

GODFREY, L. G. (1988). *Misspecification tests in econometrics*. Cambridge: Cambridge University Press.

GOLDBERGER, A. S. (1991). *A course in econometrics*. Cambridge: Harvard University Press.

GOURIEOUX, C., MONFORT, A., RENAULT, E., and TROGNON, A. (1987a). Generalized residuals. *Journal of Econometrics, 34*, 5-32.

GOURIEOUX, C., MONFORT, A., RENAULT, E., and TROGNON, A. (1987b). Simulated residuals. *Journal of Econometrics, 34*, 201-252.

GOURIEROUX, C., and MONTFORT, A. (1995). *Statistics and econometric models*. Cambridge: Cambridge University Press.

GREENE, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River: Prentice Hall.

GRILICHES, Z. (1986). Economic data issues. In Z. Griliches & M. Intriligator (Eds.), *Handbook of econometrics*. Amsterdam: North Holland.

GULATI, R. (1998). Alliances and networks. *Strategic Management Journal, 19*(4), 293-317.

HACHE, J. (2005). *Les enjeux de la biotechnologie*. Colombelle: EMS éditions management et société.

HAGEDOORN, J., and SADOWSKI, B. (1999). The transition from strategic technology alliances to mergers and acquisitions: an exploratory study. *Journal of Management Studies, 36*(1), 87-107.

HALL, B. H., and VAN REENEN, J. (2000). How effective are fiscal incentives for R&D? A review of the evidence. *Research Policy, 29*, 449-469.

HAUCK, W. W. J., and DONNER, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association, 72*(360), 851-853.

HENDERSON, B. D. (1998). The product portfolio. In C. W. Stern & G. Stalk (Eds.), *Perspectives on Strategy from the Boston Consulting Group*. New York: John Wiley & Sons.

HENDRICKS, W. A. (1949). Adjustment for bias by non-response in mailed surveys. *Agricultural Economics Research, 1*(52).

HOSMER, D. W., and LEMESHOW, S. (1989). *Applied logistic regression*. New York: John Wiley.

JAFFE, A. B. (1998). The importance of "spillovers" in the policy mission of the advanced technology program. *Journal of Technology Transfer, 23*(2), 11-19.

JAFFE, A. B., TRAJTENBERG, M., and HENDERSON, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics, 108*(3), 577-598.

JOSEPH, R., BORDT, M., and HAMDANI, D. (2005). Characteristics of business incubation in Canada, 2005. *Statistics Canada, catalogue no. 88F0006XIE, No. 007.*

JUDGE, G. G., GRIFFITHS, W. E., HILL, R. C., and LEE, T. C. (1985). *The theory and practice of econometrics* (2nd ed.). New York: John Wiley.

KALE, P., SINGH, H., and PERLMUTTER, H. (2000). Learning and protection of proprietary assets in strategic alliances: building relational capital. *Strategic Management Journal, 21*, 217-237.

KALTON, G. (1983). *Compensating for Missing Survey Data.* Ann Arbor: Research Report Series, University of Michigan.

KANNG, N., and SAKAI, K. (2000). International strategic alliances - their role in industrial globalization. *STI Working Paper, 2005/5, OECD, Paris.*

KAUFMAN, R. L. (1996). Comparing effects in dichotomous logistic regression: a variety of standardized coefficients. *Social Science Quarterly, 77*, 90-109.

KIM, J. W., and HIGGINS, M. C. (2007). Where do alliances come from? The effects of upper echelons on alliance formation. *Research Policy, 36*, 499-514.

KISH, L. (1967). *Survey Sampling* (2nd ed.). New York: Wiley.

KLEIN, R., and SPADY, R. (1993). An efficient semiparametric estimator for discrete choice models. *Econometrica, 61*, 387-421.

KRUGMAN, P. (1991). *Geography and trade.* Cambridge, Ma: MIT Press.

LAITILA, T. (1993). A pseudo-R^2 measure for limited and qualitative dependent variable models. *Journal of Econometrics, 56*, 341-356.

LEVY, P. S., and LEMESHOW, S. (1999). *Sampling of populations Methods and applications*: Wiley.

LITTLE, R. J. A. (1987). *Statistical Analysis With Missing Data.* New York: Wiley.

LONG, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks: Sage Publications.

LONG, J. S., and FREESE, J. (2006). *Regression models for categorical dependent variables using Stata.* College Station: Stata Press.

LONMO, C., and MCNIVEN, C. (2007). Selected results of the Biotechnology Use and Development Survey 2005. *Statistics Canada.*

MADDALA, G. (1983). *Limited dependent and qualitative variables in econometrics.* New York: Cambridge University Press.

MALERBA, F. (2002). Sectoral systems of innovation and production. *Research Policy, 31*, 247-264.

MALERBA, F., and ORSENIGO, L. (1993). Technological regimes and firm behavior. *Industrial and corporate change, 2*, 45-71.

MATZKIN, R. (1993). Nonparametric identification and estimation of polytomous choice models. *Journal of Econometrics, 58*, 137-168.

MCCULLAGH, P. (1986). The conditional distribution of goodness-of-fit statistics for discrete data. *Journal of the American Statistical Association, 81*, 104-107.

MCCULLAGH, P., and NELDER, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.

MCFADDEN, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. New York: Academic Press.

MCKELVEY, R., and ZAVOINA, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology, 4*, 103-120.

MERKLE, L., and ZIMMERMANN, K. F. (1992). The demographics of labor turnover: a comparison ordinal probit and censored count data models. *Reserches Economiques de Louvain, 58*, 283-307.

MIOTTI, L., and SACHWALD, F. (2003). Cooperative R&D: why and with whom? An integrated framework analysis. *Research Policy, 32*(8), 1481-1499.

Mobilizing science and technology to Canada's advantage. (2007). Government of Canada.

MONTGOMERY, D. C. (1997). *Design and analysis of experiments* (4th ed.). New York: Wiley.

MOORMAN, C., and MINER, A. S. (1997). The impact of organizational memory on new product performance. *Journal of Marketing Research, 34*, 91-107.

MOWERY, D. C. (1998). Collaborative R&D: how effective is it? *Issues in Science and Technology, 15*(1), 37-44.

MYTELKA, L. (1999). New trends in biotechnology networking. *International Journal of Biotechnology, 1*(1), 30-41.

NEGASSI, S. (2004). R&D co-operation and innovation a microeconometric study on French firms. *Research Policy, 33*(3), 365-384.

NELSON, R., and WINTER, S. (1974). Neoclassical vs evolutionary theories of economic growth: critique and prospectus. *Economic Journal, 84*, 886-905.

NEWEY, W., and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. In E. R. & D. McFadden (Eds.), *Handbook of econometrics* (Vol. IV).

NICHOLLS-NIXON, C. (1993). *Absorptive capacity and technological sourcing: Implications for the responsiveness of established firms.* Purdue University.

NIOSI, J. (2003). Alliances are not enough explaining rapid growth in biotechnology firms. *Research Policy, 32*, 737-750.

NIOSI, J., and BAS, T. G. (2001). The competencies of regions - Canada's clusters in biotechnology. *Small Business Economics, 17*, 31-42.

OECD. (2002). *Frascati Manual* (6th ed.).

OECD. (2005). *Oslo Manual* (3rd ed.).

OLIVER, A. L. (2004). On the duality of competition and collaboration: network-based knowledge relations in the biotechnology industry. *Scandinavian Journal of Management, 20*, 151-171.

OLIVER, R. W. (2003). *The biotech age: the business of biotech and how to profit from it.* New York: McGraw-Hill.

OLSON, E. M., WALKER, O. C., and RUEKERT, R. W. (1995). Organizing for effective new product development: the moderating role of product innovativeness. *Journal of Marketing, 59*, 48-62.

PIERCE, D. A., and SCHAFER, D. W. (1986). Residuals in generalized linear models. *Journal of the American Statistical Association, 81*, 977-986.

PINDYCK, R. S., and RUBINFELD, D. L. (1998). *Econometric models and economic forecasts*. Boston: McGraw Hill.

PISANO, G. P. (1991). The governance of innovation: vertical integration and collaborative arrangements in the biotechnology industry. *Research Policy, 20*, 237-249.

PISANO, G. P. (1997). *The development factory*. Boston, MA: Harvard Business School Press.

PISANO, G. P., and TEECE, D. J. (1989). Collaborative arrangements and global technology strategy. In R. Rosenbloom & R. A. Burgelman (Eds.), *Research on technological innovation, management and policy* (Vol. 4, pp. 227-256). Greenwich, CT: JAI Press.

POWELL, W. W., and BRANTLEY, P. (1992). Competitive cooperation in biotechnology: learning through networks? In N. Nohria & R. G. Eccles (Eds.), *Networks and organiza-*

tions: structure, form, and action. Boston, Massachusetts: Harvard Business School Press.

POWELL, W. W., KOPUT, K. W., and SMITH-DOERR, L. (1996). Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology. *Administrative Science Quarterly, 41*(1), 116-145.

POWELL, W. W., WHITE, D. R., KOPUT, K. W., and OWEN-SMITH, J. (2005). Networks dynamics and fields evolution: the growth of interorganizational collaboration in the life sciences. *American Journal of Sociology, 110*, 1132-1205.

PREGIBON, D. (1981). Logistic regression diagnostics. *Annals of Statistics, 9*, 705-724.

PREVEZER, M. (1997). The dynamics of industrial clustering in biotechnology. *Small Business Economics, 9*, 255-271.

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

RAO, J. N. K., and SCOTT, A. J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *The Annals of Statistics, 12*(1), 46-60.

RIOLLI-SALTZMAN, L., and LUTHANS, F. (2001). After the bubble burst: how to small high-tech firms can keep in front of the wave. *The academy of management executive, 15*(3), 114-124.

ROBBINS-ROTH, C. (2001). *Le business des biotechnologies*. Paris: Dunod.

ROSENBERG, N. (1990). Why do firms do basic research (with their own money)? *Research Policy, 19*, 165-174.

ROSENBERG, N., and NELSON, R. (1994). American universities and technical advance in industry. *Research Policy, 23*, 323-348.

ROTHENBERG, T. J. (1984). Hypothesis testing in linear models when the error covariance matrix is nonscalar. *Econometrica, 52*, 827-842.

RYANS, J. K. J., and SHANKLIN, W. L. (1989). Marketing to nonexistent markets. In *Customer-driven marketing*: Lexington Books.

SCHMIDT, T. (2007). *Motives for innovation co-operation - evidence from the Canadian Survey of Innovation.*

SHAN, W., WALKER, G., and KOGUT, B. (1994). Interfirm cooperation and startup innovation in the biotechnology industry. *Strategic Management Journal, 15*, 387-394.

SOETE, L. L. G. (1979). Firm size and inventive activity. *European Economic review, 12*, 319-340.

STATISTICSCANADA. (2005). *Canadian Trends in Biotechnology, 2nd edition*: Statistics Canada.

STUART, T. E., and DING, W. W. (2006). When do scientists become entrepreneurs? The social structure antecedents of commercial activity in the academic life sciences. *American Journal of Sociology, 112*(1), 97-144.

STUART, T. E., OZDEMIR, S. Z., and DING, W. W. (2007). Vertical alliance networks: the case of university-biotechnology-pharmaceutical alliance chains. *Research Policy, 36*, 477-498.

TRAJTENBERG, M. (1989). The welfare analysis of product innovations, with an application to computed tomography scanners. *Journal of Political Economy, 97*(21), 444-479.

TRAORÉ, N. (2004). Biotechnology Use and Development Survey: methodology, issues and responses. *Statistics Canada, Catalogue no. 88F0006XIE2004006.*

VEALL, M. R., and ZIMMERMANN, K. F. (1996). Pseudo-R^2 measures for some common limited dependent variable models. *Sonderforschungsbereich 386, Paper 18.*

VON HIPPEL, E. (1989). Cooperation between rivals: informal know how trading. In B. Carlsson (Ed.), *Industrial dynamics*. Dordrecht: Kluwer Academic Publishers.

WEISBERG, S. (1980). *Applied linear regression*. New York: John Wiley.

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica, 53*, 1-16.

WOLFE, D. A., and GERTLER, M. S. (2004). Clusters from the inside and out: local dynamics and global linkages. *Urban studies, 41*, 1071-1093.

ZARKOVICH, S. S. (1963). *Sampling Methods and Censuses;* (Vol. II, Quality of Statistical Data). Rome: FAO.

ZUCKER, L. G., DARBY, M. R., and BREWER, M. B. (1998). Intellectual human capital and the birth of U.S. biotechnology enterprises. *The American Economic Review, 88*(1), 290-306.

ANNEX A

VARIABLES

The next tables list all the variables we have drawn from the Biotechnology Use and Development surveys. For each variable, a synthetic description is provided. For binary, count and Lickert-scale variables, the type is indicated too. When not evident, the coding is provided. Finally, information is given on which surveys each variable is available.

Table A.1 : Collaborative behavior-related variables.

Variable	Description	Type	Coding	05	03	01	99
nec	Number of collaborative arrangements	integer		X	X	X	
ec	Firm involved in at least one collaborative arrangement	binary	1: nec > 0 0: otherwise	X	X	X	X
necrd	Number of partnerships to conduct R&D	integer		X		X	
necrc	Number of partnerships for regulatory affairs	integer		X		X	
necab	Number of partnerships to gain access to external patents	integer		X		X	
necpi	Number of partnerships to gain access to external intellectual property	integer		X		X	
neccc	Number of partnerships to gain access to external knowledge and competence	integer		X			
neccnd	Number of partnerships to gain access to knowledge non available internally	integer			X		
neces	Number of partnerships to gain access to external scientific expertise	integer			X		
necdcrd	Number of partnerships to reduce R&D costs	integer			X		
necdcrc	Number of partnerships to reduce regulatory/clinical costs	integer			X		
necc	Number of partnerships concerning knowledge	integer		X	X	X	

Table A.1 : Collaborative behavior-related variables (continuation).

Variable	Description	Type	Coding	05	03	01	99
necp	Number of partnerships concerning production/manufacturing	integer		X		X	
necmd	Number of partnerships to gain access to markets/distribution channels	integer		X		X	
neck	Number of partnerships to gain access to external capital	integer		X		X	
necdd	Number of partnerships to lower expenses	integer		X			
necdc	Number of partnerships to lower costs	integer			X		
necdcp	Number of partnerships for production cost reduction	integer			X		
necpc	Number of partnerships concerning production/commercialisation	integer		X	X	X	
neceo	Number of partnerships as a precursor to a formal agreement	integer			X		
necrrv	Number of partnerships to reduce risk/exposure	integer			X		
necar	Number of partnership for other reasons	integer		X	X	X	
ecrd	The firm has at least one partnership to conduct R&D	binary	1: necrd > 0 0: otherwise	X		X	X
ecrc	The firm has at least one partnership for regulatory affairs	binary	1: necrc > 0 0: otherwise	X		X	X
ecab	The firm has at least one partnership to gain access to external patents	binary	1: necab > 0 0: otherwise	X		X	
ecpi	The firm has at least one partnership to gain access to the partner's intellectual property	binary	1: necpi > 0 0: otherwise	X		X	X
eccc	The firm has at least one partnership to gain access to external knowledge/skill	binary	1: neccc > 0 0: otherwise	X			X

Table A.1 : Collaborative behavior-related variables (continuation).

Variable	Description	Type	Coding	05	03	01	99
eccnd	The firm has at least one partnership to gain access to knowledge not available internally	binary	1: neccnd > 0 0: otherwise		X		
eces	The firm has at least one partnership to gain access to external scientific expertise	binary	1: neces > 0 0: otherwise		X		
ecdcrd	The firm has at least one partnership for R&D cost reduction	binary	1: necrcrd > 0 0: otherwise		X		
ecdcrc	The firm has at least one partnership for regulatory/clinical affairs cost reduction	binary	1: necrcrc > 0 0: otherwise		X		
ecc	The firm has at least one partnership concerning knowledge	binary	1: necc > 0 0: otherwise	X	X	X	X
ecp	The firm has at least one partnership concerning production/manufacturing	binary	1: necp > 0 0: otherwise	X		X	X
ecmd	The firm has at least one partnership to gain access to markets/distribution channels	binary	1: necmd > 0 0: otherwise	X		X	X
eck	The firm has at leas one partnership to gain access to capital	binary	1: neck > 0 0: otherwise	X		X	X
ecdd	The firm has at least one partnership to reduce expenditures	binary	1: necdd > 0 0: otherwise	X			
ecdcp	The firm has at least one partnership to reduce production costs	binary	1: necdpc > 0 0: otherwise		X		
ecpc	The firm has at least one partnership to reduce production and commercialisation costs	binary	1: necpc > 0 0: otherwise	X	X	X	X
eceo	The firm has at least one partnership as a precursor to a formal agreement	binary	1: neceo > 0 0: otherwise		X		
ecrrv	The firm has at least one partnership to reduce risk/exposure	binary	1: necrrv > 0 0: otherwise		X		
ecar	The firm has at least one partnership for other reasons	binary	1: necar > 0 0: otherwise	X	X	X	X
neceb	Number of partnerships with another biotechnology firm	integer		X	X	X	

Table A.1 : Collaborative behavior-related variables (continuation).

Variable	Description	Type	Coding	05	03	01	99
necep	Number of partnerships with a pharmaceutical company	integer		X			
necabp	Number of partnerships with a non-biotechnology and non-pharmaceutical firm	integer		X			
neceab	Number of partnerships with a non-biotechnology firm	integer			X	X	
necepri	Number of partnerships with another firm	integer		X	X	X	
necuh	Number of partnerships with an academic institution/hospital	integer		X	X	X	
neclg	Number of partnerships with a Government lab or agency	integer		X	X	X	
necipub	Number of partnerships with a public institution	integer		X	X	X	
neccpri	Number of partnerships concerning knowledge with another firm	integer		X	X	X	
neccpub	Number of partnerships concerning knowledge with a public institution	integer		X	X	X	
necpcpri	Number of partnerships concerning production/commercialisation with another firm	integer		X	X	X	
necpcpub	Number of partnerships concerning production/commercialisation with a public institution	integer		X	X	X	
eceb	The firm has at least one partnership with a biotechnology firm	binary	1: neceb > 0 0: otherwise	X	X	X	
ecep	The firm has at least one partnership with a pharmaceutical company	binary	1: necep > 0 0: otherwise	X			
ecabp	The firm has at least one partnership with a non-biotechnology and non-pharmaceutical firm	binary	1: necabp > 0 0: otherwise	X			
eceab	The firm has at least one partnership with a non-biotechnology firm	binary	1: neceab > 0 0: otherwise		X	X	

Table A.1 : Collaborative behavior-related variables (continuation).

Variable	Description	Type	Coding	05	03	01	99
eceei	The firm has at least one partnership with another smaller or equal sized firm	binary					X
ecets	The firm has at least one partnership with a larger firm	binary					X
ecepri	The firm has at least one partnership with another firm	binary	1: necepri > 0 0: otherwise	X	X	X	X
ecuh	The firm has at least one partnership with an academic institution/hospital	binary	1: necuh > 0 0: otherwise	X	X	X	X
eclg	The firm has at least one partnership with a government lab or agency	binary	1: neclg > 0 0: otherwise	X	X	X	X
ecipub	The firm has at least one partnership with a public institution	binary	1: ecipub > 0 0: otherwise	X	X	X	X
eccpri	The firm has at least one partnership concerning knowledge with another firm	binary	1: neccpri > 0 0: otherwise	X	X	X	
eccpub	The firm has at least one partnership concerning knowledge with a public institution	binary	1: neccpub > 0 0: otherwise	X	X	X	
ecpcpri	The firm has at least one partnership concerning manufacturing/production with another firm	binary	1: necpcpri > 0 0: otherwise	X	X	X	
ecpcpub	The firm has at least one partnership concerning manufacturing/production with a public institution	binary	1: necpcpub > 0 0: otherwise	X	X	X	
ecebus	The firm has at least one partnership with a biotechnology firm in the U.S.	binary		X	X		
ecepus	The firm has at least one partnership with a pharmaceutical company in the U.S.	binary		X			
eceabpus	The firm has at least one partnership with a non-biotechnology/non-pharmaceutcal firm in the U.S.	binary		X			
eceabus	The firm has at least one partnership with a non-biotechnology firm in the U.S.	binary			X		
eceeius	The firm has at least one partnership with another smaller or equal sized firm in the U.S.	binary					X

Table A.1 : Collaborative behavior-related variables (continuation).

Variable	Description	Type	Coding	05	03	01	99
ecesus	The firm has at least one partnership with a larger firm in the U.S.	binary					X
eceprius	The firm has at least one partnership with a firm in the U.S.	binary		X	X		X
ecuhus	The firm has at least one partnership with an academic institution/hospital in the U.S.	binary		X	X		X
eclgus	The firm has at least one partnership with a government lab or agency in the U.S.	binary		X	X		X
ecipubus	The firm has at least one partnership with a public institution in the U.S.	binary		X	X		X
ecus	The firm has at least one partnership in the U.S.	binary		X	X		X
ecebeu	The firm has at least one partnership with a biotechnology firm in Europe	binary		X	X		
ecepeu	The firm has at least one partnership with a pharmaceutical company in Europe	binary		X			
eceabpeu	The firm has at least one partnership with a non-biotechnology/non-pharmaceutical firm in Europe	binary		X			
eceabeu	The firm has at least one partnership with a non-biotechnology firm in Europe	binary			X		
eceeieu	The firm has at least one partnership with an-other smaller or equal sized firm in Europe	binary					X
eceseu	The firm has at least one partnership with a larger firm in Europe	binary					X
eceprieu	The firm has at least one partnership with a firm in Europe	binary		X	X		X
ecuheu	The firm has at least one partnership with an academic institution/hospital in Europe	binary		X	X		X
eclgeu	The firm has at least one partnership with a government lab or agency in Europe	binary		X	X		X
ecipubeu	The firm has at least one partnership with a public institution in Europe	binary		X	X		X
eceu	The firm has at least one partnership in Europe	binary		X	X		X
eceba	The firm has at least one partnership with a biotechnology firm in Asia	binary		X	X		

Table A.1 : Collaborative behavior-related variables (end).

Variable	Description	Type	Coding	05	03	01	99
ecepa	The firm has at least one partnership with a pharmaceutical company in Asia	binary		X			
eceabpa	The firm has at least one partnership with a non-biotechnology/non-pharmaceutical firm in Asia	binary		X			
eceaba	The firm has at least one partnership with a non-biotechnology firm in Europe	binary			X		
eceeia	The firm has at least one partnership with another smaller or equal sized firm in Asia	binary					X
ecesa	The firm has at least one partnership with a larger firm in Europe	binary					X
ecepria	The firm has at least one partnership with a firm in Asia	binary		X	X		X
ecuha	The firm has at least one partnership with an academic institution/hospital in Asia	binary		X	X		X
eclga	The firm has at least one partnership with a government lab or agency in Asia	binary		X	X		X
ecpuba	The firm has at least one partnership with a public institution in Asia	binary		X	X		X
eca	The firm has at least one partnership in Asia	binary		X	X		X

Table A.2 : Variables related to reasons leading to collaborative behavior.

Variable	Description	Type	05	03	01	99
raiscolhcrd	Importance of collaboration with a foreigner firm for R&D	lickert	X			
raiscolhcrc	Importance of collaboration with a foreigner firm for regulatory affairs	lickert	X			
raiscolhcp	Importance of collaboration with a foreigner firm for production/manufacturing	lickert	X			
raiscolhcmd	Importance of collaboration with a foreigner firm to gain access to markets/distribution channels	lickert	X			
raiscolhck	Importance of collaboration with a foreigner firm to gain access to capital	lickert	X			
raiscolhcpi	Importance of collaboration with a foreigner firm to gain access to intellectual property	lickert	X			
raiscolhccc	Importance of collaboration with a foreigner firm to gain access to knowledge/skill	lickert	X	X		
raiscolhcdc	Importance of collaboration with a foreigner firm to gain access to knowledge not available internally	lickert		X		
raiscolhcrcrd	Importance of collaboration with a foreigner firm for R&D cost reduction	lickert		X		
raiscolhcrcrc	Importance of collaboration with a foreigner firm for regulatory affairs cost reduction	lickert		X		
raiscolhcrcp	Importance of collaboration with a foreigner firm for production cost reduction	lickert		X		
raiscolhceo	Importance of collaboration with a foreigner firm as a precursor to a formal agreement	lickert		X		
raiscolhcrr	Importance of collaboration with a foreigner firm for risk/exposure reduction	lickert		X		
raiscolhca	Importance of collaboration with a foreigner firm for another reason	lickert	X	X		

Table A.3 : Location variables and province codes.

Variable	Code	Condition
province	**59**	If firm's operations are in British Columbia
	48	If firm's operations are in Alberta
	47	If firm's operations are in Saskatchewan
	46	If firm's operations are in Manitoba
	35	If firm's operations are in Ontario
	24	If firm's operations are in Québec
	13	If firm's operations are in New Brunswick
	12	If firm's operations are in Nova Scotia
	11	If firm's operations are in Prince Edward Island
	10	If firm's operations are in Newfoundland and Labrador
province2	**59**	If firm's operations are in British Columbia
	48	If firm's operations are in Alberta
	47	If firm's operations are in Saskatchewan
	46	If firm's operations are in Manitoba
	35	If firm's operations are in Ontario
	24	If firm's operations are in Québec
	15	If firm's operations are in Nouveau Brunswick, Nova Scotia, Prince Edward Island or Newfoundland and Labrador
province3	**59**	If firm's operations are in British Columbia
	35	If firm's operations are in Ontario
	24	If firm's operations are in Québec
	15	If firm's operations are in Nouveau Brunswick, Nova Scotia, Prince Edward Island or Newfoundland and Labrador
	45	If firm's operations are in Alberta, Saskatchewan or Manitoba

Table A.4 : Location dummy variables.

Variable	Description	Type	Coding
dcb	Firm's location: British Columbia	binary	1: firm's operations are in British Columbia 0: otherwise
dal	Firm's location: Alberta	binary	1: firm's operations are in Alberta 0: otherwise
dsa	Firm's location: Saskatchewan	binary	1: firm's operations are in Saskatchewan 0: otherwise
dma	Firm's location: Manitoba	binary	1: firm's operations are in Manitoba 0: otherwise
don	Firm's location: Ontario	binary	1: firm's operations are in Ontario 0: otherwise
dqc	Firm's location: Québec	binary	1: firm's operations are in Québec 0: otherwise
dnb	Firm's location: New Brunswick	binary	1: firm's operations are in New Brunswick 0: otherwise
dns	Firm's location: Nova Scotia	binary	1: firm's operations are in Nova Scotia 0: otherwise
dpe	Firm's location: Prince Edward Island	binary	1: firm's operations are in Prince Edward Island 0: otherwise
dnf	Firm's location: Newfoundland and Labrador	binary	1: firm's operations are in Newfoundland and Labrador 0: otherwise
dmar	Firm's location: New Brunswick, Nova Scotia, Prince Edward Island or Newfoundland and Labrador	binary	1: firm's operations are in New Brunswick, Nova Scotia, Prince Edward Island or Newfoundland and Labrador 0: otherwise
dpra	Firm's location: Alberta, Manitoba or Saskatchewan	binary	1: firm's operations are in Alberta, Manitoba or Saskatchewan 0: otherwise

Table A.5 : Size-related variables.

Variable	Description	Type	Coding	05	03	01	99
e	Number of firm's employees in Canada	integer		X	X	X	X
taille	Dummy for firm's size	possible values: 1 2 3	1: e < 50 2: 50 ≤ e < 150 3: e ≥ 150	X	X	X	X
dpetite	Dummy for size: small firms	binary	1: taille = 1 0: otherwise	X	X	X	X
dmoy	Dummy for size: medium firms	binary	1: taille = 2 0: otherwise	X	X	X	X
dpme	Dummy for size: small and medium firms	binary	1: taille = 1 or 2 0: otherwise	X	X	X	X
dgrande	Dummy for size: large firms	binary	1: taille = 3 0: otherwise	X	X	X	X
eb	Number of employees with biotechnology-related responsibilities	integer		X	X	X	X
ebper	Percentage of employees with biotechnology-related responsibilities			X	X	X	X

Table A.6 : Age-related variables.

Variable	Description	Type	05	03	01	99
age	Firm's age	integer	X	X	X	
ageipo	Firm's age since IPO	integer	X	X	X	
agef	Firm's age since merger	integer	X	X	X	

Table A.7 : Variables related to the financial situation

Variable	Description	05	03	01	99
rt_1	Total revenue, preceding year	X	X	X	X
rt_0	Total revenue, current year	X	X	X	X
rbper1	Percentage of revenues from biotechnology, preceding year	X	X	X	X
rbper0	Percentage of revenues from biotechnology, current year	X	X	X	X
rdt_1	Total R&D expenditures, preceding year	X	X	X	X
rdt_0	Total R&D expenditures, current year	X	X	X	X
rdbper1	Percentage of R&D expenditures for biotechnology, preceding year	X	X	X	X
rdbper0	Percentage of R&D expenditures for biotechnology, current year	X	X	X	X

Table A.8 : Variables related to firm type

Variable	Description	Type	05	03	01	99
dpub	The firm is public	binary	X	X	X	
dfus	The firm merged with another firm	binary	X	X	X	
dsubs	The firm is a Canadian owned company	binary	X			
dsubsi	The firm has branches outside Canada	binary	X	X	X	
drdi	The firm conducts R&D outside Canada	binary	X			
dspin	The firm is a spin-off	binary	X	X	X	X
dspinhu	The firm is a spin-off from a university/hospital	binary	X	X	X	X
dspineb	The firm is a spin-off from another biotechnology company	binary	X	X	X	X
dspinea	The firm is a spin-off from a non-biotechnology company	binary	X	X	X	X
dspinlg	The firm is a spin-off from a government agency/lab	binary	X	X	X	X

Table A.9 : Variables related to biotechnology products.

Variable	Description	Coding	05	03	01	99
dprod	The firm is currently developing products requiring the use of biotechnology	binary	X	X	X	X
dproc	The firm is currently developing processes requiring the use of biotechnology	binary	X	X	X	X
nprodsh	Number of biotechnology products/processes in human health	integer	X	X	X	X
nprodba	Number of biotechnology products/processes in agriculture	integer	X	X	X	X
nprodbaapa	Number of biotechnology products/processes in aquaculture, agriculture and food processing	integer	X	X	X	X
nprodrn	Number of biotechnology products/processes in natural resources	integer	X	X	X	X
nprode	Number of biotechnology products/processes in environment	integer	X	X	X	X
nproda	Number of biotechnology products/processes in aquaculture	integer	X	X	X	X
nprodbi	Number of biotechnology products/processes in bioinformatics	integer	X	X	X	X
nprodpa	Number of biotechnology products/processes in food processing	integer	X	X	X	X
nprodrd	Number of products/processes at R&D stage	integer	X	X	X	X
nprodpc	Number of products/processes at preclinical trials/confined field trials/pre-market stage	integer	X	X	X	X
nprodrc	Number of products/processes at regulatory phase/unconfined release assessment/final pre-market assessment stage	integer	X	X	X	X
nprodpm	Number of products/processes at approved/ on market/ production stage	integer	X	X	X	X

Table A.10 : Contracts-related variables.

Variable	Description	Type	05	03	01	99
ncontorb	Number of contracts with a research organisation	integer	X	X		
ncontofb	Number of contracts with a manufacturing organisation	integer	X			
ncontuhb	Number of contracts with an university/hospital	integer	X	X		
ncontlgb	Number of contracts with a government lab	integer	X	X		
ncontebb	Number of contracts with another biotechnology firm	integer	X	X		
ncontab	Number of contracts with another organization	integer	X	X		
ncont	Total number of contracts	integer	X	X	X	
vtcontrdb	Total value of R&D contracts		X	X		
vtcontrcb	Total value of clinical/regulatory affairs contracts		X	X		
vtcontpb	Total value of production contracts		X	X		
vtcontab	Total value of contracts for other reasons		X	X		
vcontorb	Total value of contracts with a research organization		X	X		
vcontofb	Total value of contracts with a manufacturing organization		X			
vcontuhb	Total value of contracts with a university/hospital		X	X	X	
vcontlgb	Total value of contracts with a government lab		X	X		
vcontenb	Total value of contracts with another biotechnology firm		X	X		

Table A.10 : Contracts-related variables (continuation).

Variable	Description	Type	05	03	01	99
vcontab	Total value of contracts with another organization		X	X		
vcont	Total value of contracts		X	X	X	
raiscontes	Reason for contracting out: access to expertise/knowledge	lickert	X	X	X	
raiscontde	Reason for contracting out: faster completion of the work	lickert	X			
raiscontrf	Reason for contracting out: lower risk	lickert	X	X	X	
raiscontcm	Reason for contracting out: increase physical capacity	lickert	X			
raiscontdc	Reason for contracting out: activity area outside core competence	lickert	X	X	X	
raiscontrcrd	Reason for contracting out: R&D cost reduction	lickert	X	X	X	
raiscontrcrc	Reason for contracting out: regulatory/clinical affairs cost reduction	lickert	X	X	X	
raiscontrcp	Reason for contracting out: production cost reduction	lickert	X	X	X	
nacontaeb	Number of contracts provided to another biotechnology firm	integer	X	X		
nacontlp	Number of contracts provided to a private research lab	integer		X		
nacontep	Number of contracts provided to a pharmaceutical firm	integer	X			
naconteabp	Number of contracts provided to a firm other than biotechnology or pharmaceutical	integer	X			
nacontuh	Number of contracts provided to a university/hospital	integer	X	X		
nacontlg	Number of contracts provided to a government lab	integer	X	X		

Table A.10 : Contracts-related variables (end).

Variable	Description	Type	05	03	01	99
nacontlc	Number of contracts provided to a routine lab	integer			X	
nacontls	Number of contracts provided to a specialized lab	integer			X	
nacontp	Number of contracts provided to production/manufacturing services	integer			X	
naconta	Number of contracts provided to other organizations	integer	X		X	
nacont	Total number of provided contracts	integer	X	X	X	
racontaeb	Revenues from provided contracts to another biotechnology firm		X	X		
racontlp	Revenues from provided contracts to a private research lab			X		
racontep	Revenues from provided contracts to a pharmaceutical firm		X			
raconteabp	Revenues from provided contracts to a firm other than biotechnology or pharmaceutical		X			
racontuh	Revenues from provided contracts to a university/hospital		X	X		
racontlg	Revenues from provided contracts to a government lab		X	X		
racontlc	Revenues from provided contracts to a routine lab				X	
racontls	Revenues from provided contracts to a specialized lab				X	
racontp	Revenues from provided contracts to production/manufacturing services				X	
raconta	Revenues from provided contracts to other organizations		X		X	
racont	Total revenues from provided contracts		X		X	

Table A.11 : Variables related to Intellectual Property.

Variable	Description	Type	05	03	01	99
bb	The firm has biotechnology-related patents or pending patents	binary	X	X	X	X
nbe	Number of existing patents	integer	X	X	X	X
nba	Number of pending patents	integer	X	X	X	X
nbf	Number of expired patents	integer	X			
npb_1	Number of patented products/processes in the preceding year	integer	X			
npb_0	Number of patented products/processes in the current year	integer	X			
nbp_1	Number of submitted patents during the preceding year	integer	X	X	X	X
nbp_0	Number of submitted patents during the current year	integer	X	X	X	X
nba_1	Number of granted patents during the preceding year	integer	X			
nba_0	Number of granted patents during the current year	integer	X			
ndroitalec	Number of licensing agreements assigned/licensed to another Canadian firm	integer	X	X	X	
ndroitabec	Number of patent assignment assigned/licensed to another Canadian firm	integer	X	X	X	
ndroitatec	Number of technology transfer agreements assigned/licensed to another Canadian firm	integer	X	X	X	
ndroitaec	Total number of right assignments/licensing assigned/licensed to another Canadian firm	integer	X	X	X	

Table A.11 : Variables related to Intellectual Property (continuation).

Variable	Description	Type	05	03	01	99
ndroitaleus	Number of licensing agreements assigned/licensed to another U.S. firm	integer	X	X	X	
ndroitabeus	Number of patent assignment assigned/licensed to another U.S. firm	integer	X	X	X	
ndroitateus	Number of right assignments/licensing assigned/licensed to another U.S. firm	integer	X	X	X	
ndroitaeus	Total number of right assignments/licensing assigned/licensed to another U.S. firm	integer	X	X	X	
ndroitaleap	Number of licensing agreements assigned/licensed to a firm in another country	integer	X	X	X	
ndroitabeap	Number of patent assignment assigned/licensed to a firm in another country	integer	X	X	X	
ndroitateap	Number of right assignments/licensing assigned/licensed to a firm in another country	integer	X	X	X	
ndroitaeap	Total number of right assignments/licensing assigned/licensed to a firm in another country	integer	X	X	X	
ndroitale	Total number of assigned/licensed licensing agreements	integer	X	X	X	
ndroitabe	Total number of assigned/licensed patent assignments	integer	X	X	X	
ndroitate	Total number of assigned/licensed technology transfer agreements	integer	X	X	X	
ndroitae	Total number of assigned/licensed right assignments/licensing	integer	X	X	X	
rdroital	Total income from assigned/licensed licensing agreements		X	X	X	

Table A.11 : Variables related to Intellectual Property (continuation).

Variable	Description	Type	05	03	01	99
rdroitab	Total income from assigned/licensed patent assignments		X	X	X	
rdroitat	Total income from assigned/licensed technology transfer agreements		X	X	X	
rdroita	Total income from assigned/licensed right assignments/licensing		X	X	X	
ndroitolec	Number of licensing agreements acquired from another Canadian firm	integer	X	X	X	
ndroitobec	Number of patent assignments acquired from another Canadian firm	integer	X	X	X	
ndroitotec	Number of technology transfer agreements acquired from another Canadian firm	integer	X	X	X	
ndroitoec	Total number of right assignments/licensing acquired from another Canadian firm	integer	X	X	X	
ndroitoleus	Number of licensing agreements acquired from another U.S. firm	integer	X	X	X	
ndroitobeus	Number of patent assignments acquired from another U.S. firm	integer	X	X	X	
ndroitoteus	Number of technology transfer agreements acquired from another U.S. firm	integer	X	X	X	
ndroitoeus	Total number of right assignments/licensing acquired from another U.S. firm	integer	X	X	X	
ndroitoleap	Number of licensing agreements acquired from a firm in another country	integer	X	X	X	
ndroitobeap	Number of patent assignments acquired from a firm in another country	integer	X	X	X	
ndroitoteap	Number of technology transfer agreements acquired from a firm in another country	integer	X	X	X	

Table A.11 : Variables related to Intellectual Property (end).

Variable	Description	Type	05	03	01	99
ndroitoeap	Total number of right assignments/licensing acquired from a firm in another country	integer	X	X	X	
ndroitole	Total number of acquired licensing agreements	integer	X	X	X	
ndroitobe	Total number acquired patent assignments	integer	X	X	X	
ndroitote	Total number of acquired right assignments/licensing	integer	X	X	X	
ndroitoe	Total number of acquired right assignments/licensing	integer	X	X	X	
cdroitol	Total cost for acquiring licensing agreements	integer	X	X	X	
cdroitob	Total cost for acquiring patent assignments	integer	X	X	X	
cdroitot	Total cost for acquiring technology transfer agreements	integer	X	X	X	
cdroito	Total cost for acquiring IP rights	integer	X	X	X	

Table A.12 : Variables related to capital financing.

Variable	Description	Type	05	03	01	99
tentek	The firm attempted to raise capital for biotechnology-related purposes during the current year	binary	X	X	X	X
reusk	The firm was successful in raising capital during the current year	binary	X	X	X	X
fkreun	Total capital raised during the current year		X	X	X	X
objkatt	The firm reached its target in raising capital during the current year	binary	X	X	X	
raistkrd	Reason for attempting to raise capital: R&D purposes/expand R&D capacity	binary	X	X		
raistkremb	Reason for attempting to raise capital: repay current investors	binary	X	X		
raistkcommprd	Reason for attempting to raise capital: commercialize current R&D projects	binary	X	X		
raistkcr	Reason for attempting to raise capital: clinical/regulatory expenses	binary	X	X		
raistkp	Reason for attempting to raise capital: develop production/manufacturing capability	binary	X	X		
raistkcomm	Reason for attempting to raise capital: commercialization expenses	binary	X	X		
raisrefpad	Reason the lender/investor refused capital request: biotechnology products/processes not sufficiently developed	binary	X	X	X	
raisrefgtl	Reason the lender/investor refused capital request: biotechnology portfolio limited in scope	binary	X	X	X	
raisrefegi	Reason the lender/investor refused capital request: insufficient specific management skills/expertise	binary	X	X	X	
raisrefknd	Reason the lender/investor refused capital request: capital not available due to market conditions	binary	X	X	X	

Table A.12 : Variables related to capital financing (end).

Variable	Description	Type	05	03	01	99
raisrefndpa	Reason the lender/investor refused capital request: further product development required	binary	X	X	X	
raisrefid	Reason the lender/investor refused capital request: uncertainties of market demand for product	binary	X	X	X	
raisrefppd	Reason the lender/investor refused capital request: lender does not fun development projects	binary	X	X	X	
sourcekvcc	Percentage of funding provided by Canadian-based venture capital		X	X	X	
sourcekvcus	Percentage of funding provided by American-based venture capital		X	X	X	
sourcekvceu	Percentage of funding provided by European-based venture capital		X	X	X	X
sourcekvcap	Percentage of funding provided by venture capital from another country		X	X	X	
sourcekdet	Percentage of funding from debt capital		X	X	X	X
sourcekfam	Percentage of funding from angel investors/family		X	X	X	X
sourcekgov	Percentage of funding from government sources		X	X	X	X
sourcekgovper	Capital raised from government sources		X	X	X	X
sourcekpp	Percentage of funding from private placements		X	X	X	X
sourcekipo	Percentage of funding from Initial Public Offering (IPO)		X	X	X	X
sourcekspo	Percentage of funding from Secondary Public Offering (SPO)		X	X	X	X
sourcekvec	Percentage of funding from collaborative arrangements/alliances		X	X	X	X

Table A.13 : Variables related to import/export.

Variable	Description	Type	05	03	01	99
exportpb	The firm exports biotechnology products	binary	X	X	X	X
totrexp_1	Total export revenues during preceding year		X	X	X	X
totrexp_0	Total export revenues during current year		X	X	X	X
rexpperb_1	Percentage of export revenues from biotechnology products during preceding year		X	X	X	
rexpperb_0	Percentage of export revenues from biotechnology products during current year		X	X	X	X
importpb	The firm imports biotechnology products	binary	X	X	X	X
totdimp_1	Total import expenditures during preceding year		X	X	X	X
totdimp_0	Total import expenditures during current year		X	X	X	X
dimpperb_1	Percentage of import expenditures for biotechnology during preceding year		X	X	X	
dimpperb_0	Percentage of import expenditures for biotechnology during current year		X	X	X	X

Table A.14 : Variables related to business strategy.

Variable	Description	Type	05	03	01	99
strautilconi	Knowledge development strategies: knowledge captured from external industry sources	lickert	X	X	X	
strautilconul	Knowledge development strategies: public sources	lickert	X	X	X	
stradevconal	Knowledge development strategies: through collaborative arrangements	lickert	X	X	X	
strautilbd	Knowledge development strategies: databases of scientific information	lickert	X	X	X	
strapi	Knowledge development strategies: developed firm practices and policies for knowledge / IP	lickert	X	X	X	
straform	Knowledge development strategies: developed/encouraged staff education/upgrading	lickert	X	X	X	
stravpi	Knowledge development strategies: conducted IP audit to ensure product/processes protection	lickert	X	X	X	
strataille	Business strategies: increase size	lickert	X	X	X	
straract	Business strategies: downsize operations of the firm	lickert	X	X	X	
straservaent	Business strategies: provide R&D-based products to other firms	lickert	X	X	X	
straessai	Business strategies: increase market penetration	lickert	X	X	X	
stranouvrd	Business strategies: begin new R&D projects	lickert	X	X	X	
stramkthorsc	Business strategies: expand into foreign markets	lickert	X	X	X	

Table A.15 : Variables related to population by province.

Variable	Province Code	2005	2003	2001	1999
popprov	**59**	4257.8	4155.4	4078.4	4011.3
	48	3227.6	3161.4	3056.7	2953.3
	47	990	994.7	1000.1	1014.7
	46	1174.1	1161.9	1151.3	1142.5
	35	12558.6	12262.6	11897.6	11506.4
	24	7597.8	7494	7397	7323.3
	13	751.5	751.2	749.9	750.6
	12	936.1	936.5	932.4	933.8
	11	138.2	137.3	136.7	136.3
	10	514	518.4	522	533.4
popprov2	**59**	4257.8	4155.4	4078.4	4011.3
	48	3227.6	3161.4	3056.7	2953.3
	47	990	994.7	1000.1	1014.7
	46	1174.1	1161.9	1151.3	1142.5
	35	12558.6	12262.6	11897.6	11506.4
	24	7597.8	7494	7397	7323.3
	15	2339.8	2343.4	2341	2354.1
popprov3	**59**	4257.8	4155.4	4078.4	4011.3
	45	5391.7	5318	5208.1	5110.5
	35	12558.6	12262.6	11897.6	11506.4
	24	7597.8	7494	7397	7323.3
	15	2339.8	2343.4	2341	2354.1

Table A.16 : Biotechnology R&D personnel by province.

Variable	Description	Type	2005	2003	2001	1999
nebp	Total R&D personnel with biotechnology-related responsibilities, by province	integer	X	X	X	X

Table A.17 : Number of universities, by province.

Variable	Description	Type	Province code	2005	2003	2001	1999
nunivp	Number of universities by province	integer	59	9	9	9	9
			48	12	6	6	6
			47	2	2	2	2
			46	7	4	4	4
			35	20	20	20	20
			24	19	19	19	19
			13	9	4	4	4
			12	11	10	10	10
			11	1	1	1	1
			10	1	1	1	1
nunivp2	Number of universities by province2	integer	59	9	9	9	9
			48	12	6	6	6
			47	2	2	2	2
			46	7	4	4	4
			35	20	20	20	20
			24	19	19	19	19
			15	22	16	16	16
nunivp3	Number of universities by province3	integer	59	9	9	9	9
			45	21	12	12	12
			35	20	20	20	20
			24	19	19	19	19
			15	22	16	16	16

Table A.18 : Biotechnology patents, by province[27].

Variable	Description	Type	Province code	2005	2003	2001	1999
nbprov	Total assigned biotechnology-related patents by province	integer	59	232	183	133	94
			48	264	228	177	125
			47	91	78	64	50
			46	26	20	16	11
			35	1095	963	749	525
			24	490	432	330	226
			13	3	2	2	2
			12	18	14	9	4
			11	4	4	4	4
			10	1	1	1	1
nbprov2	Total assigned biotechnology-related patents by province2	integer	59	232	183	133	94
			48	264	228	177	125
			47	91	78	64	50
			46	26	20	16	11
			35	1095	963	749	525
			24	490	432	330	226
			15	26	21	16	11
nbprov3	Total assigned biotechnology-related patents by province3	integer	59	232	183	133	94
			45	381	326	257	186
			35	1095	963	749	525
			24	490	432	330	226
			15	26	21	16	11

In order to introduce a control for the environment, in each model is employed the variable:

```
nbppopprov = nbprov/popprov
```

which represents the number of biotechnology-related assigned patents for each province, divided by the population of that province.

[27] Data drawn from the USPTO website.

ANNEX B

REGRESSION MODELS

In the following, the models employed for each survey are presented. Each model is fitted with three variations, depending on which subsample is considered:

- Small firms only
- Small and medium firms only
- Full sample: small, medium and large firms

The dependent variables used are those gathered in Table 2.3 (binary dependent variable for logit models) and Table 2.4 (count variables for negative binomial regression). The models include independent variables in order to measure size, biotechnology orientation, level of innovativeness, biotechnology products/processes, IP protection, export orientation and public support. In the following, all the models implemented are presented.

B.1 FIRST MODEL

The first model uses the following independent variables:

- `e` as a measure of the size of the firm
- `ebper` as a measure of the biotechnology-orientation
- `dspin` and `dsubs` as a measure of the type of firm
- `dprod` and `dproc` as a measure of innovativeness
- `nbe` and `nba` as a measure of the use of IP protection
- `rexpperb_0` as a measure of export activity
- `sourcekgov` and `lsourcekgovtot` as a measure of public funding
- `nbppopprov` to control for the environment

The vector of the parameters for the first model (2005 survey) is:

$\underline{\theta}_{2005}$ = (e ebper dspin dsubs dprod dproc nbe nba rexpperb_0 sourcekgov lsourcekgovtot nbppopprov)

As in the 2003, 2001 and 1999 surveys the variable dsubs is not available, the parameters vector reduces to the following:

$\underline{\theta}_{2003} = \underline{\theta}_{2001} = \underline{\theta}_{1999}$ = (e ebper dspin dprod dproc nbe nba rexpperb_0 sourcekgov lsourcekgovtot nbppopprov)

Only for the first model, and only for the 2005 survey, we report the commands used in STATA 10 to estimate the regressions.

A. Logistic regression using the whole sample (dependent variable ec):

```
svy: logit ec e ebper dspin dprod dproc dsubs ///
  nbe nba rexpperb_0 sourcekgov lsourcekgovtot nbppopprov
```

B. Logistic regression using the subsample including small and medium firms only (dependent variable ec):

```
svy,subpop(dpme):logit ec e ebper dspin dprod dproc dsubs ///
  nbe nba rexpperb_0 sourcekgov lsourcekgovtot nbppopprov
```

C. Logistic regression using the subsample including small firms only (dependent variable ec):

```
svy,subpop(dpetite):logit ec e ebper dspin dprod dproc dsubs ///
  nbe nba rexpperb_0 sourcekgov lsourcekgovtot nbppopprov
```

D. Negative binomial regression using the whole sample (dependent variable nec):

```
svy: nbreg nec e ebper dspin dprod dproc dsubs ///
  nbe nba rexpperb_0 sourcekgov lsourcekgovtot nbppopprov
```

E. Negative binomial regression using the subsample including small and medium firms only (dependent variable nec):

```
svy,subpop(dpme):nbreg nec e ebper dspin dprod dproc dsubs ///
  nbe nba rexpperb_0 sourcekgov lsourcekgovtot nbppopprov
```

F. Negative binomial regression using the subsample including small firms only (dependent variable nec):

```
svy,subpop(dpetite):nbreg nec e ebper dspin dprod ///
 dproc dsubs nbe nba rexpperb_0 sourcekgov ///
 lsourcekgovtot nbppopprov
```

B.2 SECOND MODEL

The second model uses the following independent variables:

- lrt_0 as a measure of the size of the firm
- rbper0 as a measure of the biotechnology-orientation
- dspin and dsubs as a measure of the type of firm
- dprod and dproc as a measure of innovativeness
- nbe and nba as a measure of the use of IP protection
- rexpperb_0 as a measure of export activity
- sourcekgov and lsourcekgovtot as a measure of public funding
- nbppopprov to control for the environment

Therefore, the parameters vector for the 2005 survey is:

$\underline{\theta}_{2005}$=(lrt_0 rbper0 dspin dsubs dprod dproc nbe nba rexpperb_0 sourcekgov lsourcekgovtot nbppopprov)

As in the other surveys the variable dsubs in not available in the other surveys, the parameters vector reduces to the following:

$\underline{\theta}_{2003} = \underline{\theta}_{2001} = \underline{\theta}_{1999} =$ (lrt_0 rbper0 dspin dprod dproc nbe nba rexpperb_0 sourcekgov lsourcekgovtot nbppopprov)

B.3 THIRD MODEL

The third model uses the following independent variables:

- e as a measure of the size of the firm
- ebper as a measure of the biotechnology-orientation
- dspin and dsubs as a measure of the type of firm
- nprodrd, nprodpc, nprodrc and nprodpm as a measure of innovativeness
- nbe and nba as a measure of the use of IP protection
- rexpperb_0 as a measure of export activity
- sourcekgov and lsourcekgovtot as a measure of public funding
- nbppopprov to control for the environment

This way, the parameters vector for the third model is (2005 survey):

$\underline{\theta}_{2005} =$ (e ebper dspin dsubs nprodrd nprodpc nprodrc nprodpm nbe nba rexpperb_0 sourcekgov lsourcekgovtot nbppopprov)

As in the other surveys the variable dsubs in not available in the other surveys, the parameters vector reduces to the following:

$\underline{\theta}_{2003} = \underline{\theta}_{2001} = \underline{\theta}_{1999} =$ (e ebper dspin nprodrd nprodpc nprodrc nprodpm nbe nba rexpperb_0 sourcekgov lsourcekgovtot nbppopprov)

B.4 FOURTH MODEL

The fourth model uses the following independent variables:

- `e` as a measure of the size of the firm
- `ebper` as a measure of the biotechnology-orientation
- `dspin` and `dsubs` as a measure of the type of firm
- `dprod` and `dproc` as a measure of innovativeness
- `ndroitaec` and `ndroitoec` as a measure of the use of IP protection
- `rexpperb_0` as a measure of export activity
- `sourcekgov` and `lsourcekgovtot` as a measure of public funding
- `nbppopprov` to control for the environment

This way, the parameters vector for the third model is (2005 survey):

$\underline{\theta}_{2005}$ = (e ebper dspin dsubs dprod dproc ndroitaec ndroitoec rexpperb_0 sourcekgov lsourcekgovtot nbppopprov)

As in the other surveys the variable `dsubs` in not available in the other surveys, the parameters vector reduces to the following:

$\underline{\theta}_{2003} = \underline{\theta}_{2001}$ = (e ebper dspin dprod dproc ndroitaec ndroitoec rexpperb_0 sourcekgov lsourcekgovtot nbppopprov)

In the 1999 survey, the variables `ndroitaec` and `ndroitoec` are not available, and the fourth model thus collapses.

B.5 FIFTH MODEL

The fifth model uses the following independent variables:

- e as a measure of the size of the firm
- ebper as a measure of the biotechnology-orientation
- dspin and dsubs as a measure of the type of firm
- dprod and dproc as a measure of innovativeness
- ncont and nacont as a measure of contract-related activities
- rexpperb_0 as a measure of export activity
- sourcekgov and lsourcekgovtot as a measure of public funding
- nbppopprov to control for the environment

This way, the parameters vector for the third model is (2005 survey):

$\underline{\theta}_{2005}$ = (e ebper dspin dsubs dprod dproc ncont nacont rexpperb_0 sourcekgov lsourcekgovtot nbppopprov)

As in the other surveys the variable dsubs in not available in the other surveys, the parameters vector reduces to the following:

$\underline{\theta}_{2003} = \underline{\theta}_{2001}$ = (e ebper dspin dprod dproc ncont nacont rexpperb_0 sourcekgov lsourcekgovtot nbppopprov)

In the 1999 survey, the variables ncont and nacont are not available, and the fourth model thus collapses.

B.6 SIXTH MODEL

The sixth model uses the following independent variables:

- `e` as a measure of the size of the firm
- `ebper` and `rdbper0` as a measure of the biotechnology-orientation
- `dspin` and `dsubs` as a measure of the type of firm
- `lrdt_0` as a measure of innovativeness
- `rexpperb_0` as a measure of export activity
- `sourcekgov` and `lsourcekgovtot` as a measure of public funding
- `nbppopprov` to control for the environment

This way, the parameters vector for the third model is (2005 survey):

$\underline{\theta}_{2005}$ = (e ebper rdbper0 dspin dsubs lrdt_0 rexpperb_0 sourcekgov lsourcekgovtot nbppopprov)

As in the other surveys the variable `dsubs` in not available in the other surveys, the parameters vector reduces to the following:

$\underline{\theta}_{2003} = \underline{\theta}_{2001} = \underline{\theta}_{1999}$ = (e ebper rdbper0 dspin lrdt_0 rexpperb_0 sourcekgov lsourcekgovtot nbppopprov)

B.7 SEVENTH MODEL

The seventh models aims at evaluating the effect of whether firms were successful at raising capital (variable `reusk`) and of the total capital raised (variable `fkreun`) on the propensity to collaborate. The seventh model differs consistently from the others, because it has been created to respond to the precise need of investigating on the relation-

ship between firms' attempts to raise capital and collaborative behavior. This model uses the following independent variables:

- `e` as a measure of the size of the firm
- `ebper` as a measure of the biotechnology-orientation
- `dspin` and `dsubs` as a measure of the type of firm
- `reusk` to measure the firm successfulness in raising capital
- `lfkreun` as a measure of the total raised capital
- `rexpperb_0` as a measure of export activity
- `sourcekgov` as a measure of public funding
- `nbppopprov` to control for the environment

This way, the parameters vector for the third model is (2005 survey):

$\underline{\theta}_{2005}$ = (e ebper dspin dsubs reusk lfkreun rexpperb_0 sourcekgov nbppopprov)

As in the other surveys the variable `dsubs` in not available in the other surveys, the parameters vector reduces to the following:

$\underline{\theta}_{2003} = \underline{\theta}_{2001} = \underline{\theta}_{1999}$ = (e ebper dspin reusk lfkreun rexpperb_0 sourcekgov nbppopprov)

ANNEX C

SUMMARY STATISTICS TABLES

Table C.1 : Total number of biotech firms by size and province, 1999 to 2005.

Year	Size	Atlantic provinces	Québec	Ontario	Prairie provinces	British Columbia
1999	Small	18	66	83	30	63
	Medium	n/a	27	16	12	5
	Large	n/a	13	12	9	3
2001	Small	18	88	71	38	53
	Medium	0	22	17	12	10
	Large	7	18	13	5	4
2003	Small	22	104	92	69	65
	Medium	n/a	29	13	16	17
	Large	n/a	13	24	14	9
2005	Small	24	133	99	59	82
	Medium	n/a	32	26	15	8
	Large	0	16	19	14	3

Table C.2 : Total number of collaborative biotech firms by size and province, 1999 to 2005.

Year	Size	Atlantic provinces	Québec	Ontario	Prairie provinces	British Columbia
1999	Small	13	44	26	19	42
	Medium	n/a	24	12	9	4
	Large	n/a	11	9	9	2
2001	Small	10	50	43	19	34
	Medium	n/a	12	7	6	7
	Large	3	17	8	5	4
2003	Small	11	51	45	34	36
	Medium	n/a	12	7	13	5
	Large	n/a	8	7	14	5
2005	Small	20	66	48	26	43
	Medium	n/a	18	13	13	3
	Large	n/a	11	6	12	2

Table C.3 : Percentage of biotechnology firms involved in at least one collaborative arrangement, by size and province, 1999 to 2005.

Year	Size	Atlantic provinces	Québec	Ontario	Prairie provinces	British Columbia
1999	Small	71.9%	67.2%	31.9%	64.2%	66.5%
	Medium	n/a	87.2%	72.1%	78.7%	74.9%
	Large	n/a	82.4%	74.1%	100.0%	66.7%
2001	Small	58.7%	57.3%	60.3%	50.4%	65.1%
	Medium	n/a	53.1%	41.0%	49.4%	70.7%
	Large	49.0%	92.2%	65.8%	100.0%	100.0%
2003	Small	50.3%	49.2%	48.4%	49.9%	54.5%
	Medium	n/a	41.2%	51.5%	79.1%	29.0%
	Large	n/a	61.8%	30.8%	100.0%	61.1%
2005	Small	82.3%	49.5%	48.5%	43.4%	51.9%
	Medium	n/a	55.9%	48.3%	88.5%	42.0%
	Large	n/a	66.4%	32.4%	88.1%	51.6%

Table C.4 : Number of collaborative arrangements, by firm size and province, 1999 to 2005.

Year	Size	Atlantic provinces	Québec	Ontario	Prairie provinces	British Columbia
1999	Small	23	152	49	35	135
	Medium	0	56	35	25	14
	Large	0	63	55	31	13
2001	Small	23	186	168	58	185
	Medium	0	48	26	25	56
	Large	5	194	118	34	22
2003	Small	33	178	148	97	112
	Medium	2	63	153	33	33
	Large	2	24	92	52	10
2005	Small	60	337	256	96	204
	Medium	0	75	180	46	28
	Large	0	65	16	41	23

Table C.5 : Number of collaborative arrangements per collaborative firm by size and province, 1999-2005.

Year	Size	Atlantic provinces	Québec	Ontario	Prairie provinces	British Columbia
	Small	1.82	3.42	1.84	2.18	3.24
1999	Medium	n/a	2.39	2.99	2.73	3.70
	Large	n/a	5.75	6.14	3.52	6.50
	Small	2.27	3.69	3.95	3.03	5.41
2001	Medium	n/a	4.09	3.71	4.24	7.70
	Large	1.44	11.52	14.05	7.27	5.05
	Small	2.99	3.46	3.31	2.83	3.13
2003	Medium	n/a	5.40	23.39	2.57	6.59
	Large	n/a	3.00	12.35	3.76	1.98
	Small	3.06	5.12	5.34	3.76	4.79
2005	Medium	n/a	4.18	14.38	3.42	8.03
	Large	n/a	6.08	2.62	3.33	15.00

Table C.6 : Contingency table for variable `nec` : total number of partnerships, 1999.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	55.9%	20.8%	23.3%
Ontario	34.9%	25.4%	39.7%
Prairie Provinces	42.9%	25.5%	31.6%
British Columbia	83.3%	8.7%	8.0%
Total	57.8%	18.9%	23.4%

Table C.7 : Contingency table for variable nec : total number of partnerships, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	83.2%	n/a	n/a
Québec	43.5%	11.2%	45.3%
Ontario	53.8%	8.4%	37.7%
Prairie Provinces	49.4%	21.8%	28.9%
British Columbia	70.3%	21.3%	8.3%
Total	54.0%	13.6%	32.4%

Table C.8 : Contingency table for variable nec : total number of partnerships, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	91.2%	n/a	n/a
Québec	66.9%	23.9%	9.2%
Ontario	37.6%	38.9%	23.5%
Prairie Provinces	53.0%	18.4%	28.6%
British Columbia	72.1%	21.2%	6.7%
Total	54.9%	27.6%	17.5%

Table C.9 : Contingency table for variable nec : total number of partnerships, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	70.8%	15.7%	13.6%
Ontario	56.6%	39.8%	3.6%
Prairie Provinces	52.7%	25.0%	22.3%
British Columbia	79.9%	11.0%	9.1%
Total	66.8%	23.0%	10.2%

Table C.10 : Total number of firms collaborating for knowledge-related reasons (variable ecc), by size and province, 1999.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	12.8	0.0	0.0
Québec	42.0	23.7	11.0
Ontario	24.0	11.8	9.0
Prairie Provinces	16.8	9.1	8.8
British Columbia	40.4	3.8	2.0
Total	135.9	48.4	30.8

Table C.11 : Total number of firms collaborating for knowledge-related reasons (variable ecc), by size and province, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	10.3	0.0	3.3
Québec	46.3	10.2	15.3
Ontario	38.3	7.1	8.4
Prairie Provinces	16.1	6.0	4.6
British Columbia	29.6	7.3	4.3
Total	140.6	30.6	35.9

Table C.12 : Total number of firms collaborating for knowledge-related reasons (variable ecc), by size and province, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	1.5	0.0	0.0
Québec	32.8	10.1	4.8
Ontario	38.5	4.8	5.9
Prairie Provinces	20.0	5.0	6.6
British Columbia	24.9	5.0	5.2
Total	117.6	25.0	22.4

Table C.13 : Total number of firms collaborating for knowledge-related reasons (variable ecc), by size and province, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	16.7	0.0	0.0
Québec	62.8	17.9	10.6
Ontario	46.4	12.5	6.3
Prairie Provinces	24.1	13.4	12.2
British Columbia	39.9	3.5	0.0
Total	189.8	47.2	29.2

Table C.14 : Total number of knowledge-related alliances (variable necc), by size and province, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	21.9	0.0	4.7
Québec	127.3	32.7	191.9
Ontario	111.2	24.8	73.7
Prairie Provinces	44.5	21.1	27.7
British Columbia	161.1	50.3	18.9
Total	466.0	128.8	316.8

Table C.15 : Total number of knowledge-related alliances (variable necc), by size and province, 2003 survey.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	1.6	0.0	1.5
Québec	88.2	21.7	14.6
Ontario	110.6	73.7	32.1
Prairie Provinces	43.0	11.8	22.3
British Columbia	62.4	29.6	5.2
Total	305.8	136.9	75.7

Table C.16 : Total number of knowledge-related alliances (variable necc), by size and province, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	34.3	0.0	0.0
Québec	275.6	63.3	60.2
Ontario	186.7	169.2	16.0
Prairie Provinces	64.6	43.7	29.7
British Columbia	160.3	22.7	0.0
Total	721.5	299.0	105.8

Table C.17 : Total number of firms collaborating for production/commercialization-related reasons (variable ecpc), by size and province, 1999.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	7.8	0.0	0.0
Québec	26.9	21.2	5.8
Ontario	20.0	2.6	2.5
Prairie Provinces	8.3	4.2	2.3
British Columbia	24.9	1.5	1.0
Total	87.8	29.4	11.7

Table C.18 : Total number of firms collaborating for production/commercialization-related reasons (variable ecpc), by size and province, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	0.0	0.0	0.0
Québec	23.6	1.5	0.0
Ontario	18.7	1.6	4.2
Prairie Provinces	8.4	1.4	1.8
British Columbia	8.9	2.9	1.5
Total	59.6	7.5	7.6

Table C.19 : Total number of firms collaborating for production/commercialization-related reasons (variable ecpc), by size and province, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	4.6	0.0	0.0
Québec	26.2	3.8	6.4
Ontario	13.8	3.4	3.1
Prairie Provinces	17.2	3.6	1.6
British Columbia	18.5	1.6	1.7
Total	80.3	12.4	12.8

Table C.20 : Total number of firms collaborating for production/commercialization-related reasons (variable ecpc), by size and province, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	10.0	0.0	0.0
Québec	17.7	3.5	3.1
Ontario	16.8	3.4	1.2
Prairie Provinces	9.0	3.8	9.2
British Columbia	18.9	1.8	1.6
Total	72.4	12.5	15.0

Table C.21 : Total number of production/commercialization-related alliances (variable necpc), by size and province, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	0.0	0.0	0.0
Québec	57.3	15.1	0.0
Ontario	57.2	1.6	44.4
Prairie Provinces	13.1	4.3	1.8
British Columbia	15.0	5.9	3.0
Total	142.6	27.0	49.2

Table C.22 : Total number of production/commercialization-related alliances (variable necpc), by size and province, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	9.3	0.0	0.0
Québec	61.2	20.9	9.8
Ontario	29.2	75.7	12.3
Prairie Provinces	32.9	5.6	7.7
British Columbia	31.0	1.6	3.4
Total	163.7	103.7	33.2

Table C.23 : Total number of production/commercialization-related alliances (variable necpc), by size and province, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	26.3	0.0	0.0
Québec	42.2	11.2	4.4
Ontario	27.8	10.8	0.5
Prairie Provinces	31.9	2.0	11.1
British Columbia	43.5	5.3	23.3
Total	171.7	29.3	39.3

Table C.24 : Total number of firms collaborating with another biotechnology firm (variable eceb), by size and province, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	3.3	n/a	1.8
Québec	22.0	5.9	6.2
Ontario	23.2	4.1	7.0
Prairie Provinces	9.0	4.6	3.2
British Columbia	16.2	7.3	4.3
Total	73.7	21.9	22.5

Table C.25 : Total number of firms collaborating with another biotechnology firm (variable eceb), by size and province, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	1.6	1.6	0
Québec	13.8	5.4	0
Ontario	15.4	1.7	3
Prairie Provinces	17.0	6.9	3.4
British Columbia	16.9	3.3	1.7
Total	64.7	18.9	8.1

Table C.26 : Total number of firms collaborating with another biotechnology firm (variable eceb), by size and province, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	3.5	0	n/a
Québec	28.2	11.7	7.7
Ontario	18.8	6.3	3.6
Prairie Provinces	14.8	3.0	3.0
British Columbia	25.3	1.8	0
Total	90.6	22.8	14.3

Table C.27 : Total number of firms collaborating with another non-biotechnology firm (variable eceab), by size and province, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	1.1	n/a	0
Québec	9.3	0	0
Ontario	3.0	0	0
Prairie Provinces	2.6	0	0
British Columbia	4.5	0	0
Total	20.6	0	0

Table C.28 : Total number of firms collaborating with another non-biotechnology firm (variable eceab), by size and province, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	1.5	0	0
Québec	8.0	0	1.7
Ontario	9.3	3.4	0
Prairie Provinces	0	0	0
British Columbia	3.1	0	0
Total	21.9	3.4	1.7

Table C.29 : Total number of firms collaborating with a pharmaceutical firm (variable ecep), by size and province, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	5.4	0	n/a
Québec	4.7	5.9	3.1
Ontario	12.4	2.7	1.8
Prairie Provinces	4.2	3.7	1.5
British Columbia	6.9	0	1.6
Total	33.6	12.3	8.0

Table C.30 : Total number of firms collaborating with another non-biotechnology and non/pharmaceutical firm (variable ecabp), by size and province, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	8.8	0	n/a
Québec	16.3	0	4.9
Ontario	10.4	4.4	1.8
Prairie Provinces	4.0	1.5	4.6
British Columbia	9.8	0	0
Total	49.3	5.9	11.3

Table C.31 : Total number of firms collaborating with another firm (variable `ecepri`), by size and province, 1999.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	8.8	0	0
Québec	18.5	17.8	4.7
Ontario	11.1	4.4	3.7
Prairie Provinces	8.3	8.3	1
British Columbia	19.2	1	0
Total	65.8	31.6	9.4

Table C.32 : Total number of firms collaborating with another firm (variable `ecepri`), by size and province, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	4.4	0	1.8
Québec	31.3	5.9	6.2
Ontario	26.2	4.1	7.0
Prairie Provinces	11.6	4.6	3.2
British Columbia	20.7	7.3	4.3
Total	94.3	21.8	22.6

Table C.33 : Total number of firms collaborating with another firm (variable `ecepri`), by size and province, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	3.1	1.6	0
Québec	21.9	5.4	1.7
Ontario	24.7	5.0	3
Prairie Provinces	17.0	6.9	3.4
British Columbia	20.0	3.3	1.7
Total	86.5	22.3	9.8

Table C.34 : Total number of firms collaborating with another firm (variable `ecepri`), by size and province, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	16.6	0	0
Québec	44.3	13.2	9.2
Ontario	31.0	7.8	3.6
Prairie Provinces	18.4	6.7	6.1
British Columbia	33.5	1.8	1.6
Total	143.8	29.5	20.4

Table C.35 : Total number of firms collaborating with a hospital/university (variable `ecuh`), by size and province, 1999.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	4.1	n/a	0
Québec	30.4	11.4	7.3
Ontario	14.3	8.3	8
Prairie Provinces	8.5	2	6.3
British Columbia	17.5	3.8	2
Total	72.7	25.5	23.6

Table C.36 : Total number of firms collaborating with a hospital/university (variable `ecuh`), by size and province, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	7.8	n/a	0
Québec	11.8	2.9	9.1
Ontario	16.0	2.7	5.6
Prairie Provinces	11.6	0	3.3
British Columbia	9.0	1.4	1.7
Total	56.2	7	19.7

Table C.37 : Total number of firms collaborating with a hospital/university (variable ecuh), by size and province, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	1.6	0	0
Québec	10.7	1.9	3.4
Ontario	22.8	1.7	5.9
Prairie Provinces	12.2	0	3.1
British Columbia	9.3	5.0	1.8
Total	56.6	8.6	14.2

Table C.38 : Total number of firms collaborating with a hospital/university (variable ecuh), by size and province, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	3.0	0	n/a
Québec	33.3	8.8	6.3
Ontario	31.3	9.1	3.3
Prairie Provinces	8.9	4.1	7.3
British Columbia	19.0	3.5	0
Total	95.9	25.5	16.9

Table C.39 : Total number of firms collaborating with a government laboratory (variable eclg), by size and province, 1999.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	5.5	n/a	0
Québec	9.0	9.9	8.0
Ontario	6.8	9.6	2.2
Prairie Provinces	3	5.4	7.3
British Columbia	20.5	1	1
Total	44.8	25.9	18.5

Table C.40 : Total number of firms collaborating with a government laboratory (variable `eclg`), by size and province, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	3.3	n/a	0
Québec	14.7	1.4	4.7
Ontario	12.8	2.5	4.2
Prairie Provinces	3.0	1.4	1.4
British Columbia	11.9	0	1.7
Total	45.7	5.3	12.0

Table C.41 : Total number of firms collaborating with a government laboratory (variable `eclg`), by size and province, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	1.5	0	0
Québec	10.6	4.7	3.4
Ontario	3.0	0	4.3
Prairie Provinces	0	0	6.6
British Columbia	4.7	0	0
Total	19.8	4.7	14.3

Table C.42 : Total number of firms collaborating with a government laboratory (variable `eclg`), by size and province, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	8.6	0	n/a
Québec	17.9	6.1	6.6
Ontario	11.1	5.9	3.0
Prairie Provinces	7.9	7.0	10.7
British Columbia	11.6	0	0
Total	57.1	19.0	20.3

Table C.43 : Total number of firms collaborating with a public institution (variable `ecipub`), by size and province, 1999.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	7.5	0	0
Québec	31.9	19.3	9.0
Ontario	18.5	10.8	9.0
Prairie Provinces	9.5	5.4	7.3
British Columbia	31.4	3.8	2.0
Total	98.7	39.3	27.3

Table C.44 : Total number of firms collaborating with a public institution (variable `ecipub`), by size and province, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	7.8	0	0
Québec	20.5	4.3	9.1
Ontario	20.4	4.1	7.0
Prairie Provinces	11.6	1.4	3.3
British Columbia	16.5	1.4	1.7
Total	76.8	11.2	21.0

Table C.45 : Total number of firms collaborating with a public institution (variable `ecipub`), by size and province, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	3.1	0	0
Québec	18.2	6.6	5.1
Ontario	22.8	1.7	5.9
Prairie Provinces	12.2	0	6.6
British Columbia	10.9	5.0	1.8
Total	67.1	13.2	19.3

Table C.46 : Total number of firms collaborating with a public institution (variable `ecipub`), by size and province, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	11.6	0	0
Québec	46.8	11.5	8.0
Ontario	36.3	10.5	4.5
Prairie Provinces	14.0	9.7	12.2
British Columbia	23.6	3.5	0
Total	132.3	35.1	24.8

Table C.47 : Contingency table for variable `neceb` : total number of partnerships with another biotech firm, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	68.8%	20.2%	11.0%
Ontario	63.2%	8.6%	28.2%
Prairie Provinces	34.4%	40.7%	24.9%
British Columbia	66.7%	25.6%	6.8%
Total	62.5%	21.6%	15.9%

Table C.48 : Contingency table for variable `neceb` : partnerships with another biotech firm, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	n/a	n/a	n/a
Ontario	26.5%	64.9%	8.6%
Prairie Provinces	61.5%	14.0%	24.4%
British Columbia	71.8%	21.0%	7.2%
Total	54.9%	35.3%	9.8%

Table C.49 : Contingency table for variable neceb : partnerships with another biotech firm, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	76.4%	17.7%	5.9%
Ontario	53.7%	40.7%	6.7%
Prairie Provinces	70.2%	24.5%	5.4%
British Columbia	87.2%	n/a	n/a
Total	71.6%	23.9%	4.5%

Table C.50 : Contingency table for variable eccc : collaboration to gain access to external knowledge/skill, 1999.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	45.8%	41.9%	12.3%
Ontario	49.8%	28.8%	21.5%
Prairie Provinces	34.5%	46.6%	18.9%
British Columbia	85.3%	10.3%	4.4%
Total	56.9%	30.5%	12.6%

Table C.51 : Contingency table for variable eccc : collaboration to gain access to external knowledge/skill, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	77.0%	12.84%	10.2%
Ontario	64.2%	28.5%	7.3%
Prairie Provinces	52.3%	13.0%	34.7%
British Columbia	85.5%	14.5%	0.0%
Total	69.0%	16.9%	14.1%

Table C.52 : Contingency table for variable eccnd : collaboration to gain access to external knowledge, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	59.7%	29.6%	10.7%
Ontario	82.6%	9.0%	8.5%
Prairie Provinces	65.2%	10.0%	24.9%
British Columbia	68.6%	10.2%	21.2%
Total	70.4%	15.3%	14.3%

Table C.53 : Number of firms collaborating for knowledge-related reasons, 1999.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	12.7	0	0
Québec	42.0	23.7	11.0
Ontario	24.0	11.8	9.0
Prairie Provinces	16.8	9.1	8.8
British Columbia	40.4	3.8	2
Total	135.9	48.4	30.8

Table C.54 : Contingency table for variable ecc : knowledge-related collaboration, 1999.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	54.8%	30.9%	14.3%
Ontario	53.5%	26.4%	20.1%
Prairie Provinces	48.3%	26.4%	25.3%
British Columbia	87.4%	8.3%	4.3%
Total	63.2%	22.5%	14.3%

Table C.55 : Number of firms collaborating for knowledge-related reasons, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	10.3	0	3.3
Québec	46.3	10.2	15.3
Ontario	38.3	7.1	8.4
Prairie Provinces	16.1	6.0	4.6
British Columbia	29.6	7.3	4.3
Total	140.6	30.6	35.9

Table C.56 : Contingency table for variable ecc : knowledge-related collaboration, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	64.5%	14.2%	21.3%
Ontario	71.2%	13.2%	15.6%
Prairie Provinces	60.3%	22.4%	17.3%
British Columbia	71.8%	17.7%	10.5%
Total	67.9%	14.8%	17.3%

Table C.57 : Number of firms collaborating for knowledge-related reasons, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	1.5	0	0
Québec	32.8	10.1	4.8
Ontario	38.5	4.8	5.9
Prairie Provinces	20.0	5.0	6.6
British Columbia	24.9	5.0	5.2
Total	117.6	25.0	22.4

Table C.58 : Contingency table for variable ecc : knowledge-related collaboration, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	68.8%	21.3%	10.0%
Ontario	78.2%	9.9%	11.9%
Prairie Provinces	63.4%	15.8%	20.8%
British Columbia	70.9%	14.2%	14.9%
Total	71.3%	15.1%	13.6%

Table C.59 : Total number of firms collaborating for knowledge-related reasons, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	16.7	0	0
Québec	62.8	17.9	10.6
Ontario	46.4	12.5	6.3
Prairie Provinces	24.1	13.4	12.2
British Columbia	39.9	3.5	0
Total	189.8	47.2	29.2

Table C.60 : Contingency table for variable ecc : knowledge-related collaboration, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	n/a	n/a	n/a
Québec	68.8%	19.6%	11.7%
Ontario	71.1%	19.2%	9.7%
Prairie Provinces	48.5%	26.9%	24.6%
British Columbia	92.0%	8.0%	0%
Total	71.3%	17.7%	11.0%

Table C.61 : Number of firms collaborating for production/commercialization-related reasons, 1999.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	7.8	0	0
Québec	26.9	21.2	5.8
Ontario	20.0	2.6	2.5
Prairie Provinces	8.3	4.2	2.3
British Columbia	24.9	1.5	1
Total	87.8	29.4	11.7

Table C.62 : Contingency table for variable ecpc : production/commercialization-related collaboration, 1999.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	100.0%	0.0%	0.0%
Québec	49.9%	39.3%	10.8%
Ontario	79.9%	10.2%	10.0%
Prairie Provinces	55.9%	28.4%	15.8%
British Columbia	90.9%	5.5%	3.7%
Total	68.1%	22.8%	9.0%

Table C.63 : Number of firms collaborating for production/commercialization-related reasons, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	0	0	0
Québec	23.6	1.5	0
Ontario	18.7	1.6	4.2
Prairie Provinces	8.4	1.4	1.8
British Columbia	8.9	2.9	1.5
Total	59.6	7.5	7.6

Table C.64 : Contingency table for variable ecpc : production/commercialization-related collaboration, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	0.0%	0.0%	0.0%
Québec	94.0%	6.0%	0%
Ontario	76.3%	6.5%	17.2%
Prairie Provinces	71.8%	12.4%	15.8%
British Columbia	67.0%	21.7%	11.3%
Total	79.9%	10.0%	10.1%

Table C.65 : Number of firms collaborating for production/commercialization-related reasons, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	4.6	0.0	0.0
Québec	26.2	3.8	6.4
Ontario	13.8	3.4	3.1
Prairie Provinces	17.2	3.6	1.6
British Columbia	18.5	1.6	1.7
Total	80.3	12.4	12.8

Table C.66 : Contingency table for variable ecpc : production/commercialization-related collaboration, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	100.0%	0.0%	0.0%
Québec	71.9%	10.4%	17.7%
Ontario	68.0%	16.6%	15.3%
Prairie Provinces	76.8%	16.0%	7.2%
British Columbia	84.7%	7.5%	7.8%
Total	76.1%	11.7%	12.2%

Table C.67 : Number of firms collaborating for production/commercialization-related reasons, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	10.0	0.0	0.0
Québec	17.7	3.5	3.1
Ontario	16.8	3.4	1.2
Prairie Provinces	9.0	3.8	9.2
British Columbia	18.9	1.8	1.6
Total	72.4	12.5	15.0

Table C.68 : Contingency table for variable ecpc : production/commercialization-related collaboration, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	100.0%	0.0%	0.0%
Québec	72.9%	14.4%	12.7%
Ontario	78.4%	15.9%	5.7%
Prairie Provinces	40.9%	17.4%	41.7%
British Columbia	85.2%	7.9%	7.0%
Total	72.5%	12.5%	15.1%

Table C.69 : Total number of knowledge-related collaborative arrangements by province and firm size, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	21.9	0.0	4.7
Québec	127.3	32.7	191.9
Ontario	111.2	24.8	73.7
Prairie Provinces	44.5	21.1	27.7
British Columbia	161.1	50.3	18.9
Total	466.0	128.8	316.8

Table C.70 : Total number of knowledge-related collaborative arrangements by province and firm size, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	1.6	0.0	1.5
Québec	88.2	21.7	14.6
Ontario	110.6	73.7	32.1
Prairie Provinces	43.0	11.8	22.3
British Columbia	62.4	29.6	5.2
Total	305.8	136.9	75.7

Table C.71 : Total number of knowledge-related collaborative arrangements by province and firm size, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	34.3	0.0	0.0
Québec	275.6	63.3	60.2
Ontario	186.7	169.2	16.0
Prairie Provinces	64.6	43.7	29.7
British Columbia	160.3	22.7	0.0
Total	721.5	299.0	105.8

Table C.72 : Total number of manufacturing/commercialization-related collaborative arrangements by province and firm size, 2001.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	0.0	0.0	0.0
Québec	57.3	15.1	0.0
Ontario	57.2	1.6	44.4
Prairie Provinces	13.1	4.3	1.8
British Columbia	15.0	5.9	3.0
Total	142.6	27.0	49.2

Table C.73 : Total number of manufacturing/commercialization-related collaborative arrangements by province and firm size, 2003.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	9.3	0.0	0.0
Québec	61.2	20.9	9.8
Ontario	29.2	75.7	12.3
Prairie Provinces	32.9	5.6	7.7
British Columbia	31.0	1.6	3.4
Total	163.7	103.7	33.2

Table C.74 : Total number of manufacturing/commercialization-related collaborative arrangements by province and firm size, 2005.

Province	Size		
	Small Firms	Medium Firms	Large Firms
Atlantic province	26.3	0.0	0.0
Québec	42.2	11.2	4.4
Ontario	27.8	10.8	0.5
Prairie Provinces	31.9	2.0	11.1
British Columbia	43.5	5.3	23.3
Total	171.7	29.3	39.3

ANNEX D

RESULTS – LOGIT MODELS

Table D.1 : First logit model, small firms, dependent variable ec, 1999 to 2005.

Variable	1999		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0387**	0.0166	0.0130	0.0169	0.0626***	0.0156	0.0314**	0.0140
ebper	0.0028	0.0052	0.0032	0.0072	0.0140**	0.0058	0.0079	0.0050
dspin	0.3250	0.4236	0.1572	0.3492	0.7867**	0.3071	1.0490***	0.2896
dprod	0.6698	0.5114	0.1357	0.5385	0.8602*	0.4459	0.4707	0.4888
dproc	0.4414	0.4525	0.2922	0.3603	0.3844	0.3623	0.3464	0.3114
dsubs	-	-	-	-	-	-	0.4507	0.4367
nbe	0.0478	0.0415	0.0866**	0.0350	-0.0136	0.0175	0.0014	0.0069
nba	0.0446	0.0257	0.0121	0.0156	0.0201*	0.0120	0.0042	0.0076
rexpperb_0	0.0040*	0.0048	0.0025	0.0040	-0.0002	0.0039	-0.0006	0.0037
sourcekgov	-0.0061	0.0207	0.0147	0.0096	0.0119	0.0069	0.0134	0.0150
lsourcekgovtot	0.0689	0.0934	0.0466	0.0406	-0.0578	0.0448	-0.0276	0.0791
nbppopprov	-34.0682	15.8701	-15.7316	10.2463	-2.0933	7.2221	-17.4940***	6.6460
Intercept	-0.8504**	0.9100	-0.1139	0.7445	-3.3809***	0.8126	-1.4371	0.8883
Observations	163		180		227		266	
Weighted observations	259		267		352		397	
F-statstic	2.12**		1.56		3.12***		2.44***	
R^2_{MZ}	99.8%		98.4%		99.9%		99.6%	

Table D.2 : First logit model, small and medium firms, dependent variable ec, 1999 to 2005.

Variable	1999		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0143	0.0096	-0.0019	0.0049	0.0117**	0.0051	0.0162***	0.0055
ebper	-0.0002	0.0050	0.0040	0.0066	0.0109**	0.0046	0.0093**	0.0042
dspin	0.1788	0.4170	0.4366	0.3059	0.7333***	0.2759	0.8680***	0.2681
dprod	0.7089	0.4713	0.5076	0.4332	0.7170*	0.4099	0.3697	0.4386
dproc	0.5631	0.4246	0.3805	0.3042	0.4913	0.3297	0.2202	0.2981
dsubs	-	-	-	-	-	-	0.1971	0.3895
nbe	0.0445	0.0359	-0.0128	0.0102	-0.0134*	0.0072	0.0022	0.0056
nba	0.0553**	0.0247	0.0145	0.0102	0.0100	0.0069	0.0001	0.0047
rexpperb_0	0.0020	0.0043	0.0010	0.0035	0.0015	0.0037	-0.0018	0.0033
sourcekgov	0.0200	0.0220	0.0154	0.0098	0.0097	0.0066	0.0167	0.0119
lsourcekgovtot	-0.0308	0.0722	0.0295	0.0402	-0.0403	0.0377	-0.0470	0.0560
nbppopprov	-40.4152**	15.8233	-11.9354	9.4657	4.8777	6.4913	-16.2013***	6.0564
Intercept	0.0056	0.8325	-0.2490	0.6775	-2.7946***	0.6959	-0.9103	0.7827
Observations	194		222		271		318	
Weighted observations	319		329		428		480	
F-statstic	2.36***		1.66**		2.77***		2.64***	
R^2_{MZ}	98.6%		49.0%		97.4%		98.7%	

Table D.3 : Second logit model, small firms, dependent variable ec, 1999 to 2005.

Variable	1999		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
lrt_0	0.0751	0.0702	0.0682	0.0659	-0.0148	0.0537	0.0367	0.0327
rbper0	0.0014	0.0053	-0.0018	0.0045	0.0043	0.0040	-0.0021	0.0039
dspin	0.4112	0.4395	0.2722	0.3713	0.7174**	0.2975	1.2010***	0.2980
dprod	0.6249	0.5174	0.1973	0.5299	0.8214*	0.4318	0.6058	0.4677
dproc	0.5412	0.4642	0.2752	0.3573	0.3985	0.3497	0.4044	0.3159
dsubs	-	-	-	-	-	-	0.2768	0.4417
nbe	0.0547	0.0434	0.0853**	0.0344	0.0014	0.0177	0.0041	0.0063
nba	0.0594	0.0260	0.0156	0.0162	0.0293**	0.0130	0.0090	0.0088
rexpperb_0	0.0016	0.0057	0.0022	0.0043	-0.0002	0.0043	0.0004	0.0040
sourcekgov	-0.0035	0.0229	0.0136	0.0101	0.0114*	0.0068	0.0099	0.0138
lsourcekgovtot	0.0570	0.0990	0.0500	0.0418	-0.0474	0.0461	-0.0058	0.0718
nbppopprov	-43.3249	16.0282	-16.6734	10.3275	-0.3255	6.8921	-16.0364**	6.4263
Intercept	-0.2698	0.7643	-0.1242	0.7506	-1.7626**	0.7703	-0.8817	0.8853
Observations	163		180		227		266	
Weighted observations	259		267		352		397	
F-statstic	1.98**		1.61*		1.88**		2.23**	
R^2_{MZ}	92.9%		95.5%		60.7%		23.7%	

Table D.4 : Second logit model, small and medium firms, dependent variable ec, 1999 to 2005.

Variable	1999		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. err.	Coeff.	Std. err.
lrt_0	0.1887***	0.0632	0.0318	0.0506	-0.0027	0.0460	0.0252	0.0284
rbper0	-0.0032	0.0049	0.0011	0.0038	0.0071**	0.0035	-0.0017	0.0036
dspin	0.4440	0.4562	0.5409*	0.3145	0.7319***	0.2757	0.9411***	0.2679
dprod	0.8517	0.5253	0.6189	0.4185	0.8290**	0.4125	0.4246	0.4107
dproc	0.8037*	0.4767	0.3409	0.3013	0.5658*	0.3190	0.3113	0.2979
dsubs	-	-	-	-	-	-	0.1229	0.3866
nbe	0.0501	0.0349	-0.0141	0.0102	-0.0136*	0.0074	0.0082	0.0060
nba	0.0641**	0.0257	0.0151	0.0101	0.0143**	0.0065	0.0027	0.0044
rexpperb_0	-0.0012	0.0053	-0.0003	0.0038	-0.0023*	0.0039	0.0004	0.0035
sourcekgov	0.0171	0.0212	0.0155	0.0100	0.0117	0.0067	0.0113	0.0112
lsourcekgovtot	-0.0188	0.0743	0.0317	0.0404	-0.0360	0.0385	-0.0192	0.0534
nbppopprov	-47.4934***	16.8855	-13.1691	9.5063	4.1375	6.4865	-15.1464**	5.9025
Intercept	-0.6238	0.7155	-0.4489	0.6729	-2.1416***	0.7127	-0.2867	0.7675
Observations	194		222		271		318	
Weighted observations	319		329		428		480	
F-statstic	2.58***		1.63*		2.31***		1.94**	
R^2_{MZ}	92.7%		17.2%		24.3%		15.3%	

Table D.5 : Third logit model, small firms, dependent variable ec, 1999 to 2005.

Variable	1999		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. err.	Coeff.	Std. err.
e	0.0293*	0.0174	0.0130	0.0175	0.0597***	0.0158	0.0377**	0.0147
ebper	0.0011	0.0073	0.0030	0.0073	0.0149***	0.0056	0.0094*	0.0050
dspin	0.5344	0.4457	0.2076	0.3508	0.7745**	0.3056	1.0172***	0.2975
nprodrd	0.1073	0.0672	0.0118	0.0081	-0.0067	0.0054	-0.0020	0.0014
nprodpc	0.3601**	0.1552	0.0394	0.1087	0.0192	0.0358	-0.0116	0.0342
nprodrc	0.1561	0.2106	-0.0163***	0.0050	0.1562	0.1698	-0.1283*	0.0762
npprodpm	-0.0006	0.0010	0.0769	0.0455	-0.0037	0.0075	-0.0106	0.0072
dsubs	-	-	-	-	-	-	0.4482	0.4515
nbe	0.0398	0.0410	0.0975**	0.0418	-0.0152	0.0174	0.0019	0.0076
nba	0.0395	0.0268	0.0147	0.0178	0.0215*	0.0121	0.0069	0.0075
rexpperb_0	0.0034	0.0051	-0.0012	0.0045	0.0005	0.0041	0.0014	0.0038
sourcekgov	0.0000	0.0216	0.0169*	0.0101	0.0129**	0.0064	0.0101	0.0148
lsourcekgovtot	0.0498	0.0949	0.0449	0.0408	-0.0644	0.0438	-0.0101	0.0795
nbppopprov	-43.7800***	16.5487	-14.9778	10.2053	-0.9365	7.4558	-17.1576***	6.5657
Intercept	-0.2526	0.8453	-0.0820	0.6301	-2.4572***	0.6687	-0.8328	0.7200
Observations	163		180		227		266	
Weighted observations	259		267		352		397	
F-statstic	2.54***		2.14**		2.67***		2.15***	
R^2_{MZ}	99.9%		99.6%		99.9%		99.8%	

Table D.6 : Third logit model, small and medium firms, dependent variable ec, 1999 to 2005.

Variable	1999		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. err.	Coeff.	Std. err.
e	0.0169*	0.0102	-0.0021	0.0051	0.0107**	0.0053	0.0174***	0.0060
ebper	-0.0015	0.0056	0.0070	0.0064	0.0123***	0.0045	0.0111***	0.0042
dspin	0.2670	0.4335	0.4498	0.3089	0.7208***	0.2751	0.8312***	0.2734
nprodrd	0.0935	0.0610	0.0032	0.0048	-0.0045	0.0063	-0.0009	0.0016
nprodpc	0.3125**	0.1422	0.0646	0.0674	0.0059	0.0373	-0.0016	0.0307
nprodrc	-0.0198	0.1650	-0.0538	0.0487	0.1584	0.1405	-0.0245	0.0506
npprodpm	-0.0003	0.0009	0.0646*	0.0342	0.0004	0.0022	-0.0115*	0.0065
dsubs	-	-	-	-	-	-	0.1030	0.3882
nbe	0.0254**	0.0340	-0.0119	0.0108	-0.0140*	0.0080	0.0030	0.0057
nba	0.0546	0.0274	0.0143	0.0109	0.0104	0.0074	-0.0001	0.0047
rexpperb_0	0.0010	0.0045	-0.0031	0.0039	0.0018	0.0036	0.0007	0.0033
sourcekgov	0.0253	0.0216	0.0172*	0.0102	0.0103*	0.0061	0.0161	0.0119
lsourcekgovtot	-0.0390	0.0674	0.0305	0.0402	-0.0430	0.0367	-0.0471	0.0555
nbppopprov	-46.6966**	16.1599	-10.6001	9.0907	5.1890	6.6716	-16.3134***	6.0778
Intercept	0.5936	0.7602	0.1197	0.5408	-1.9381***	0.5659	-0.4209	0.6029
Observations	194		222		271		318	
Weighted observations	319		329		428		480	
F-statstic	2.75***		1.64*		2.26***		2.34***	
R^2_{MZ}	99.7%		99.2%		96.9%		98.9%	

Table D.7 : Fourth logit model, small firms, dependent variable ec, 2001 to 2005.

Variable	1999		2001[28]		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. err.	Coeff.	Std. err.
e			0.0217	0.0162	0.0691***	0.0145	0.0314**	0.0136
ebper			0.0056	0.0072	0.0148**	0.0058	0.0070	0.0050
dspin			0.1691	0.3317	0.8426***	0.3069	1.0797***	0.2941
dprod			0.0854	0.5566	0.9151**	0.4478	0.4483	0.5001
dproc			0.3856	0.3556	0.4525	0.3656	0.2978	0.3126
ndroitaec			1.3473	0.9185	0.2357	0.5371	0.0175	0.0654
ndroitoec			-	-	0.1619	0.1932	0.9071*	0.4934
dsubs			-	-	-	-	0.3445	0.4216
rexpperb_0			0.0020	0.0039	-0.0017	0.0039	0.0000	0.0038
sourcekgov			0.0153	0.0095	0.0106	0.0068	0.0160	0.0141
lsourcekgovtot			0.0326	0.0416	-0.0535	0.0415	-0.0366	0.0714
nbppopprov			-9.2329	10.2110	-3.8075	7.2901	-17.4511***	6.7215
Intercept			-0.4257	0.7688	-3.4480***	0.7976	-1.2828	0.8973
Observations			168		227		266	
Weighted observations			249		352		397	
F-statstic			1.16		3.48***		2.59***	
R^2_{MZ}			99.3%		99.9%		99.6%	

[28] ndroitoec $\neq$ 0 predicts success perfectly (12 observations not used in the 2001 survey).

Table D.8 : Fourth logit model, small and medium firms, dependent variable ec, 2001 to 2005.

Variable	1999		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. err.	Coeff.	Std. err.
e			-0.0023	0.0051	0.0119***	0.0045	0.0168***	0.0053
ebper			0.0064	0.0064	0.0112**	0.0046	0.0087**	0.0042
dspin			0.4271	0.3008	0.7245***	0.2737	0.8842***	0.2687
dprod			0.3607	0.4424	0.6803*	0.4117	0.3235	0.4448
dproc			0.4219	0.3134	0.4627	0.3223	0.1773	0.3015
ndroitaec			0.1366	0.3979	0.3625	0.3222	0.0181	0.0669
ndroitoec			1.2758	0.8805	0.2694	0.2024	0.8571**	0.4100
dsubs			-	-	-	-	0.1708	0.3776
rexpperb_0			0.0003	0.0035	0.0013	0.0037	-0.0010	0.0034
sourcekgov			0.0136	0.0100	0.0097	0.0064	0.0178	0.0114
lsourcekgovtot			0.0273	0.0406	-0.0390	0.0361	-0.0501	0.0524
nbppopprov			-8.1716	9.6813	3.4694	6.5475	-16.5401***	6.1529
Intercept			-0.3659	0.6616	-2.7024***	0.6823	-0.8365	0.7830
Observations			222		271		318	
Weighted observations			329		428		480	
F-statstic			1.66*		2.72***		2.82***	
R^2_{MZ}			94.8%		97.5%		98.8%	

Table D.9 : Fifth logit model, small firms, dependent variable ec, 1999 to 2005.

Variable	1999		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. err.	Coeff.	Std. err.
e			0.0194	0.0159	0.0707***	0.0149	0.0284**	0.0138
ebper			0.0014	0.0074	0.0152**	0.0059	0.0066	0.0050
dspin			0.2778	0.3261	0.8926**	0.3112	1.0387***	0.2870
dprod			0.0217	0.5289	0.9217	0.4533	0.4042	0.4872
dproc			0.2202	0.3527	0.4670	0.3714	0.3377	0.3096
ncont			0.1513*	0.0830	-0.0014	0.0181	0.0356	0.0246
nacont			0.0212	0.0162	-0.0317**	0.0171	0.0036	0.0095
dsubs			-	-	-	-	0.3135	0.4243
rexpperb_0			0.0014	0.0038	-0.0016	0.0038	-0.0009	0.0037
sourcekgov			0.0162	0.0099	0.0098	0.0068	0.0135	0.0143
lsourcekgovtot			0.0442	0.0406	-0.0521	0.0417	-0.0205	0.0745
nbppopprov			-12.7574	9.9729	-3.2416	7.2345	-16.9777***	6.5698
Intercept			-0.2310	0.7407	-3.4583***	0.8023	-1.2209	0.8817
Observations			180		227		266	
Weighted observations			267		352		397	
F-statstic			1.41		3.60***		2.38***	
R^2_{MZ}			98.9%		99.9%		99.6%	

Table D.10 : Fifth logit model, small and medium firms, dependent variable ec, 1999 to 2005.

Variable	1999		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. err.	Coeff.	Std. err.
e			-0.0014	0.0048	0.0130***	0.0047	0.0145***	0.0054
ebper			0.0055	0.0063	0.0118**	0.0046	0.0080*	0.0042
dspin			0.4604	0.3023	0.7885***	0.2819	0.8226***	0.2663
dprod			0.3722	0.4332	0.7213*	0.4101	0.3154	0.4358
dproc			0.3529	0.3048	0.5027	0.3220	0.2134	0.2981
ncont			0.0160	0.0191	-0.0088	0.0123	0.0292*	0.0176
nacont			0.0233	0.0162	-0.0280**	0.0109	0.0001	0.0078
dsubs			-	-	-	-	0.1177	0.3775
rexpperb_0			-0.0011	0.0036	0.0020	0.0036	-0.0015	0.0033
sourcekgov			0.0162*	0.0097	0.0085	0.0065	0.0164	0.0115
lsourcekgovtot			0.0350	0.0398	-0.0313	0.0364	-0.0413	0.0536
nbppopprov			-11.2194	9.2771	4.8970	6.5416	-15.8467***	6.0349
Intercept			-0.2378	0.6566	-2.8034***	0.6879	-0.7577	0.7769
Observations			222		271		318	
Weighted observations			329		428		480	
F-statstic			1.66*		3.13***		2.47***	
R^2_{MZ}			64.6%		97.9%		98.3%	

Table D.11 : Sixth logit model, small firms, dependent variable ec, 1999 to 2005.

Variable	1999		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. err.	Coeff.	Std. err.
e	0.0238	0.0175	0.0161	0.0179	0.0568***	0.0146	0.0247*	0.0139
ebper	-0.0039	0.0066	0.0004	0.0074	0.0114*	0.0060	0.0056	0.0052
dspin	0.2906	0.4042	0.1813	0.3311	0.7126**	0.3130	0.9014***	0.2879
lrdt_0	0.3415**	0.1686	0.1419	0.1153	0.1426	0.0950	0.1092	0.0737
rdbper0	0.0071	0.0069	0.0035	0.0066	0.0092*	0.0053	0.0037	0.0055
dsubs	-	-	-	-	-	-	0.4290	0.4281
rexpperb_0	0.0004	0.0215	0.0016	0.0040	-0.0003	0.0039	0.0003	0.0036
sourcekgov	0.0481	0.1005	0.0134	0.0092	0.0109	0.0069	0.0115	0.0135
lsourcekgovtot	0.0073	0.0046	0.0405	0.0398	-0.0517	0.0428	-0.0179	0.0694
nbppopprov	-41.7350***	15.8664	-10.7058	10.0943	-3.6729	7.5724	-16.5172**	6.6369
Intercept	-1.3635	1.1214	-0.8498	1.0065	-3.4684***	0.8503	-2.1462**	1.0205
Observations	163		180		227		266	
Weighted observations	259		267		352		397	
F-statstic	2.23**		1.30		4.34***		2.62***	
R^2_{MZ}	99.4%		98.3%		99.9%		99.4%	

Table D.12 : Sixth logit model, small and medium firms, dependent variable ec, 1999 to 2005.

Variable	1999		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. err.	Coeff.	Std. err.
e	0.0126	0.0104	-0.0053	0.0053	0.0089*	0.0052	0.0144**	0.0058
ebper	-0.0039	0.0057	0.0025	0.0064	0.0072	0.0048	0.0084*	0.0043
dspin	0.1900	0.3842	0.3728	0.3097	0.6268**	0.2765	0.7965***	0.2657
lrdt_0	0.2650**	0.1191	0.1962**	0.0922	0.1201	0.0733	0.0524	0.0614
rdbper0	0.0043	0.0062	0.0037	0.0057	0.0082*	0.0046	0.0012	0.0048
dsubs	-	-	-	-	-	-	0.1884	0.3802
rexpperb_0	0.0231	0.0206	0.0021	0.0037	0.0025	0.0035	-0.0008	0.0032
sourcekgov	-0.0377	0.0687	0.0161*	0.0091	0.0102	0.0062	0.0161	0.0114
lsourcekgovtot	0.0043	0.0041	0.0318	0.0395	-0.0391	0.0361	-0.0464	0.0526
nbppopprov	-42.8928***	15.9517	-12.0308	9.4326	3.9238	6.7331	-15.7458***	5.9918
Intercept	-0.2530	0.9155	-0.9743	0.8796	-2.7883***	0.6817	-1.0810	0.8740
Observations	194		222		271		318	
Weighted observations	319		329		428		480	
F-statstic	1.99**		1.99**		3.21***		2.98***	
R^2_{MZ}	97.8%		86.1%		95.6%		98.3%	

Table D.13 : Seventh logit model, small firms, dependent variable ec, 1999 to 2005.

Variable	1999[29]		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. err.	Coeff.	Std. err.
e	0.0770**	0.0308	0.0239	0.0160	0.0664***	0.0152	0.0355**	0.0140
ebper	0.0222	0.0140	0.0023	0.0070	0.0150**	0.0054	0.0095*	0.0050
dspin	0.7421	0.6978	0.1937	0.3195	0.8072**	0.3111	1.0289***	0.2897
reusk	-	-	1.0411	2.9796	0.3706	2.2869	4.6285*	2.5188
lfkreun	0.1661	0.1358	-0.0342	0.2057	0.0357	0.1607	-0.3074*	0.1723
dsubs	-	-	-	-	-	-	0.2979	0.4335
rexpperb_0	0.0115	0.0152	0.0019	0.0038	-0.0005	0.0040	0.0006	0.0036
sourcekgov	0.0140	0.0094	0.0167	0.0103	0.0047	0.0062	0.0026	0.0068
nbppopprov	22.0893	28.5380	-11.5937	10.0946	-3.1157	7.2240	-17.1758***	6.5633
Intercept	-5.7459*	2.9786	0.0241	0.6120	-2.6202***	0.6635	-0.8377	0.7020
Observations	79		180		227		266	
Weighted observations	118		267		352		397	
F-statstic	1.18		1.31		4.78***		3.36***	
R^2_{MZ}	86.1%		99.2%		99.9%		99.7%	

[29] Variable reusk dropped due to collinearity in the 1999 survey.

Table D.14 : Seventh logit model, small and medium firms, dependent variable ec, 1999 to 2005.

Variable	1999[30]		2001		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. err.	Coeff.	Std. err.
e	0.0124	0.0163	-0.0014	0.0047	0.0109**	0.0047	0.0172***	0.0055
ebper	0.0202*	0.0119	0.0046	0.0064	0.0113**	0.0044	0.0101**	0.0042
dspin	0.5091	0.5685	0.4023	0.2999	0.6670**	0.2739	0.8421***	0.2662
reusk	-	-	-0.3229	2.5969	0.0954	1.9102	3.1231	2.1188
lfkreun	0.1163	0.1111	0.0593	0.1753	-	-	-0.1991	0.1418
dsubs	-	-	-	-	0.0513	0.1315	0.1159	0.3733
rexpperb_0	0.0105	0.0163	0.0015	0.0034	0.0023	0.0036	-0.0005	0.0032
sourcekgov	0.0138*	0.0073	0.0185	0.0099	0.0040	0.0055	0.0023	0.0064
nbppopprov	17.6406	25.4925	-11.9757	9.1799	3.3268	6.5538	-16.5888***	6.0136
Intercept	-3.6348	2.4510	0.2820	0.5372	-1.9769***	0.5481	-0.4536	0.5921
Observations	96		222		271		318	
Weighted observations	134		329		428		480	
F-statstic	0.99		1.82*		3.58***		3.49***	
R^2_{MZ}	38.8%		36.6%		97.0%		98.8%	

[30] Variable reusk dropped due to collinearity in the 1999 survey.

Table D.15 : First logit model, dependent variable `ecepri`, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0227	0.0156	0.0237*	0.0130	0.0066	0.0057	0.0119**	0.0051
ebper	0.0036	0.0059	0.0079	0.0053	0.0048	0.0056	0.0117**	0.0046
dspin	0.2443	0.3458	0.6086**	0.2957	0.4798	0.3141	0.6249**	0.2758
dprod	-0.0057	0.5162	0.2207	0.4756	0.2182	0.5127	0.1224	0.4360
dproc	0.4274	0.4466	0.0037	0.3077	0.5205	0.4297	-0.0366	0.3003
dsubs	-	-	0.0177	0.4267	-	-	0.0720	0.4006
nbe	-0.0234	0.0182	-0.0058	0.0077	-0.0222***	0.0068	0.0011	0.0065
nba	0.0229**	0.0094	0.0048	0.0080	0.0222***	0.0070	-0.0019	0.0043
rexpperb_0	-0.0033	0.0049	-0.0011	0.0038	-0.0014	0.0042	-0.0039	0.0035
sourcekgov	0.0032	0.0080	0.0074	0.0128	0.0053	0.0074	0.0064	0.0103
lsourcekgovtot	-0.0269	0.0490	-0.0351	0.0657	0.0001	0.0394	-0.0252	0.0474
nbppopprov	6.1098	8.8197	-14.8642	6.6240	6.1814	8.1955	-14.6299**	6.0350
Intercept	-2.6827	0.8269	-1.0167	0.8403	-3.0469***	0.8495	-1.0686	0.7832
Observations	227		266		271		318	
Weighted observations	352		397		428		479	
F-statstic	1.44		1.26		2.01**		1.81**	
R^2_{MZ}	99.3%		99.4%		92.4%		97.6%	

Table D.16 : Third logit model, dependent variable `ecepri`, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0203	0.0155	0.0281**	0.0133	0.0047	0.0064	0.0122**	0.0052
ebper	0.0045	0.0059	0.0086	0.0053	0.0056	0.0056	0.0123***	0.0047
dspin	0.2120	0.3555	0.6088**	0.2973	0.4488	0.3207	0.6122**	0.2743
nprodrd	-0.0008	0.0083	-0.0018	0.0014	-0.0012	0.0089	-0.0008	0.0012
nprodpc	-0.0025	0.0480	0.0376	0.0320	0.0007	0.0504	0.0344	0.0297
nprodrc	0.0990	0.1935	-0.0597	0.0723	0.0895	0.1640	0.0178	0.0473
npprodpm	0.0050	0.0076	-0.0162	0.0084	0.0038	0.0026	-0.0126**	0.0062
dsubs	-	-	0.0016	0.4274	-	-	0.0200	0.4091
nbe	-0.0247	0.0186	-0.0066	0.0077	-0.0226***	0.0071	0.0001	0.0063
nba	0.0240***	0.0093	0.0053	0.0079	0.0227***	0.0072	-0.0023	0.0043
rexpperb_0	-0.0040	0.0049	-0.0002	0.0038	-0.0021	0.0041	-0.0028	0.0034
sourcekgov	0.0043	0.0080	0.0081	0.0132	0.0065	0.0075	0.0067	0.0104
lsourcekgovtot	-0.0319	0.0494	-0.0422	0.0670	-0.0080	0.0403	-0.0305	0.0465
nbppopprov	6.5807	8.9114	-15.5706**	6.7694	6.6574	8.3541	-15.4185**	6.1867
Intercept	-2.4469***	0.7149	-0.8223	0.7084	-2.5234***	0.6596	-0.9567	0.6308
Observations	227		226		271		318	
Weighted observations	352		397		428		480	
F-statstic	1.29		1.37		1.91		1.80**	
R^2_{MZ}	99.1%		99.6%		87.1%		97.7%	

Table D.17 : Fourth logit model, dependent variable `ecepri`, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0340**	0.0136	0.0232*	0.0126	0.0110**	0.0044	0.0116**	0.0046
ebper	0.0048	0.0060	0.0079	0.0053	0.0065	0.0055	0.0115**	0.0047
dspin	0.3006	0.3403	0.6067**	0.2927	0.5317*	0.3097	0.6190**	0.2717
dprod	0.1315	0.5154	0.2056	0.4762	0.3666	0.5126	0.0671	0.4334
dproc	0.5161	0.4551	-0.0222	0.3078	0.4610	0.3962	-0.0641	0.3048
ndroitaec	-0.6380	0.7554	0.0505	0.0744	-0.1842	0.3214	0.0577	0.0772
ndroitoec	0.0240	0.2101	0.2134	0.3390	0.0391	0.2048	0.2381	0.3265
dsubs	-	-	-0.0692	0.4191	-	-	0.0696	0.3939
rexpperb_0	-0.0058	0.0050	-0.0010	0.0039	-0.0028	0.0042	-0.0034	0.0035
sourcekgov	0.0003	0.0080	0.0103	0.0127	0.0030	0.0073	0.0064	0.0102
lsourcekgovtot	-0.0067	0.0462	-0.0476	0.0646	0.0117	0.0376	-0.0243	0.0464
nbppopprov	3.5942	8.5158	-14.7389**	6.5993	6.2446	7.8462	-15.1533**	6.0479
Intercept	-2.7926***	0.8027	-0.9313	0.8420	-3.1657***	0.8229	-1.0006	0.7820
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	1.20		1.29		1.34		1.87**	
R^2_{MZ}	99.7%		99.4%		97.1%		97.4%	

Table D.18 : Fifth logit model, dependent variable `ecepri`, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0328**	0.0138	0.0200	0.0132	0.0106**	0.0045	0.0086*	0.0050
ebper	0.0047	0.0062	0.0073	0.0054	0.0066	0.0056	0.0106**	0.0048
dspin	0.2957	0.3438	0.5703*	0.2926	0.5297*	0.3124	0.5327*	0.2751
dprod	0.0650	0.5245	0.1805	0.4751	0.3435	0.5111	0.0389	0.4324
dproc	0.5523	0.4594	0.0012	0.3050	0.4697	0.3962	-0.0559	0.3018
dsubs	-	-	0.0199	0.0170	-	-	0.0242*	0.0126
ncont	0.0058	0.0173	-0.0026	0.0086	0.0033	0.0123	-0.0027	0.0080
nacont	-0.0263	0.0196	-0.1051	0.4194	-0.0110	0.0095	-0.0052	0.3918
rexpperb_0	-0.0052	0.0049	-0.0008	0.0039	-0.0024	0.0041	-0.0031	0.0036
sourcekgov	0.0005	0.0079	0.0094	0.0127	0.0029	0.0073	0.0054	0.0102
lsourcekgovtot	-0.0144	0.0454	-0.0409	0.0649	0.0112	0.0378	-0.0166	0.0460
nbppopprov	3.4216	8.5153	-14.3377**	6.6112	5.9264	7.7842	-14.6223**	6.0719
Intercept	-2.7169***	0.7974	-0.8487	0.8450	-3.1235***	0.8167	-0.8431	0.7837
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	1.22		1.24		1.42		1.93**	
R^2_{MZ}	99.7%		99.1%		96.9%		95.4%	

Table D.19 : Sixth logit model, dependent variable `ecepri`, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0104	0.0156	0.0187	0.0134	0.0030	0.0061	0.0096*	0.0051
ebper	-0.0022	0.0071	0.0067	0.0053	-0.0001	0.0063	0.0114**	0.0046
dspin	0.1199	0.3466	0.5464**	0.2898	0.3947	0.3151	0.5997**	0.2683
lrdt_0	0.3274***	0.1220	0.0668	0.0624	0.2688***	0.0944	0.0510	0.0620
rdbper0	0.0094	0.0071	0.0004	0.0055	0.0077	0.0062	-0.0030	0.0050
dsubs	-	-	-0.0473	0.4202	-	-	0.0922	0.3980
rexpperb_0	-0.0029	0.0050	-0.0008	0.0037	-0.0004	0.0040	-0.0033	0.0034
sourcekgov	0.0030	0.0077	0.0091	0.0128	0.0057	0.0070	0.0062	0.0102
lsourcekgovtot	-0.0206	0.0424	-0.0451	0.0647	0.0012	0.0349	-0.0263	0.0463
nbppopprov	1.6661	8.9014	-14.3852**	6.6733	4.0273	8.2106	-14.6661	6.1204
Intercept	-4.1723***	1.0802	-1.4741	0.9250	-4.0980***	0.9124	-1.3544**	0.9484
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	1.79*		1.43		2.35**		2.13**	
R^2_{MZ}	96.8%		99.0%		73.1%		96.3%	

Table D.20 : Seventh logit model, dependent variable `ecepri`, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0284*	0.0151	0.0233*	0.0129	0.0099*	0.0051	0.0107**	0.0048
ebper	0.0044	0.0061	0.0083	0.0052	0.0059	0.0055	0.0111**	0.0046
dspin	0.2237	0.3469	0.5960**	0.2871	0.4594	0.3118	0.5987**	0.2654
reusk	0.3509	2.0312	2.2192	2.4229	0.4709	1.7331	0.9827	1.9868
lfkreun	0.0509	0.1417	-0.1435	0.1666	0.0446	0.1176	-0.0515	0.1330
dsubs	-	-	-0.1201	0.4232	-	-	0.0468	0.3926
rexpperb_0	-0.0046	0.0047	-0.0007	0.0037	-0.0018	0.0039	-0.0033	0.0034
sourcekgov	-0.0005	0.0074	-0.0027	0.0070	0.0033	0.0059	-0.0017	0.0068
nbppopprov	2.6788	8.6181	-14.9212**	6.5862	5.0220	8.1712	-15.3617**	6.0985
Intercept	-2.5829***	0.7136	-0.7435	0.7026	-2.8155***	0.6770	-0.9347	0.6203
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	2.48**		1.72*		3.16***		2.45**	
R^2_{MZ}	99.6%		99.4%		96.4%		97.0%	

Table D.21 : Fourth logit model, dependent variable `ecipub`, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003[31]		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	-0.0047	0.0148	0.0265**	0.0134	0.0069	0.0059	0.0124**	0.0054
ebper	0.0180**	0.0076	0.0042	0.0059	0.0267***	0.0077	0.0053	0.0045
dspin	0.2458	0.3649	1.1555***	0.3041	0.0940	0.3343	0.7479***	0.2713
dprod	0.9027	0.6070	0.3949	0.5566	1.0039*	0.5845	0.5057	0.4756
dproc	1.1161*	0.5744	0.1948	0.3360	1.0871**	0.5291	0.0618	0.3100
ndroitaec	-	-	0.0135	0.0750	-0.0341	0.3399	-0.0156	0.0802
ndroitoec	0.2721	0.2500	0.8261*	0.4712	0.2766	0.2363	0.6620*	0.3722
dsubs	-	-	0.2634	0.4512	-	-	0.0838	0.3670
rexpperb_0	-0.0062	0.0057	0.0011	0.0039	-0.0029	0.0049	0.0013	0.0034
sourcekgov	0.0105	0.0079	0.0073	0.0115	0.0125	0.0077	0.0056	0.0105
lsourcekgovtot	-	-	0.0084	0.0538	-0.0805*	0.0454	0.0071	0.0469
nbppopprov	8.0277	10.2731	-11.7135	7.3286	8.3531	9.2337	-9.2257	6.4044
Intercept	-5.1311***	1.1761	-2.0649	1.0078	-6.2093***	1.2311	-1.7269**	0.8476
Observations	207		266		271		318	
Weighted observations	321		397		428		480	
F-statstic	1.95**		2.45***		2.45***		1.61*	
R^2_{MZ}	88.6%		99.5%		92.9%		97.8%	

[31]Variable `ndroitaec` $\neq$ `0` predicts failure perfectly. Variable `ndroitaec` dropped and 6 observations not used; Variable `lsourcekgovtot` $\neq$ `0` predicts failure perfectly. Variable `lsourcekgovtot` dropped and 14 observations not used.

Table D.22 : Fourth logit model, dependent variable ecc, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0520***	0.0144	0.0354***	0.0137	0.0085	0.0042	0.0195***	0.0055
ebper	0.0058	0.0059	0.0073	0.0050	0.0086	0.0049	0.0089**	0.0042
dspin	0.3571	0.3323	1.0206***	0.2924	0.3221	0.2871	0.8439***	0.2669
dprod	1.1512**	0.6095	0.3621	0.4876	0.7640	0.5144	0.2759	0.4433
dproc	0.4568	0.4022	0.2931	0.3145	0.6501	0.3701	0.1918	0.3026
ndroitaec	-1.0448	1.2991	0.0211	0.0670	0.1726	0.2563	0.0219	0.0686
ndroitoec	0.0270	0.2130	0.9771*	0.5203	0.1398	0.1963	0.9215**	0.4242
dsubs	-	-	0.7277*	0.4390	-	-	0.4353	0.3775
rexpperb_0	-0.0064	0.0044	-0.0029	0.0038	-0.0025	0.0038	-0.0033	0.0034
sourcekgov	-0.0054	0.0071	0.0162	0.0141	-0.0033	0.0066	0.0173	0.0112
lsourcekgovtot	-0.0226	0.0432	-0.0374	0.0719	0.0030	0.0348	-0.0483	0.0517
nbppopprov	16.2609*	8.5649	-14.0542**	6.9347	19.8739	7.5921	-13.7726**	6.2635
Intercept	-4.2770***	1.0097	-1.9192**	0.9312	-4.0696	0.9018	-1.3582*	0.8078
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	2.60***		2.62***		1.94**		2.88***	
R^2_{MZ}	99.9%		99.7%		95.3%		99.1%	

Table D.23 : Seventh logit model, dependent variable ecc, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0494***	0.0148	0.0389***	0.0141	0.0069	0.0045	0.0198***	0.0057
ebper	0.0067	0.0055	0.0094*	0.0050	0.0092*	0.0048	0.0100**	0.0043
dspin	0.3086	0.3232	0.9693***	0.2850	0.2871	0.2864	0.7937***	0.2633
reusk	0.3900	2.2077	5.0598**	2.5531	-0.8738	1.7666	3.6413*	2.1371
lfkreun	0.0154	0.1536	-0.3273*	0.1749	-	-	-0.2262	0.1429
dsubs	-	-	0.6705	0.4486	0.1062	0.1207	0.3630	0.3719
rexpperb_0	-0.0050	0.0044	-0.0022	0.0036	-0.0011	0.0037	-0.0026	0.0032
sourcekgov	-0.0092	0.0073	0.0011	0.0067	-0.0037	0.0057	0.0007	0.0063
nbppopprov	16.5458*	8.6570	-14.2088**	6.7332	19.5758**	7.6378	-14.0118**	6.1098
Intercept	-3.2000***	0.7335	-1.5193**	0.7472	-3.1032***	0.6752	-0.9833	0.6199
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	3.15***		3.36***		2.56**		3.54***	
R^2_{MZ}	99.9%		99.8%		92.8%		99.1%	

Table D.24 : Fourth logit model, dependent variable ecpc, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0205	0.0142	0.0030	0.0136	0.0050	0.0054	0.0020	0.0062
ebper	0.0083	0.0065	-0.0021	0.0064	0.0117*	0.0063	-0.0034	0.0052
dspin	0.7336**	0.3368	0.5843	0.3587	0.8696***	0.3220	0.5324	0.3350
dprod	0.0494	0.4795	0.4228	0.6369	0.1266	0.4684	0.3264	0.5655
dproc	0.2958	0.4293	0.1083	0.3732	0.2320	0.4031	0.0790	0.3654
ndroitaec	0.4919	0.4928	0.0122	0.0694	0.0404	0.3388	0.0144	0.0717
ndroitoec	0.1001	0.2208	0.6782**	0.3325	-	-	0.6937**	0.3406
dsubs	-	-	-0.2494	0.4908	0.1285	0.2124	-0.5600	0.4124
rexpperb_0	0.0024	0.0045	0.0032	0.0045	0.0027	0.0042	0.0029	0.0040
sourcekgov	0.0081	0.0074	-0.0067	0.0156	0.0094	0.0069	0.0076	0.0114
lsourcekgovtot	-0.0009	0.0393	0.0818	0.0731	0.0229	0.0352	0.0075	0.0509
nbppopprov	-3.7202	7.5185	-10.4622	7.9988	-1.3422	7.5348	-11.7539	7.3223
Intercept	-2.7613***	0.7681	-1.4352	1.1163	-3.1661***	0.7911	-0.8594	0.9824
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	1.12		1.05		2.07**		1.01	
R^2_{MZ}	99.1%		74.9%		87.0%		60.6%	

Table D.25 : Seventh logit model, dependent variable ecpc, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0181	0.0151	0.0096	0.0132	0.0048	0.0060	0.0033	0.0059
ebper	0.0078	0.0064	-0.0005	0.0063	0.0114*	0.0063	-0.0021	0.0052
dspin	0.6853**	0.3381	0.5432	0.3598	0.8393***	0.3200	0.5346	0.3354
reusk	0.8739	2.0898	2.7032	3.0483	0.9310	1.8819	3.7876	2.4853
lfkreun	-0.0154	0.1454	-0.1606	0.2091	-0.0239	0.1293	-0.2376	0.1671
dsubs	-	-	-0.2532	0.5011	-	-	-0.6504	0.4110
rexpperb_0	0.0026	0.0043	0.0036	0.0044	0.0034	0.0040	0.0036	0.0039
sourcekgov	0.0078	0.0064	0.0023	0.0077	0.0111**	0.0054	0.0011	0.0074
nbppopprov	-2.9713	7.4707	-11.7450	8.2185	-0.9799	7.6316	-12.4403	7.5838
Intercept	-2.6725***	0.6629	-1.0964	0.8776	-3.0795***	0.6808	-0.5244	0.7079
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	2.00**		0.96		3.27***		1.20	
R^2_{MZ}	98.9%		96.4%		86.0%		76.2%	

ANNEX E

RESULTS - NEGATIVE BINOMIAL MODELS

Table E.1 : First negative binomial model, dependent variable nec, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0235*	0.0128	0.0243***	0.0094	0.0066	0.0049	0.0149***	0.0043
ebper	0.0136***	0.0052	0.0090**	0.0044	0.0148***	0.0047	0.0083**	0.0036
dspin	0.0507	0.2455	0.5333**	0.2561	0.0463	0.2398	0.5052**	0.2527
dprod	-0.1288	0.3994	0.9060**	0.4001	0.0306	0.3895	0.9154***	0.3439
dproc	0.4557	0.3262	0.3709	0.2644	0.6858**	0.3144	0.2811	0.2430
dsubs	-	-	0.2365	0.3055	-	-	0.0709	0.2945
nbe	-0.0129	0.0095	-0.0048	0.0040	-0.0017	0.0091	-0.0035	0.0028
nba	0.0156***	0.0049	-0.0068	0.0047	0.0133***	0.0039	-0.0029	0.0028
rexpperb_0	-0.0025	0.0037	-0.0064*	0.0033	0.0002	0.0038	-0.0039	0.0031
sourcekgov	0.0042	0.0052	-0.0026	0.0088	0.0060	0.0051	0.0036	0.0073
lsourcekgovtot	-0.0582*	0.0319	0.0451	0.0462	-0.0081	0.0292	0.0008	0.0337
nbppopprov	3.5952	5.7494	-1.9018	4.8119	4.4566	5.9157	2.4855	4.3018
Intercept	-1.9935***	0.6909	-1.5390**	0.6889	-2.3839***	0.6617	-1.4739**	0.6018
Alpha	2.5212		2.4668		2.8309		2.5107	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	2.90***		2.36***		3.44***		2.58***	

Table E.2 : Third negative binomial model, dependent variable nec, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0234*	0.0123	0.0244***	0.0095	0.0038	0.0046	0.0118***	0.0040
ebper	0.0152***	0.0053	0.0102**	0.0046	0.0161***	0.0047	0.0099***	0.0036
dspin	0.0587	0.2500	0.4335	0.2706	0.0116	0.2475	0.3688	0.2549
nprodrd	-0.0054	0.0040	-0.0021	0.0020	-0.0037	0.0044	-0.0007*	0.0003
nprodpc	0.0191	0.0264	0.0706	0.0550	0.0038	0.0273	0.0448	0.0455
nprodrc	-0.0080	0.1251	-0.0060	0.0784	0.0726	0.1203	0.0596*	0.0319
nprodpm	0.0082	0.0089	0.0002	0.0073	0.0036***	0.0009	-0.0034	0.0058
dsubs	-	-	0.2734	0.3065	-	-	0.1099	0.2812
nbe	-0.0146	0.0099	-0.0063*	0.0032	-0.0006	0.0092	-0.0054**	0.0027
nba	0.0163***	0.0049	-0.0059	0.0045	0.0133***	0.0040	-0.0008	0.0024
rexpperb_0	-0.0042	0.0039	-0.0052*	0.0032	-0.0024	0.0036	-0.0033	0.0027
sourcekgov	0.0054	0.0053	-0.0038	0.0078	0.0092	0.0052	0.0013	0.0068
lsourcekgovtot	-0.0651**	0.0300	0.0427	0.0390	-0.0288	0.0270	0.0023	0.0328
nbppopprov	2.6437	5.9300	-3.4648	4.6536	4.9431	5.9678	-0.9526	4.2725
Intercept	-1.8069***	0.6736	-0.5288	0.4924	-1.8659***	0.6173	-0.3818	0.4737
Alpha	2.5336		2.5081				2.4826	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	2.59***		2.17***		6.47***		3.17***	

Table E.3 : Fourth negative binomial model, dependent variable nec, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0304***	0.0105	0.0201**	0.0089	0.0152***	0.0056	0.0134***	0.0040
ebper	0.0131***	0.0048	0.0069	0.0046	0.0172***	0.0046	0.0070*	0.0039
dspin	0.1037	0.2424	0.3992	0.2594	0.0637	0.2379	0.4092	0.2513
dprod	-0.0104	0.3878	0.8337**	0.3886	0.0905	0.4011	0.8403**	0.3365
dproc	0.5078	0.3259	0.3061	0.2758	0.7934	0.3170	0.2064	0.2414
ndroitaec	-1.2392**	0.5376	-0.0192	0.1189	-0.7980***	0.2564	-0.0245	0.1161
ndroitoec	0.1392	0.1618	0.5160**	0.2480	0.1524	0.1659	0.4813**	0.2372
dsubs	-	-	0.2608	0.3027	-	-	0.1257	0.2762
rexpperb_0	-0.0056	0.0035	-0.0050	0.0035	-0.0022	0.0038	-0.0027	0.0031
sourcekgov	0.0018	0.0051	0.0003	0.0097	0.0040	0.0052	0.0047	0.0076
lsourcekgovtot	-0.0447	0.0330	0.0289	0.0494	-0.0073	0.0295	-0.0030	0.0350
nbppopprov	1.5424	5.8702	-0.5440	4.8197	6.1652	6.0865	3.6103	4.2604
Intercept	-1.9038***	0.6590	-1.4238	0.7008	-2.7285***	0.6574	-1.4448**	0.6149
Alpha	2.6269		2.4346		2.9910		2.4838	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	2.95***		2.69***		3.03***		2.93***	

Table E.4 : Fifth negative binomial model, dependent variable nec, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0243**	0.0109	0.0174*	0.0096	0.0147**	0.0058	0.0094***	0.0035
ebper	0.0125**	0.0049	0.0047	0.0036	0.0173***	0.0048	0.0070**	0.0031
dspin	0.1245	0.2435	0.4791**	0.2158	0.1126	0.2433	0.3214	0.2220
dprod	-0.0841	0.3849	0.7006*	0.3728	0.0657	0.3980	0.7867**	0.3279
dproc	0.4340	0.3457	0.2371	0.2498	0.7497**	0.3335	0.1586	0.2279
ncont	0.0215*	0.0117	0.0419*	0.0235	0.0072	0.0106	0.0193*	0.0113
nacont	-0.0210	0.0138	0.0011	0.0034	-0.0150	0.0114	-0.0002	0.0036
dsubs	-	-	0.1553	0.3007	-	-	0.0109	0.2714
rexpperb_0	-0.0044	0.0036	-0.0045	0.0028	-0.0015	0.0037	-0.0040	0.0026
sourcekgov	0.0023	0.0050	0.0000	0.0086	0.0035	0.0052	0.0034	0.0072
lsourcekgo~t	-0.0488	0.0330	0.0356	0.0437	-0.0046	0.0302	0.0058	0.0328
nbppopprov	2.8690	5.6929	-1.2273	4.7039	6.9176	5.9460	1.7443	4.2984
Intercept	-1.7968***	0.6588	-1.0823	0.6788	-2.7435***	0.6601	-1.0621*	0.5752
Alpha	2.5547		2.3455		3.0291		2.3638	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	2.75***		2.27***		2.83***		2.64***	

Table E.5 : Seventh negative binomial model, dependent variable nec, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0235**	0.0120	0.0228**	0.0090	0.0171***	0.0062	0.0132***	0.0045
ebper	0.0123***	0.0047	0.0074*	0.0039	0.0183***	0.0048	0.0081**	0.0034
dspin	0.2610	0.2405	0.3901	0.2559	0.1257	0.2440	0.4552*	0.2389
reusk	-0.9757	1.4536	1.5755	1.8906	0.1729	1.4889	1.6375	1.5197
lfkreun	0.1337	0.1003	-0.0674	0.1343	0.0428	0.1009	-0.0773	0.1014
dsubs	-	-	0.2704	0.3085	-	-	0.1147	0.2714
rexpperb_0	-0.0049	0.0035	-0.0026	0.0029	-0.0034	0.0036	-0.0005	0.0027
sourcekgov	0.0041	0.0054	-0.0010	0.0060	0.0049	0.0047	-0.0030	0.0057
nbppopprov	0.3352	6.0433	-3.8498	4.5402	6.6437	6.2028	-0.0135	4.1690
Intercept	-1.8670***	0.5648	-0.4439	0.5085	-2.5682***	0.5879	-0.5021	0.4889
Alpha	2.4508		2.5070		2.9882		2.5549	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	5.01***		3.02***		5.63***		3.12***	

Table E.6 : Fourth negative binomial model, dependent variable `necepri`, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0313***	0.0119	0.0092	0.0105	0.0169***	0.0065	0.0112**	0.0056
ebper	0.0077	0.0056	0.0064	0.0050	0.0119**	0.0052	0.0066	0.0046
dspin	0.1909	0.3123	0.1151	0.2876	0.1794	0.2899	0.3492	0.3064
dprod	0.0121	0.4592	1.2874***	0.4940	0.1041	0.4830	1.0268**	0.4387
dproc	0.1709	0.3794	0.5060	0.3274	0.4617	0.3687	0.3683	0.3077
ndroitaec	-1.0391**	0.5744	-0.0771	0.0810	-0.7805***	0.2743	-0.0657	0.0851
ndroitoec	0.1467	0.2717	0.3053	0.3122	0.1471	0.2618	0.2869	0.3107
dsubs	-	-	-0.1457	0.3291	-	-	-0.1900	0.3518
rexpperb_0	-0.0049	0.0043	-0.0079**	0.0040	-0.0024	0.0044	-0.0037	0.0040
sourcekgov	-0.0053	0.0061	-0.0058	0.0120	-0.0043	0.0057	0.0003	0.0102
lsourcekgovtot	-0.0087	0.0331	0.0046	0.0564	0.0133	0.0303	-0.0297	0.0450
nbppopprov	2.1108	7.4365	-0.5162	5.5929	10.0289	7.4148	3.5468	5.1271
Intercept	-1.8125**	0.8395	-1.7411**	0.8217	-2.8102***	0.7940	-1.8273**	0.7606
Alpha	4.3346		3.7605		4.8113		4.1400	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	1.36		1.45		1.86**		1.54	

Table E.7 : Seventh negative binomial model, dependent variable `necepri`, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0282**	0.0141	0.0144	0.0117	0.0205***	0.0070	0.0118*	0.0065
ebper	0.0060	0.0055	0.0068	0.0046	0.0119**	0.0053	0.0084*	0.0043
dspin	0.2569	0.3220	0.0541	0.3087	0.1956	0.3019	0.3726	0.3103
reusk	-0.7908	1.8728	1.8799	2.4389	0.7026	1.9172	2.6566	1.9485
lfkreun	0.1351	0.1296	-0.0864	0.1730	0.0194	0.1295	-0.1529	0.1331
dsubs	-	-	-0.0896	0.3480	-0.0026	0.0044	-0.0790	0.3328
rexpperb_0	-0.0033	0.0041	-0.0058	0.0035	-0.0020	0.0050	-0.0014	0.0036
sourcekgov	-0.0013	0.0062	-0.0095	0.0064	12.9115*	7.2777	-0.0105	0.0067
nbppopprov	4.4828	7.2854	-4.5684	5.8187	-3.1279***	0.6723	-0.0392	5.3341
Intercept	-2.3291***	0.7084	-0.2596	0.5738	4.5329		-0.7807	0.5935
Alpha	3.6189		3.8832		4.5329		4.2580	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	3.37***		1.19		3.57***		1.51	

Table E.8 : Fourth negative binomial model, dependent variable necipub, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0195	0.0146	0.0270***	0.0104	0.0146***	0.0049	0.0179***	0.0049
ebper	0.0218***	0.0059	0.0071	0.0052	0.0293***	0.0057	0.0076*	0.0042
dspin	-0.0881	0.3118	0.7799***	0.2930	-0.0609	0.3096	0.5416**	0.2698
dprod	0.7225	0.5623	0.3842	0.4818	0.8214	0.5854	0.7420*	0.4420
dproc	1.6911***	0.5242	0.0550	0.2986	1.7189***	0.4946	0.1064	0.2782
ndroitaec	-13.1751***	0.9330	0.0381	0.1312	-0.9231***	0.3403	0.0181	0.1320
ndroitoec	0.1695	0.1499	0.7992***	0.2967	0.1673	0.1467	0.7093***	0.2568
dsubs	-	-	0.6813	0.3795	-	-	0.4493	0.3312
rexpperb_0	-0.0119**	0.0053	-0.0011	0.0037	-0.0046	0.0045	-0.0018	0.0034
sourcekgov	0.0109	0.0076	-0.0009	0.0109	0.0125	0.0076	0.0059	0.0089
lsourcekgovtot	-0.3081***	0.0704	0.0608	0.0587	-0.0291	0.0415	0.0207	0.0397
nbppopprov	-0.5146	8.4086	-0.8243	5.4139	0.5722	8.4985	3.2929	5.3562
Intercept	-4.8983***	0.9807	-2.5256***	0.8594	-5.8514***	0.9205	-2.7006***	0.8067
Alpha	3.439022		2.9316		4.3567		3.2187	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	24.60***		3.44***		5.19***		2.99***	

Table E.9 : Seventh negative binomial model, dependent variable `necipub`, 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0091	0.0165	0.0284***	0.0104	0.0114**	0.0051	0.0160***	0.0048
ebper	0.0258***	0.0061	0.0076	0.0049	0.0322***	0.0058	0.0087**	0.0041
dspin	0.1642	0.3116	0.7787***	0.2745	0.0783	0.3101	0.5889**	0.2563
reusk	-2.4973	2.0725	1.1051	2.0303	-1.0977	2.1093	0.3129	1.8127
lfkreun	0.2213	0.1484	-0.0403	0.1377	0.1208	0.1481	0.0141	0.1215
dsubs	-	-	0.6955*	0.3765	-	-	0.3408	0.3368
rexpperb_0	-0.0080	0.0055	0.0004	0.0035	-0.0035	0.0044	0.0005	0.0031
sourcekgov	0.0063	0.0070	0.0048	0.0072	0.0092	0.0060	0.0039	0.0066
nbppopprov	-0.2963	7.7853	-3.2813	5.2625	3.1046	7.7603	-0.5360	4.7879
Intercept	-3.4521***	0.6904	-2.0839***	0.6906	-4.2889***	0.7014	-1.7500***	0.5923
Alpha	4.4316		3.1628		4.9682		3.3692	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	4.23***		4.05***		6.39***		3.62***	

Table E.10 : Fourth negative binomial model, dependent variable necc 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0309***	0.0095	0.0189**	0.0088	0.0125	0.0047	0.0160***	0.0044
ebper	0.0079	0.0049	0.0101**	0.0048	0.0111	0.0047	0.0100**	0.0040
dspin	0.2509	0.2404	0.2911	0.2714	0.2581	0.2306	0.3184	0.2615
dprod	0.8096	0.4964	0.7380*	0.3810	0.5393	0.4279	0.7990**	0.3309
dproc	0.5456*	0.3074	0.4040	0.2824	0.7651	0.3117	0.3458	0.2463
ndroitaec	-1.3080	0.9698	-0.0347	0.0916	0.2228	0.1598	-0.0477	0.0865
ndroitoec	-0.1746	0.1700	0.5335**	0.2400	-0.0944	0.1367	0.4801**	0.2310
dsubs	-	-	0.3429	0.3409	-	-	0.3058	0.2943
rexpperb_0	-0.0094***	0.0034	-0.0070**	0.0034	-0.0049	0.0035	-0.0045	0.0031
sourcekgov	-0.0130**	0.0057	-0.0010	0.0092	-0.0130	0.0056	0.0026	0.0070
lsourcekgovtot	-0.0060	0.0306	0.0164	0.0485	0.0332	0.0302	-0.0107	0.0320
nbppopprov	15.6791	6.2421	-0.4958	4.9749	19.2799	6.7055	3.9583	4.4100
Intercept	-3.2927***	0.7472	-1.9070***	0.7343	-3.5968	0.7196	-2.1470***	0.6465
Alpha	2.0145		2.2762		2.6051		2.3302	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	4.46***		3.10***		3.75***		3.33***	

Table E.11 : Seventh negative binomial model, dependent variable necc 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0330***	0.0104	0.0228**	0.0092	0.0165***	0.0051	0.0163***	0.0050
ebper	0.0096**	0.0043	0.0102**	0.0040	0.0121***	0.0046	0.0106***	0.0035
dspin	0.3908*	0.2309	0.2419	0.2646	0.3385	0.2264	0.3279	0.2499
reusk	0.5341	1.6326	2.1188	1.8918	0.6150	1.5371	1.9035	1.5127
lfkreun	0.0127	0.1134	-0.0983	0.1352	0.0007	0.1059	-0.0899	0.1010
dsubs	-	-	0.3680	0.3386	-	-	0.3000	0.2850
rexpperb_0	-0.0076	0.0036	-0.0045	0.0028	-0.0046	0.0035	-0.0018	0.0026
sourcekgov	-0.0109*	0.0061	-0.0063	0.0057	-0.0089*	0.0044	-0.0078	0.0054
nbppopprov	14.6318**	6.3215	-3.9400	4.6615	20.1645***	6.6152	0.4866	4.2350
Intercept	-2.7251***	0.5849	-0.9599*	0.5488	-3.0221***	0.6031	-1.1554**	0.4947
Alpha	2.0711		2.3093		2.7656		2.3925	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	5.03***		3.17***		4.77***		3.28***	

Table E.12 : Fourth negative binomial model, dependent variable necpc 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0126	0.0107	-0.0017	0.0156	0.0145**	0.0069	0.0039	0.0052
ebper	0.0114**	0.0055	-0.0010	0.0055	0.0190***	0.0059	0.0004	0.0048
dspin	0.6811**	0.3024	0.4628	0.3127	0.5807*	0.3006	0.5455*	0.3026
dprod	0.5602	0.3750	0.7178	0.6991	0.7493*	0.4020	0.4383	0.5938
dproc	0.6312*	0.3611	-0.1201	0.3775	0.8485**	0.3933	-0.2063	0.3654
ndroitaec	0.2223	0.3323	0.0586	0.1725	-0.1792	0.3016	0.0872	0.2178
ndroitoec	0.0348	0.2070	0.6951**	0.3498	0.0195	0.2142	0.7244**	0.3553
dsubs	-	-	-0.0477	0.5183	-	-	-0.4659	0.4493
rexpperb_0	-0.0006	0.0038	0.0027	0.0046	0.0013	0.0042	0.0055	0.0043
sourcekgov	-0.0049	0.0052	0.0097	0.0186	-0.0014	0.0045	0.0154	0.0174
lsourcekgovtot	0.0166	0.0292	0.0389	0.0740	0.0246	0.0253	0.0054	0.0688
nbppopprov	-1.0766	6.7548	-8.2828	7.4926	8.8318	7.3963	-6.2862	7.0336
Intercept	-3.1907***	0.7646	-1.3487	1.1922	-4.8077***	0.8844	-1.1019	0.9324
Alpha	3.0214		5.2165		4.2224		5.6928	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	1.95**		1.34		3.28***		1.42	

Table E.13 : Seventh negative binomial model, dependent variable necpc 2003 and 2005.

Variable	Small firms				Small and medium firms			
	2003		2005		2003		2005	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
e	0.0105	0.0139	0.0015	0.0158	0.0194***	0.0074	0.0029	0.0050
ebper	0.0099*	0.0059	0.0019	0.0060	0.0199***	0.0062	0.0023	0.0053
dspin	0.6820**	0.2933	0.5431	0.3290	0.5581*	0.2965	0.6643**	0.3208
reusk	1.3010	1.9810	3.9599	2.5667	2.8501	1.9603	5.7284**	2.3673
lfkreun	-0.0188	0.1413	-0.2389	0.1713	-0.1407	0.1361	-0.3693**	0.1557
dsubs	-	-	-0.2547	0.5437	-	-	-0.6987	0.4587
rexpperb_0	0.0004	0.0039	0.0044	0.0050	0.0024	0.0043	0.0062	0.0047
sourcekgov	-0.0007	0.0051	0.0109	0.0106	0.0022	0.0041	0.0093	0.0107
nbppopprov	3.0253	6.6642	-13.1010	7.1103	13.1545*	7.4934	-9.5536	6.7408
Intercept	-2.8012***	0.6423	-0.6640	0.8578	-4.2869***	0.7400	-0.6292	0.7349
Alpha	2.6617		5.5500		4.2463		5.7333	
Observations	227		266		271		318	
Weighted observations	352		397		428		480	
F-statstic	4.21***		1.54		5.59***		2.02**	

ANNEX F

COEFFICIENTS OF CORRELATION

Table F.1 : Correlation coefficients, 2005 survey.

	e	ebper	dspin	dsubs	dprod	dproc	nprodrd	nprodpc	nprodrc	nprodpm	nbe	nba	ndroitaec
e	1.00												
ebper	-0.29	1.00											
dspin	-0.11	0.31	1.00										
dsubs	-0.04	0.20	0.07	1.00									
dprod	-0.17	0.23	0.11	-0.06	1.00								
dproc	0.04	0.05	-0.19	0.00	0.04	1.00							
nprod	-0.01	-0.02	-0.03	0.02	0.01	0.05							
nprodrd	-0.01	-0.02	-0.02	0.03	0.01	0.04	1.00						
nprodpc	-0.05	0.06	0.03	0.03	0.06	0.08	0.10	1.00					
nprodrc	-0.04	0.02	0.06	-0.09	0.10	0.07	0.10	0.29	1.00				
nprodpm	0.04	0.06	-0.08	-0.08	0.00	0.10	0.00	0.08	0.01	1.00			
nbe	-0.02	0.12	0.18	0.02	0.12	0.01	-0.02	0.14	0.21	-0.01	1.00		
nba	-0.02	0.12	0.19	-0.11	0.14	0.00	-0.02	0.01	0.12	-0.05	0.39	1.00	
ndroitaec	-0.02	-0.03	-0.05	0.02	0.04	0.06	**0.91**	0.09	0.09	-0.02	-0.02	0.01	1.00

Table F.1 : Correlation coefficients, 2005 survey (continuation).

	e	ebper	dspin	dsubs	dprod	dproc	nprodrd	nprodpc	nprodrc	nprodpm	nbe	nba	ndroitaec
ndroitoec	-0.01	0.06	0.09	0.04	0.04	0.05	0.00	0.01	0.03	-0.03	0.03	0.31	0.06
ncont	-0.02	0.11	0.16	0.07	0.06	0.02	0.03	0.10	**0.52**	0.01	0.12	0.12	0.02
nacont	-0.03	0.14	-0.02	0.02	0.05	0.09	0.01	0.01	0.11	0.28	-0.04	-0.06	0.00
rt_0	0.70	-0.24	-0.09	0.00	-0.19	-0.05	-0.01	-0.04	-0.04	0.00	-0.03	-0.04	-0.01
rdt_0	**0.47**	-0.11	0.00	-0.23	0.07	0.08	-0.02	-0.04	0.09	0.00	0.11	0.20	-0.02
rdbper0	-0.28	**0.46**	0.22	0.06	0.31	0.12	-0.08	0.09	0.11	-0.04	0.13	0.15	-0.08
reusk	-0.10	0.29	0.15	0.14	0.04	0.02	-0.04	0.08	0.06	0.00	0.17	0.23	-0.02
fkreun	-0.03	0.14	0.14	-0.06	0.06	0.00	-0.02	-0.02	0.07	-0.03	0.13	0.23	0.00
sourcekgov	0.00	0.15	-0.06	0.03	0.01	0.12	-0.02	0.03	-0.03	-0.01	-0.02	0.08	-0.01
lsourcekgovtot	-0.03	0.16	-0.02	0.01	0.00	0.14	-0.02	0.08	0.02	0.00	0.07	0.09	0.00
nbppopprov	0.01	0.04	0.09	-0.03	-0.12	0.05	0.05	0.07	0.10	0.01	0.04	0.06	0.06
lrt_0	0.25	-0.31	-0.26	-0.07	-0.13	0.08	0.03	0.01	0.03	0.16	-0.06	-0.04	0.03
lrdt_0	0.17	0.16	0.15	-0.07	0.32	0.23	-0.01	0.09	0.21	-0.05	0.22	0.28	0.02
lfkreun	-0.10	0.29	0.16	0.13	0.05	0.01	-0.04	0.08	0.07	0.00	0.19	0.25	-0.02
rexpperb_0	-0.07	0.14	-0.11	0.09	0.13	0.23	-0.02	0.10	0.10	0.21	0.04	-0.05	-0.04
rbper0	-0.16	0.32	0.00	0.09	0.08	0.12	-0.02	0.03	0.06	0.15	0.03	0.06	-0.03

Table F.1 : Correlation coefficients, 2005 survey (end).

	ndroitoec	ncont	nacont	rt_0	rdt_0	rdbper0	reusk	fkreun	sourcekgov	lsourcekgovtot	nbppopprov	lrt_0	lrdt_0	lfkreun	rexpperb_0	rbper0
ndroitoec	1.00															
ncont	0.06	1.00														
nacont	-0.03	0.22	1.00													
rt_0	-0.01	-0.03	-0.03	1.00												
rdt_0	0.07	0.04	-0.04	0.21	1.00											
rdbper0	0.04	0.10	0.02	-0.25	-0.10	1.00										
reusk	0.10	0.03	-0.06	-0.09	-0.02	0.25	1.00									
fkreun	**0.47**	0.04	-0.06	-0.03	0.13	0.14	**0.41**	1.00								
sourcekgov	-0.03	-0.06	0.03	-0.02	0.04	0.10	0.37	-0.02	1.00							
lsourcekgovtot	-0.02	-0.04	0.04	-0.04	0.05	0.11	**0.43**	0.11	**0.86**	1.00						
nbppopprov	0.00	0.02	-0.01	0.01	0.06	-0.02	0.08	0.06	0.01	0.02	1.00					
lrt_0	-0.11	0.04	0.10	0.23	0.09	-0.26	-0.21	-0.18	0.01	-0.04	0.05	1.00				
lrdt_0	0.09	0.16	0.01	0.10	0.37	**0.46**	0.16	0.20	0.04	0.09	-0.04	-0.02	1.00			
lfkreun	0.13	0.04	-0.07	-0.09	0.00	0.25	**0.99**	**0.49**	0.31	**0.40**	0.08	-0.22	0.19	1.00		
rexpperb_0	-0.04	0.05	0.21	-0.06	-0.05	0.13	0.02	-0.05	0.13	0.13	0.02	0.25	0.04	0.02	1.00	
rbper0	-0.05	0.12	0.17	-0.13	-0.07	0.26	0.12	-0.04	0.13	0.11	0.03	**0.47**	0.04	0.10	0.41	1.00

Table F.2 : Correlation coefficients, 2003 survey.

	e	ebper	dspin	dprod	dproc	nprod	nprodrd	nprodpc	nprodrc	nprodpm	nbe	nba	ndroitaec	ndroitoec
e	1.00													
ebper	-0.28	1.00												
dspin	-0.12	0.33	1.00											
dprod	0.04	0.07	0.05	1.00										
dproc	0.05	0.14	0.06	0.10	1.00									
nprod	0.00	-0.02	0.06	0.04	0.02	1.00								
nprodrd	-0.01	-0.05	-0.06	0.05	-0.05	0.36	1.00							
nprodpc	-0.01	-0.02	-0.04	0.05	-0.09	0.27	**0.83**	1.00						
nprodrc	-0.01	0.02	0.00	0.10	-0.04	0.04	0.16	0.25	1.00					
nprodpm	0.01	-0.01	0.08	0.03	0.04	**0.95**	0.04	-0.01	-0.02	1.00				
nbe	0.02	0.07	0.09	0.08	0.07	-0.02	0.00	0.03	0.21	-0.02	1.00			
nba	0.02	0.11	0.09	0.10	0.05	-0.02	-0.02	0.02	0.21	-0.02	**0.82**	1.00		
ndroitaec	0.01	-0.14	-0.07	0.04	0.02	0.04	0.13	0.14	0.04	-0.01	-0.02	-0.02	1.00	
ndroitoec	0.02	0.00	0.10	0.07	0.04	-0.02	-0.01	0.00	-0.04	-0.02	0.16	0.14	0.01	1.00

Table F.2 : Correlation coefficients, 2003 survey (continuation).

	e	ebper	dspin	dprod	dproc	nprod	nprodrd	nprodpc	nprodrc	nprodpm	nbe	nba	ndroitaec	ndroitoec
ncont	-0.03	0.12	0.18	0.07	0.02	-0.02	-0.03	0.00	0.30	-0.02	0.38	0.34	0.00	-0.04
nacont	-0.03	0.04	0.03	-0.02	0.07	0.03	0.00	-0.01	-0.01	0.04	-0.04	0.00	-0.02	-0.01
rt_0	**0.96**	-0.31	-0.14	0.02	0.06	-0.01	-0.01	-0.01	-0.01	0.00	0.01	-0.02	0.06	0.02
rdt_0	0.06	-0.02	0.00	0.07	-0.09	-0.01	0.00	0.00	0.21	-0.01	0.24	0.22	-0.01	0.25
rdbper0	-0.05	**0.49**	0.29	0.24	0.04	0.02	-0.04	0.02	0.09	0.03	-0.02	0.00	-0.16	0.01
reusk	-0.11	0.21	0.17	0.16	0.03	0.05	-0.06	-0.03	0.04	0.07	0.12	0.15	-0.05	0.10
fkreun	-0.01	-0.01	0.08	0.02	-0.05	0.15	-0.03	-0.01	0.06	0.17	0.27	0.27	-0.02	**0.54**
sourcekgov	-0.05	0.12	0.03	-0.06	0.09	0.00	0.02	0.00	-0.04	-0.01	-0.06	-0.07	-0.03	-0.03
lsourcekgovtot	-0.03	0.03	0.04	0.04	0.06	0.01	0.00	0.00	0.10	0.01	-0.05	-0.02	0.00	-0.02
nbppopprov	0.04	-0.04	-0.02	0.05	-0.03	0.02	0.07	0.08	0.05	-0.01	0.03	0.02	0.07	0.03
rexpperb_0	-0.05	-0.04	-0.04	0.08	0.07	0.15	-0.02	-0.05	0.11	0.16	-0.06	-0.08	-0.05	0.12
lrt_0	0.32	**-0.50**	-0.30	-0.07	0.01	0.11	0.06	0.04	0.14	0.10	0.10	0.04	0.12	0.01
lrdt_0	-0.04	0.12	0.11	0.24	0.14	0.04	0.03	0.02	0.25	0.03	0.27	0.30	0.05	0.15
lfkreun	-0.10	0.20	0.18	0.16	0.03	0.07	-0.05	-0.03	0.06	0.09	0.15	0.19	-0.05	0.13
rbper0	-0.12	0.10	0.04	-0.01	-0.01	0.07	-0.05	-0.02	0.14	0.09	0.01	0.03	-0.07	0.05

Table F.2 : Correlation coefficients, 2003 survey (end).

	ncont	nacont	rt_0	rdt_0	rdbper0	reusk	fkreun	sourcekgov	lsourcekgovtot	nbppopprov	rexpperb_0	lrt_0	lrdt_0	lfkreun	rbper0
ncont	1.00														
nacont	-0.04	1.00													
rt_0	-0.03	-0.02	1.00												
rdt_0	0.11	-0.04	0.08	1.00											
rdbper0	0.15	0.01	-0.10	-0.13	1.00										
reusk	0.22	-0.04	-0.13	0.06	0.21	1.00									
fkreun	0.20	-0.02	-0.01	0.20	0.07	**0.33**	1.00								
sourcekgov	0.00	0.01	-0.07	-0.07	0.02	0.09	-0.04	1.00							
lsourcekgovtot	0.04	0.08	-0.04	-0.03	0.03	0.37	0.05	**0.56**	1.00						
nbppopprov	0.02	0.00	0.06	0.07	-0.04	0.00	-0.02	-0.11	-0.05	1.00					
rexpperb_0	-0.02	0.11	-0.06	-0.06	0.06	0.02	0.10	-0.03	0.11	-0.07	1.00				
lrt_0	-0.01	0.11	0.38	0.06	-0.31	-0.26	0.05	-0.14	-0.04	-0.03	0.24	1.00			
lrdt_0	0.26	-0.03	-0.02	**0.48**	0.18	0.22	0.19	-0.07	0.06	0.06	-0.08	0.04	1.00		
lfkreun	0.27	-0.04	-0.13	0.08	0.22	**0.99**	0.42	0.06	0.35	0.00	0.02	-0.25	0.26	1.00	
rbper0	0.02	0.15	-0.15	-0.06	0.19	-0.03	0.03	-0.14	-0.05	0.02	0.40	0.32	-0.12	-0.03	1.00

Table F.3 : Correlation coefficients, 2001 survey.

	e	ebper	dspin	dprod	dproc	nprodrd	nprodpc	nprodrc	nprodpm	nbe	nba	ndroitaec	ndroitoec	ncont
e	1.00													
ebper	-0.19	1.00												
dspin	-0.14	0.17	1.00											
dprod	-0.04	0.16	-0.01	1.00										
dproc	0.05	-0.12	-0.01	0.24	1.00									
nprodrd	0.01	-0.07	-0.04	0.06	0.02	1.00								
nprodpc	0.02	-0.06	-0.06	0.03	0.08	-0.02	1.00							
nprodrc	0.00	-0.05	-0.08	0.03	0.06	0.22	**0.53**	1.00						
nprodpm	-0.01	-0.06	0.06	0.04	0.03	0.30	-0.02	-0.01	1.00					
nbe	0.08	-0.06	-0.05	0.06	0.06	-0.01	0.10	0.00	0.00	1.00				
nba	0.06	-0.01	0.00	0.05	0.05	-0.02	0.08	-0.01	0.00	**0.96**	1.00			
ndroitaec	-0.01	-0.05	0.07	0.03	0.06	0.31	0.07	0.04	**0.94**	0.01	-0.01	1.00		
ndroitoec	0.01	-0.04	-0.02	0.05	0.07	0.01	0.89	0.52	0.00	0.17	0.16	0.07	1.00	
ncont	-0.02	0.04	0.11	0.08	0.00	-0.03	0.03	-0.01	-0.02	0.36	0.38	0.00	-0.01	1.00

Table F.3 : Correlation coefficients, 2001 survey (continuation).

	e	ebper	dspin	dprod	dproc	nprod	nprodrd	nprodpc	nprodrc	nprodpm	nbe	nba	ndroitaec	ndroitoec	ncont
nacont	0.00	-0.06	-0.03	0.01	0.07	0.06	0.02	0.01	-0.01	0.06	-0.02	-0.03	0.06	-0.01	-0.03
rt_0	**0.91**	-0.19	-0.14	-0.02	0.03	0.00	0.01	0.07	0.03	-0.01	0.11	0.07	0.00	0.05	-0.02
rdt_0	**0.49**	-0.05	-0.05	0.05	0.00	-0.02	0.01	0.03	0.00	-0.02	0.26	0.33	-0.02	0.04	0.16
rdbper0	-0.33	0.24	0.22	0.13	-0.03	-0.03	-0.17	0.00	-0.14	0.05	-0.02	0.00	0.04	0.02	0.07
reusk	-0.02	0.19	0.29	0.07	0.10	0.03	-0.02	-0.03	-0.07	0.06	-0.07	-0.04	0.09	-0.01	0.04
fkreun	-0.01	0.18	0.15	0.05	0.09	0.20	0.07	-0.01	-0.02	0.21	0.00	0.06	0.25	0.00	0.16
sourcekgov	0.15	-0.12	0.02	0.06	0.01	-0.04	-0.04	-0.04	-0.03	-0.03	-0.05	-0.06	-0.01	-0.01	-0.09
lsourcekgovtot	0.12	-0.02	0.13	0.08	0.11	-0.04	-0.03	-0.03	-0.03	-0.03	-0.05	-0.06	0.00	-0.01	-0.07
nbppopprov	-0.05	-0.04	-0.06	-0.03	0.01	-0.03	-0.05	0.04	0.01	-0.02	0.00	0.01	0.00	0.03	0.00
lrt_0	0.38	-0.36	-0.37	-0.04	0.04	0.12	0.13	0.14	0.11	0.08	0.19	0.14	0.08	0.10	0.00
lrdt_0	0.23	0.09	0.06	0.12	0.11	0.03	0.05	0.08	0.08	0.01	0.25	0.32	0.02	0.11	0.30
lfkreun	-0.02	0.21	0.29	0.07	0.10	0.05	-0.01	-0.03	-0.07	0.08	-0.07	-0.03	0.11	-0.01	0.06
rbper0	-0.16	0.04	-0.05	0.05	0.05	0.08	-0.04	-0.02	0.01	0.11	0.01	0.02	0.06	-0.06	-0.10
rexpperb_0	-0.06	-0.12	-0.09	0.00	0.07	0.11	0.01	-0.03	-0.06	0.13	-0.08	-0.12	0.12	-0.05	-0.14

Table F.3 : Correlation coefficients, 2001 survey (end).

	nacont	rt_0	rdt_0	rdbper0	reusk	fkreun	sourcekgov	lsourcekgovtot	nbppopprov	lrt_0	lrdt_0	lfkreun	rbper0	rexpperb_0
nacont	1.00													
rt_0	-0.02	1.00												
rdt_0	0.03	**0.42**	1.00											
rdbper0	-0.13	-0.35	-0.21	1.00										
reusk	-0.02	-0.06	0.00	0.20	1.00									
fkreun	-0.01	-0.03	0.11	0.12	**0.43**	1.00								
sourcekgov	-0.04	0.12	0.11	-0.08	0.15	-0.02	1.00							
lsourcekgovtot	-0.04	0.09	0.11	0.01	**0.41**	0.04	**0.56**	1.00						
nbppopprov	0.07	0.01	0.02	-0.04	0.00	0.06	-0.01	-0.02	1.00					
lrt_0	0.08	0.39	0.28	-0.35	-0.32	0.01	-0.10	-0.11	0.11	1.00				
lrdt_0	0.09	0.22	**0.58**	-0.11	0.13	0.24	-0.06	0.00	0.07	0.18	1.00			
lfkreun	-0.02	-0.06	0.02	0.20	**0.99**	**0.52**	0.11	0.37	0.01	-0.30	0.16	1.00		
rbper0	-0.03	-0.16	-0.09	0.29	-0.07	0.07	-0.11	-0.01	0.12	0.27	-0.20	-0.06	1.00	
rexpperb_0	0.02	-0.05	-0.15	0.15	-0.15	-0.07	-0.02	-0.01	0.07	0.17	-0.25	-0.16	**0.42**	1.00

Table F.4 : Correlation coefficients, 1999 survey.

	e	ebper	dspin	dprod	dproc	nprod	nprodrd	nprodpc	nprodrc	nprodpm	nbe	nba
e	1.00											
ebper	-0.27	1.00										
dspin	0.04	0.05	1.00									
dprod	-0.13	-0.08	0.05	1.00								
dproc	-0.08	-0.17	-0.11	**0.54**	1.00							
nprod	0.06	-0.03	-0.13	0.10	0.11	1.00						
nprodrd	0.07	-0.03	-0.16	0.09	0.09	**0.95**	1.00					
nprodpc	0.07	0.00	-0.12	0.09	0.08	**0.79**	**0.84**	1.00				
nprodrc	0.08	-0.02	-0.15	0.08	0.08	**0.92**	**0.97**	**0.94**	1.00			
nprodpm	0.01	-0.03	-0.03	0.10	0.11	**0.79**	**0.57**	0.38	**0.51**	1.00		
nbe	**0.65**	-0.01	0.13	-0.11	-0.09	0.01	-0.01	-0.01	-0.01	0.05	1.00	
nba	**0.42**	-0.07	0.11	0.05	0.04	0.07	0.08	0.12	0.10	0.00	**0.49**	1.00

Table F.4 : Correlation coefficients, 1999 survey (continuation).

	e	ebper	dspin	dprod	dproc	nprod	nprodrd	nprodpc	nprodrc	nprodpm	nbe	nba
rt_0	**0.82**	-0.28	0.03	-0.12	-0.08	0.00	-0.01	0.02	0.01	0.02	**0.45**	0.21
rdt_0	**0.67**	-0.01	0.03	-0.04	-0.04	0.11	0.13	0.17	0.16	0.01	**0.71**	**0.45**
rdbper0	0.09	0.13	0.02	0.02	0.09	0.08	0.07	0.06	0.06	0.08	0.07	0.00
fkreun	0.16	-0.14	-0.09	0.01	0.03	0.00	0.00	0.01	0.00	-0.01	0.21	0.12
sourcekgov	-0.12	0.08	-0.09	0.19	0.10	-0.06	-0.05	-0.04	-0.05	-0.06	-0.09	-0.07
lsourcekgovtot	-0.08	0.08	-0.14	0.20	0.11	-0.08	-0.07	-0.05	-0.06	-0.08	-0.11	-0.09
nbppopprov	-0.13	-0.06	-0.03	0.06	0.19	-0.07	-0.07	-0.07	-0.08	-0.05	-0.03	-0.04
lrt_0	**0.47**	-0.16	-0.18	-0.19	-0.06	0.13	0.10	0.10	0.11	0.15	0.18	0.02
lrdt_0	0.39	-0.01	0.11	0.05	0.03	0.09	0.13	0.15	0.14	-0.03	0.27	0.31
lfkreun	0.31	-0.15	-0.06	0.12	0.23	0.15	0.13	0.13	0.13	0.13	0.29	0.27
rexpperb_0	0.28	-0.02	-0.11	0.08	0.13	0.11	0.07	0.14	0.10	0.13	0.19	0.09
rbper0	0.21	0.11	0.05	-0.05	0.00	0.13	0.10	0.07	0.09	0.16	0.18	0.06

Table F.4 : Correlation coefficients, 1999 survey (end).

	rt_0	rdt_0	rdbper0	fkreun	sourcekgov	lsourcekgovtot	nbppopprov	lrt_0	lrdt_0	lfkreun	rexpperb_0	rbper0
rt_0	1.00											
rdt_0	**0.58**	1.00										
rdbper0	0.01	0.07	1.00									
fkreun	0.08	0.15	0.07	1.00								
sourcekgov	-0.12	-0.12	0.06	-0.06	1.00							
lsourcekgovtot	-0.13	-0.11	0.11	-0.07	**0.79**	1.00						
nbppopprov	-0.06	0.02	0.12	-0.05	0.12	0.01	1.00					
lrt_0	**0.59**	0.35	0.17	0.09	-0.03	-0.05	0.00	1.00				
lrdt_0	0.24	**0.61**	0.19	0.00	-0.09	-0.04	0.03	0.23	1.00			
lfkreun	0.12	0.31	0.08	**0.40**	-0.13	-0.06	-0.20	0.08	0.31	1.00		
rexpperb_0	0.25	0.16	0.22	0.17	0.01	0.00	-0.03	**0.46**	0.04	0.24	1.00	
rbper0	0.18	0.15	0.36	0.03	0.10	0.03	0.07	**0.59**	-0.02	0.00	**0.57**	1.00

ANNEX G

ECONOMETRIC METHODS

In this annex, we provide the necessary theoretical basis to tackle the analysis of a complex survey using models for binary outcomes and for count variables. First, an introduction to sampling methods deals with simple random sampling (as the fundamental basis for sampling procedures) and with stratification. A detailed introduction to Maximum Likelihood estimation is then presented, as this is the technique employed for fitting models for limited dependent and count variables. Logit, probit, Poisson and negative binomial regression are presented in the last subsections, along with specific testing methods and diagnostics. As we do not purport to provide a detailed and exhaustive monograph on this issue, it is recommended to consult the bibliography to deepen the understanding on the methods employed for econometric analysis.

G.1 SURVEY SAMPLING METHODS

This section deals with the main methods of sampling for surveys, with a particular focus on stratification, which is the technique employed in the Biotechnology Use and Development Surveys (BUDS) made by Statistics Canada. This group of techniques aim at reducing time and cost of computing summary statistics and of fitting models on the entire target population by using a sample. Results can be then extended to the whole population through an inference process, which intrinsically introduces a certain level of incertitude on the measures. A compromise is therefore necessary between cost and time required and the precision of the results; the way estimators are calculated must take into account the specific sampling design in order to obtain unbiased measures with the smallest possible variance. Moreover, the characteristics of the population and of its elements influence the choice of which method should be used and which should be excluded. The preliminary analysis on the population is, for econometricians, an essential step, as a refined sampling design can considerably improve the reliability of the estimates. The most useful text dealing with this topic is Levy and Lemeshow (1999): an overview on the main features of sampling of populations is provided, along with exam-

ples and applications. Many other texts are available, and in the following, they will be referenced when dealing with particular details of this topic.

G.1.1 SIMPLE RANDOM SAMPLING

Let us suppose the entire population consists of N elements, from which n elements are drawn through a completely random process. This is the simplest way to obtain a sample from a population. The main summary statistics the practitioner is usually interested in are the population estimates of mean, total, variance and, for a binary variable, proportion. The sample mean is calculated using the well-known formula:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

which is an unbiased estimator of the population mean. The standard error of the estimate of the mean is:

$$SE(\bar{x}) = \sqrt{\frac{N-n}{N}}\left(\frac{s_{\bar{x}}}{\sqrt{n}}\right)$$

where $s_{\bar{x}}$ is the estimated standard error of the sample, defined as usual as:

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

The estimate of the population total x^t is obtained simply multiplying the mean by the number of elements of the population N:

$$x^t = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and its variance is:

$$SE(x^t) = N\sqrt{\frac{N-n}{N}}\left(\frac{s_{\bar{x}}}{\sqrt{n}}\right)$$

If the variable x is binary, the mean takes the meaning of the proportion:

$$p = \frac{1}{n}\sum_{i=1}^{n} x_i$$

In this case, its standard error is takes the following form:

$$SE(p) = \sqrt{\left(\frac{N-n}{N}\right)\left(\frac{p(1-p)}{n-1}\right)}$$

Note that the value of the estimated standard errors grows as the number of the elements included in the sample reduces: intuitively, this happens because it is harder to estimate the parameter of the population by including too few elements in the sample. Conversely, as the sample size approaches to N, the standard error reduces and is 0 when $n = N$: in this case the sample includes the whole population, and the true parameter is directly calculated. This effect is even clearer if we examine the interval estimates of the population parameters (for simplicity, we only consider the case of the population mean). For example, at the $(1-\alpha)\times 100\%$ confidence, the interval estimate of the population mean is:

$$\bar{x} \pm z_{\left(1-\frac{\alpha}{2}\right)} \sqrt{\frac{N-n}{N}} \left(\frac{s_{\bar{x}}}{\sqrt{n}} \right)$$

Most programs for statistical and econometric analysis require as input the *sampling weights*, defined as the inverse of the probability of an element to be included in the sample. For instance, if the population consists of 150 elements, and 10 elements are included in the sample, the sampling weight is equal to 15. Another important coefficient often used by softwares is the *finite population correction* (fpc), defined as:

$$fpc = \frac{N-n}{N-1}$$

Although its simplicity, in practice simple random sampling is rarely used in survey designs, as it requires a-priori identification of all the elements of the population. Once they are labelled, a random sample can be drawn, but this is an expensive, and sometime just impossible, procedure. Further, especially in the case of a geographically spread population and when a direct interview is required, this method generates a sample with elements over a large area. In these cases, it could be convenient to employ clustered sampling, which we do not present here. If interested, the reader can consult the texts indicated in the bibliography.

When analyzing a survey, it is often required to provide summary statistics for subgroups of the population. In our case, for example, we investigate the differences between small, medium and large firms with respect to some of their characteristics. This requires the entire sample to be divided in subsamples, and the estimates to be calculated within each subsample. Levy and Lemeshow (1999) recommend proceeding this way: first create a dummy variable y_i which takes the value 1 if the ith element belongs to the identified subgroup, 0 otherwise. Let y be the total of the dummy variable. Clearly, this quantity is the number of the elements of the subpopulation that are included in the sam-

ple. Now suppose we are interested in computing the average of the variable x within the subsample. Firs we create the variable z defined as:

$$z = \sum_{i=1}^{n} z_i = \sum_{i=1}^{n} x_i \cdot y_i$$

Thus z is the total of the variable x within the subgroup. It can be shown that the ratio z/y is an unbiased estimator of the subgroup mean, whose variance is unfortunately unknown. However, if the number of the elements in the subgroup is greater than or equal to 20, it can be shown that the following approximation holds:

$$SE(z/y) \cong \left(\frac{\sigma_z}{\sqrt{E(y)}} \right) \sqrt{\frac{Y - E(y)}{Y - 1}}$$

where $E(Y)$ can be estimated as through y':

$$y' = y \cdot \left(\frac{N}{n} \right)$$

and the variance of the elements of the subgroup through $\hat{\sigma}_z$:

$$\hat{\sigma}_z = \sqrt{\left(\frac{y'-1}{y'} \right)} \cdot \sqrt{\sum_{i=1}^{y} \frac{(z_i - \bar{z})^2}{y-1}}$$

where $\bar{z}$ is the sample mean of z_i.

The fact that the variance of the subgroup estimate is not known is the main shortcoming of computing summary statistics for subgroups within a simple random sample. The

sampling design should take into account the need of calculating statistics for subgroups, and a more refined technique is required. A widely adopted framework implies the preliminary division of the population into subgroups, and the drawing of a random sampling within each subgroup. Next subsection deals with this technique, commonly called stratification.

G.1.2 STRATIFIED SAMPLING

In order to increase the precision of the estimates of one or more characteristics of a certain subdomain, it is possible to divide the entire population in mutually exclusive subdomains; within each subdomain, a sample is independently drawn using the simple random sampling technique discussed above. Following Levy and Lemeshow's notation, let L be the number of strata and n_h the number of elements drawn within each stratum, where h indicates the stratum. For example, let us suppose we are interested in knowing how many firms in Canada are involved in a collaborative arrangement with another firm; further, we are particularly interested in a by-province estimate of this characteristic. The total of the variable x within the stratum h is:

$$x_h^t = \sum_{i=1}^{n_h} x_{h,i}$$

The stratum mean is obtained by dividing the stratum total by the number of the elements included in the sample drawn from that specific stratum:

$$\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{h,i} = \frac{1}{n_h} x_h^t$$

Of course, this definition holds also in the case of a binary variable, and the mean is called, as usual, proportion.

The estimate of the mean of the entire population can be interpreted as the weighted average of each stratum mean, where the weights W_h are given by the proportion of the elements belonging to the stratum with respect to the entire population:

$$\bar{x} = \frac{\sum_{h=1}^{L} N_h \cdot \bar{x}_h}{N} = \sum_{h=1}^{L} W_h \cdot \bar{x}_h$$

Clearly, this form is equivalent to the one when a simple random sampling technique is adopted, as

$$\sum_{h=1}^{L} W_h \cdot \bar{x}_h = \frac{x^t}{N}$$

The proportion of a variable x for the entire population can be calculated following the same considerations as above, obtaining:

$$p_x = \frac{\sum_{h=1}^{L} N_h \cdot p_{x,h}}{N} = \sum_{h=1}^{N} W_h \cdot p_{x,h}$$

Within a stratum h, the estimated variance of the distribution of the variable x is:

$$s_{x,h}^2 = \frac{\sum_{i=1}^{N_h} \left(x_{h,i} - \bar{x}_h\right)^2}{N_h}$$

The standard error estimates for the entire population reflect the way the entire population has been divided and the proportion of elements included in the sample within each stratum. The standard error estimate of the mean for the entire population is:

$$SE(\bar{x}_{str}) = \sqrt{\sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \frac{s_{x,h}^2}{n_n} \left(\frac{N_h - n_h}{N_h}\right)}$$

This can be interpreted as a weighted sum of the variances of x within each stratum. The term $(N_h/N)^2$ takes into account the portion of the entire population represented by the stratum h. The smaller this proportion, the smaller the value of this ratio, and thus the smaller the weight of the variance of that stratum. Through the term $(N_h - n_h/N_h)$ the portion of the elements of the stratum h that are included in the sample is considered. As in the case of simple random sampling, the more n_h tends towards N_h, the more precise the estimate is, and the smaller its variance.

The estimated standard error of the total of x for the population is:

$$SE(x_{str}^t) = \sqrt{\sum_{h=1}^{L} \frac{N_h^2 \cdot s_h^2}{n_h} \cdot \left(\frac{N_h - n_h}{N_h}\right)}$$

Finally, the estimated variance of the proportion is computed as:

$$SE(p_{str}) = \sqrt{\sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \frac{p_{x,h}(1 - p_{x,h})}{n_h - 1} \cdot \left(\frac{N_h - n_h}{N_h}\right)}.$$

It has been said above that estimated means, totals and proportions of a subgroup are unbiased estimates for the corresponding measures of the population. Unfortunately, in the case of a stratified random sample, it could not hold. This comes from the fact that each stratum is representative of a certain portion of the entire population which is a-priori identified.

The main advantage of stratification is clear when we are interested in estimating the mean (or total, or proportion) of a variable among the elements belonging to the same

stratum. In other words, the subgroups are identified by the strata. In this case, estimates for each stratum are computed as if the stratum was a simple random sample, and the same formulas introduced in the preceding subsection hold. Comparing two estimates involves a t-test on two samples, with, in general, unknown variances and means.
The procedure followed in the analysis of the BUDS partially follows this framework: to obtain means and totals, the whole sample has been divided by province and by firm size. Actually, the technique used by the statisticians at Statistics Canada is slightly different, involving stratification by size, province and industrial code. If such stratification were formally implemented, it would lead to strata with an exiguous number of elements, which would invalidate any analysis. In reality, the survey was addressed to all of the firms of the population, and a post-stratification technique was adopted to take into account the effect of nonresponse rate. Next paragraph introduces this method, which has the benefit of providing a virtually stratified sample by post-weighting the observations according to the response rate.

As anticipated above, sometimes strata are formed after a simple random sample (or another type of sample) is taken. This way, every element is allocated to a specific stratum and given a sampling weight, which is computed after strata have been identified. For example, let us suppose we are interested in measuring the mean of a certain characteristic of Canadian households by province using post stratification. First, we draw a simple random sample from all of the Canadian households. After collecting data, we count the proportion of households in the sample with respect to the number of households in the entire population. The inverse of this ratio is the sampling weight associated with each element belonging to that specific stratum. For this reason post stratification is also called stratification after selection by weighting (Kish, 1967). As we will see this is not exactly the technique used in the Biotechnology Use and Development Surveys: the survey was mailed to the whole population, but not all the firms responded. In this case, the response rate determines the sampling weights, within each stratum. Therefore, the sampling weights serve only as a correction that takes into account the response rate. As ex-

plained in the Statistics Canada website, three levels of virtual stratification are used: size, province and industrial code (NAICS). This procedure could generate strata with an exiguous number of elements, leading to unreliable estimates or to a too small number of degrees of freedom. For this reason in this situation, a formal definition of strata is not used, and the pseudo-stratification is obtained using the sampling weights only.

G.1.3 MISSING OBSERVATIONS

One of the main problems a practitioner has to face when tackling a survey analysis involves missing observations, which can be divided into two broad classes: unit nonresponse and item nonresponse. The former refers to the case when for certain units no information has been obtained at all. In the case of the BUDS, this issue is taken into account using sampling weights, as described above. Item nonresponse refers to the partial lack of response in the questionnaire from a unit. In other words, one or more questions have not been answered in the questionnaire. To fix this problem, which constitutes a major issue in survey analysis, several techniques are available, each of which involves some disadvantages, as we will see. Four broad categories of methods are available to overcome the problem introduced by item nonresponse: complete case methods (only complete questionnaires are used for analysis), imputation methods (missing values are replaced), re-weighting methods (weights are modified in order to take into account missing values) and model-based methods (employing maximum-likelihood maximization procedures). Imputation methods are widely adopted because missing values decrease the number of observations available for model fitting, and a higher number of degrees of freedom are obtained through replacing nonresponses.

We first introduce the imputation method for item nonresponse called call-back, requiring the practitioner to follow-up on nonresponses. This activity is always very expensive time-consuming, and the benefit could not match the expectations. For this reason, the decision to undertake a call-back procedure should be carefully evaluated (Zarkovich, 1963). Methodologists and econometricians at Statistics Canada adopt this method to improve data reliability.

Another imputation method, which can take several forms, consists in substituting the missing value. The logic behind the substitution defines the different methods gathered under this name. For example, missing values can be replaced by the mean of non-missing items; this way the reliability of the estimate does not improve, as the information is taken from existing data: no additional new values are added. Many variations of this method are proposed, but all of them consist in replacing nonresponses using existing observations in order to create new consistent values. Sometimes specific solutions can be developed for specific situations, as it has been done in this work to analyze the BUDS, where missing values often have been replaced with 0's in the presence of dependence between two variables. For example, consider the case of a nested question: if the respondent answers positively to question A, than he is required to answer to question B, otherwise he is required to skip it. If A is negatively answered, then we expect to have a missing value for B. It is possible to replace, for most of these cases, the missing value with a zero.

As this subsection is just intended to make the reader aware of the problem of missing data, we will not present the other methods dealing with item nonresponse. If interested, the reader can consult the literature dealing with this topic. An introduction to data issue in survey is provided by Griliches (1986); for more details on imputation methods, see Levy and Lemeshow (1999), Durbin (1954), Hendricks (1949), Little (1987) and Kalton (1983).

G.2 ESTIMATION AND TEST OF FIT

Under the name of estimation methods we mean those techniques aimed at computing estimates of the parameters of a population or of the coefficients of a model. In the preceding section, we introduced some basic estimates: mean, proportion, and total of a variable for a population. When fitting a model to observed data, more refined procedures are required, building on more sophisticated theoretical bases. A variety of estima-

tion methods are available ranging from tightly parameterized to non-parametric ones[32]. Parametric estimators (Ordinary Least Squares, Maximum Likelihood, Bayesian Estimation ...) provide probabilities, marginal effects and many other post-estimation indicators, but require data to follow a given probability distribution, and lay down stringent assumptions on the errors. That is to say, they provide a host of results for interpretations; yet, those results are reliable only if the initial assumptions are not violated. On the other hand, non-parametric methods often require just an association between dependent and independent variables (they do not require stringent assumptions on the distribution of the error, as homoscedasticity or serial correlation), provide robust results, but extrapolation and other post-estimation features are not possible and the conclusions that can be drawn are less interesting. The necessity of having tractable results also in the presence of heteroscedasticity or other assumption violations is at the origin of the development of semi-parametric estimation methods, like GMM (Generalized Method of Moments). Moreover, in the case of model misspecification, a semi-parametric estimation for a parametric model provides robust results, while a parametric estimation does not. Whether to use a parametric, semi-parametric or non-parametric method for estimation depends on a variety of factors, often having an economic impact. For example, when interested in improving a manufacturing process, an engineer may be interested in developing an extremely refined model estimated through a parametric method, in order to be able to obtain a deep understanding of the effects of the independent variables on the outcome. Semi-parametric and non-parametric methods are used more and more by practitioners for the robustness of the conclusions that can be drawn. However, Maximum Likelihood Estimation (MLE) is preferred in the widest range of applications thanks to its asymptotic properties, and for the possibility to calculate appropriate asymptotic covariance matrix even when the density is misspecified. The next subsection introduced MLE providing some examples of its application and the pseudo-MLE, a variation of MLE, is presented.

[32] For further details, see Greene (2003).

G.2.1 MAXIMUM LIKELIHOOD ESTIMATION (MLE)

To introduce the MLE method it is necessary to start from the concept of joint distribution. Given X and Y, continuous random variables, the joint density function $f(x,y)$ can be defined as follows:

$$\Pr(a \le x \le b, c \le y \le d) = \int_a^b \int_c^d f(x,y)dxdy$$

With the conditions:

$$f(x,y) \ge 0$$

$$\int_X \int_Y f(x,y)dxdy = 1$$

Let us now consider the case of a random variable Y and a parameter θ. The probability density function of Y conditional to the values assumed by θ is denoted as $f(y \mid \theta)$. This is the general p.d.f. (probability density function) of the random variable Y as a function of the parameter. The extension to the multi-parameter case is straightforward. Denoting $y_1, \ldots, y_n$ a set of n observations of the random variable Y, the joint probability density, conditional to the set of parameters $\underline{\theta}$ is:

$$f(y_1, \ldots, y_n \mid \underline{\theta}) = \prod_{i=1}^{n} f(y_i \mid \underline{\theta}) = L(\underline{\theta} \mid \underline{y})$$

Where $L(\underline{\theta} \mid \underline{y})$ is called *likelihood function*, and must be interpreted as the joint probability density of the set of parameters $\underline{\theta}$ conditional on the data.

The wide use of MLE in econometrics derives from its appealing asymptotic properties, as anticipated above. If the likelihood function meets three regularity conditions[33], then the estimator has the following asymptotic properties:

- *Consistency*: as the sample size grows, the probability that the true parameter and its estimate differ, tends towards zero. Let $\underline{\theta}_0$ be the true vector of the parameters, and $\hat{\underline{\theta}}$ its estimate: if the method is consistent, then $\hat{\underline{\theta}} \rightarrow \underline{\theta}_0$ when the sample size grows up.
- *Asymptotic normality*: as the sample size grows, the parameters estimated through the Maximum Likelihood method are normally distributed.
- *Asymptotic efficiency*: as the sample size grows, the Maximum Likelihood method provides estimates whose variance is the smallest possible among consistent estimators.
- *Invariance*: ML estimates keep the preceding properties if the parameters undergo one-to-one transformations.

The demonstrations of these properties are far beyond the scopes of this thesis, and are not presented here[34].

These properties are asymptotic, as they hold as the sample size grows up and approaches infinity. The behavior of MLE is unknown with small samples. When using ML for binary and count variables regression models, Long (1997) suggests that it should not be used whit less than 100 observations and a sample of 500 observations seems to be adequate in many situations. Moreover, a minimum of 10 observations per parameter seems to be a necessary condition, while this number grows up if the data are ill-conditioned.

---

[33] See Greene (2003), pag. 474

[34] For a complete discussion on this topic, see (Crowder, 1976) and (Newey and McFadden, 1994).

For computational reasons, it is simpler to work with the log of the likelihood function, as it expression reduces to

$$\ln L(\underline{\theta} \mid \underline{y}) = \ln \prod_{i=1}^{n} f(y_i \mid \underline{\theta}) = \sum_{i=1}^{n} \ln f(y_i \mid \underline{\theta})$$

Note that $f(\underline{y} \mid \underline{\theta})$ and $L(\underline{\theta} \mid \underline{y})$ are exactly the same function, but in $L(\underline{\theta} \mid \underline{y})$ it is emphasized how the parameters are computed using the information contained in the observations. That is to say, when a set of data is collected, and a parametric model is chosen, the likelihood function measures the likelihood of the parameters of the model to generate the observed data. Maximizing $L(\underline{\theta} \mid \underline{y})$, it is therefore possible to obtain the values of the parameters $\underline{\theta}$ that maximize the likelihood of the observed data to be generated by the model.

G.2.1.1 THE GAUSSIAN CASE: UNKNOWN VARIANCE

Let us consider a set of observations which we suppose to be normally distributed with mean 0 and variance σ^2. This way, the variance is the only parameter to be estimated. Let $\underline{y}$ be the vector of observed data. The p.d.f. of a normal distribution with mean 0 and variance σ^2 is:

$$N(0, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{y^2}{\sigma^2}}$$

This way, the log of the likelihood function can be written as follows:

$$\ln L(\underline{\theta} \mid \underline{y}) = \ln L(\sigma^2 \mid \underline{y}) = \sum_{1=1}^{7} \ln \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{y_i^2}{\sigma^2}} \right] = -\frac{1}{2} \sum_{i=1}^{7} \left[\ln \sigma^2 + \ln(2\pi) + \frac{y_i^2}{\sigma^2} \right]$$

When this function is at its maximum, its first derivative is equal to 0, and its second derivative must be negative:

$$\begin{cases} \dfrac{\partial \ln L(\sigma^2 \mid \underline{y})}{\partial \sigma^2} = 0 \\ \dfrac{\partial^2 \ln L(\sigma^2 \mid \underline{y})}{\partial^2 \sigma^2} < 0 \end{cases}$$

In a more general form, the first condition can be written as

$$\frac{\partial \ln L(\underline{\theta} \mid \underline{y})}{\partial \underline{\theta}} = 0$$

which is called Likelihood Equation. The likelihood equation provides the criterion for parameter estimation. In practice, it is usually impossible to analytically write and evaluate the likelihood equation, and numerical methods must be used. A discussion on these methods is beyond the scope of this work; therefore, if interested, the reader should consult specialized texts. Long (1997) provides an overview on the most used numerical methods for MLE with nonlinear models, as logit, probit and models for counts.

In the preceding example, let us suppose the vector of the observed data $\underline{y}$ is $(-3 \quad -3 \quad -1 \quad 0 \quad 2 \quad 3 \quad 5)$. In this case, the likelihood function and its log have their maximum values when σ^2 is approximately equal to 8.1, as Figure G.1 shows.

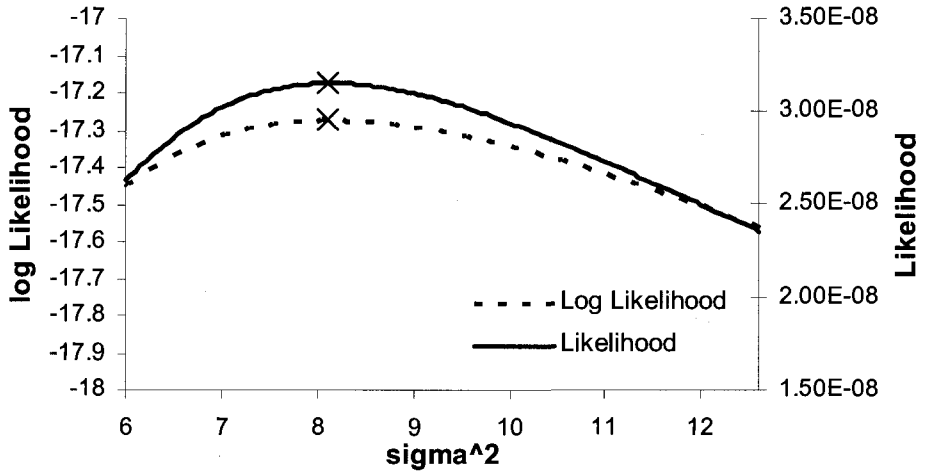

Figure G.1 : Likelihood function and its log, with respect to the parameter σ^2.

This way, the value of the unknown parameter σ^2 that maximizes the likelihood of the data to be generated from a normal process with 0 mean and variance σ^2 can be estimated.

G.2.1.2 THE GAUSSIAN CASE – UNKNOWN MEAN, KNOWN VARIANCE

Let us consider the case of a normally distributed random variable, whose variance is σ^2 and whose mean μ are not known. Let y_i, $i=1...n$ the observed data. The conditional probability density of y_i can be written as:

$$p(y_i \mid \mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i-\mu)^2}{\sigma^2}}$$

therefore

$$\ln p(y_i \mid \mu) = -\frac{1}{2}\ln\sigma^2 - \frac{1}{2}\ln 2\pi - \frac{1}{2}\frac{(y_i-\mu)^2}{\sigma^2}$$

The log-likelihood function takes the form:

$$\frac{\partial \ln L(\mu \mid \underline{y})}{\partial \mu} = \sum_{i=1}^{n}\frac{\partial \ln p(y_i \mid \mu)}{\partial \mu} = \sum_{i=1}^{n}\frac{(y_i-\hat{\mu})}{\sigma^2}$$

The estimated mean is then

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

For $\hat{\mu}$ to be a maximum, it must be that

$$\frac{\partial^2 \ln L(\mu \mid \underline{y})}{\partial \mu^2} < 0$$

In the case considered, we obtain

$$\frac{\partial^2 \ln L(\mu \mid \underline{y})}{\partial \mu^2} = -\frac{n}{\sigma^2} < 0$$

As a result, the maximum likelihood estimator for a sample mean is the geometric mean.

G.2.1.3 MULTI-PARAMETER MAXIMUM LIKELIHOOD ESTIMATION

We are now interested in the case when more than one parameter are to be estimated. The procedure is conceptually the same as above, and the likelihood function can be written as a system of equations, using matrix algebra. Let us first consider the case of a univariate Gaussian distribution, whose mean μ and variance σ^2 are unknown. The vector containing the parameters to be estimated is $\underline{\theta} = \begin{pmatrix} \mu & \sigma^2 \end{pmatrix}$. The p.d.f. of the y_i, $i = 1 \ldots n$ observations, conditional to the parameters is:

$$p(y_i \mid \underline{\theta}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(y_i-\mu)^2}{\sigma^2}}$$

Therefore, the log-likelihood function takes the following form

$$\ln L(\underline{\theta} \mid y_i) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 - \frac{1}{2}\frac{(y_i-\mu)^2}{\sigma^2}$$

The likelihood equation is thus:

$$\frac{\partial \ln L(\underline{\theta} \mid y_i)}{\partial \underline{\theta}} = \nabla_{\underline{\theta}} \ln L(\underline{\theta} \mid y_i) = \begin{bmatrix} \dfrac{(y_i-\mu)}{\sigma^2} \\ -\dfrac{1}{2\sigma^2} + \dfrac{(y_i-\mu)^2}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

From this system of two equations, we simply obtain:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

which is, once again the arithmetic mean of the sample. The estimated variance takes instead the following form

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu})^2$$

which is not the same as the variance estimate of the sample s^2, and it is actually a biased estimator of the variance for small samples, being its expected value

$$E\left[\hat{\sigma}^2\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu})^2\right] = \frac{n-1}{n}\sigma^2 .$$

The multivariate case requires little generalization and more manipulation, but the procedure remains conceptually the same. Let us consider a sample drawn from a multivariate normal population with mean $\underline{\mu}$ and covariance matrix $\underline{\underline{\Sigma}}$. The likelihood equation in this case is:

$$\sum_{i=1}^{n}\underline{\underline{\Sigma}}^{-1}\left(\underline{y}_i - \underline{\hat{\mu}}\right) = \underline{0} .$$

Thus

$$\underline{\hat{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\underline{y}_i .$$

The estimated covariate matrix is:

$$\underline{\underline{\hat{\Sigma}}} = \frac{1}{n}\sum_{i=1}^{n}\left(\underline{y}_i - \underline{\hat{\mu}}\right)\left(\underline{y}_i - \underline{\hat{\mu}}\right)^T .$$

G.2.1.4 ML ESTIMATION FOR LINEAR REGRESSION

The extension to the estimation of the parameters for linear regression is straightforward, considering that the mean of the distribution of the observations to be fitted varies depending on one parameter (the extension to the case of more parameters is then immediate), as $y_i = x_i\beta + \varepsilon_i$. Where ε_i is assumed to be normally distributed with 0 mean and constant variance σ^2. This way, the distribution of the error can be written as:

$$f(\varepsilon_i \mid \alpha + \beta x_i, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{[y_i - \alpha - \beta x_i]^2}{\sigma^2}\right)}$$

The Likelihood function is thus

$$L(\alpha, \beta, \sigma \mid \underline{y}, \underline{\underline{X}}) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2}}$$

and its log is

$$\ln L(\alpha, \beta, \sigma \mid \underline{y}, \underline{\underline{X}}) = \sum_{i=1}^{n} \ln \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2}}$$

The unknown parameters estimates α, β and σ are computed maximizing the log-likelihood function.

In the case of the linear regression model, it can be shown that the Maximum Likelihood estimates of the parameter are he same of Ordinary Least Squares. This is a particular case, as the parameters can be obtained solving a system of equations in a closed form. When estimating parameters for binary or count variables, a solution in a closed form is not available and the estimates must be computed using iterative numerical methods.

G.2.1.5 VARIANCE OF THE ML ESTIMATOR AND PSEUDO-ML

The variance of the MLE is a particularly important topic, leading to a variation of the MLE method called pseudo-MLE, which is robust to some distribution misspecifications.

Let us consider a case in which we are interested in estimating one parameter. The estimate $\hat{\theta}$ of the parameter θ_0 is obtained by maximizing the likelihood function, and $\hat{\theta}$ is a maximum if

$$\left.\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}\right|_{\theta_0} < 0$$

Moreover, if the log-likelihood function changes slowly around θ_0, and so its slope, it will be difficult to determine the correct value of θ_0 that maximizes the function. It can be shown (Davidson and MacKinnon, 1993) that the relationship between the variance of the estimate and the second derivative of the log-likelihood function is

$$\mathrm{var}(\hat{\theta}) = -E\left[\left.\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}\right|_{\theta_0}\right]^{-1}$$

The extension to the case of multi-parameter estimation is straightforward, and requires the use of matrix algebra. For example, let us consider the preceding case of the linear regression model with just one parameter and a constant term. Here α, β and σ must be estimated. Defining the information matrix $\underline{\underline{I}}(\underline{\theta})$ as the negative of the expected value of the Hessian, the covariance matrix of the estimator is the inverse of the information matrix, evaluated in θ_0

$$\operatorname{var}(\hat{\underline{\theta}}) = \begin{pmatrix} -E\left.\frac{\partial^2 \ln L(\underline{\theta})}{\partial \alpha^2}\right|_{\theta_0} & -E\left.\frac{\partial^2 \ln L(\underline{\theta})}{\partial \alpha \partial \beta}\right|_{\theta_0} & -E\left.\frac{\partial^2 \ln L(\underline{\theta})}{\partial \alpha \partial \sigma}\right|_{\theta_0} \\ -E\left.\frac{\partial^2 \ln L(\underline{\theta})}{\partial \beta \partial \alpha}\right|_{\theta_0} & -E\left.\frac{\partial^2 \ln L(\underline{\theta})}{\partial \beta^2}\right|_{\theta_0} & -E\left.\frac{\partial^2 \ln L(\underline{\theta})}{\partial \beta \partial \sigma}\right|_{\theta_0} \\ -E\left.\frac{\partial^2 \ln L(\underline{\theta})}{\partial \sigma \partial \alpha}\right|_{\theta_0} & -E\left.\frac{\partial^2 \ln L(\underline{\theta})}{\partial \sigma \partial \beta}\right|_{\theta_0} & -E\left.\frac{\partial^2 \ln L(\underline{\theta})}{\partial \sigma^2}\right|_{\theta_0} \end{pmatrix}^{-1}$$

This matrix can be directly estimated only in those cases when a close-form solution for θ_0 is available, and when some regularity conditions are met. When using MLE for binary or count outcomes, approximate numerical methods must be used, as the solution has an implicit form.

Different estimates for the covariance matrix can be used: for example, one can compute it through the outer product of the gradient, using the form

$$\operatorname{var}(\hat{\underline{\theta}}) = \left(E\left[\left.\frac{\partial \ln L(\underline{\theta})}{\partial \underline{\theta}} \frac{\partial \ln L(\underline{\theta})'}{\partial \underline{\theta}}\right|_{\theta_0} \right] \right)^{-1}$$

or the simpler BHHH method (Berndt et al., 1974) through evaluate the information matrix:

$$\underline{\underline{I}}(\underline{\theta}_0) = \sum_{i=1}^{n} \left(\frac{\partial \ln L(\underline{\theta} \mid \underline{x}_i)}{\partial \underline{\theta}} \right) \left(\frac{\partial \ln L(\underline{\theta} \mid x_i)}{\partial \underline{\theta}} \right)^T$$

in $\underline{\theta} = \underline{\theta}_0$.

The ML estimator, with some exceptions, is inconsistent if the model is misspecified. Nevertheless, the covariance matrix can be robustly estimated trough the sandwich estimator. The theoretical aspects of this topic are here omitted, and just the results are pre-

sented. If interested, the reader can consult White (1982) and (Gourieroux and Montfort, 1995) for further and more detailed explanations.

The sandwich estimator of the covariance matrix uses the pseudo-true parameter values, estimated with the assumed probability density function. The covariate matrix can thus be estimated as

$$\mathrm{var}\left(\hat{\underline{\theta}}\right) = \underline{\underline{A}}^{-1}\underline{\underline{G}}\underline{\underline{A}}^{-1}$$

where

$$\underline{\underline{A}} = -\lim_{n\to\infty}\frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n}\frac{\partial^2 L(\underline{\theta})}{\partial\underline{\theta}\partial\underline{\theta}'}\bigg|_{\underline{\theta}_0}\right]$$

$$\underline{\underline{G}} = \lim_{n\to\infty}\frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n}\frac{\partial L(\underline{\theta})}{\partial\underline{\theta}}\frac{\partial L(\underline{\theta})}{\partial\underline{\theta}'}\bigg|_{\underline{\theta}_0}\right]$$

These two matrices can be estimated in different ways using a variety of numerical methods.

G.3 MODELS FOR BINARY OUTCOMES

A binary outcome is very common in social sciences, and is a variable that can only take two values, typically 0 and 1. For example, the value 1 is assigned when a certain event takes place and 0 otherwise. For this reason, a binary variable in social sciences is often called a *discrete choice variable*. In this work, for example, the dependent binary variable `ec` takes the value 1 if the biotechnology firm is involved in a partnership, 0 otherwise.

Although the results produced by a parametric model could seem fragile, due to the full a-priori specification, they continue to be broadly used by practitioners for their simplic-

ity and interpretability. Moreover, the literature on non parametric models is not developed enough and is almost only theoretical (Matzkin, 1993). A compromise between a fully specified model and a nonparametric one is the semiparametric approach. For models for binary outcomes, for example, the density of the error is not assumed, but is estimated through the kernel density method (Klein and Spady, 1993).

The use a linear regression model for discrete choice variables is in general inappropriate. In the following, we present two very similar models for fitting binary outcomes: the binary logit model and the binary probit model. Two different approaches are followed to present the models: the non-linear probability density and the latent variable approach. Both techniques lead to the same results, starting from conceptually different assumptions.
Before introducing the logit and probit models for binary variables, we want now to justify the use of those models for fitting a binary outcome showing the shortcomings of using a linear regression model.

Let us consider a binary variable y that can assume the values 0 and 1[35] and depends on a number of parameters $\underline{\beta}$. It can be easily shown that the conditional expected value of the variable is:

$$\mathrm{E}(y_i \mid \underline{x}_i) = \Pr(y_i = 1 \mid \underline{x}_i)$$

In general, we are looking for a relationship between the conditional expected value of the binary variable and the set of parameters:

$$\Pr(y_i = 1 \mid \underline{x}_i) = F(\underline{x}, \underline{\beta})$$

[35] For computational reasons, when the variable is not available in this format, it is necessary to re-codify it this way.

A possibility is to use the linear regression model, setting

$$F(\underline{x}, \underline{\beta}) = \underline{x}_i' \underline{\beta}$$

As it has been anticipated, the classical linear regression model used for fitting a binary dependent variable, also called Linear Probability Model (LPM) has a number of shortcomings[36]:

- The errors are heteroscedastic and depend on $\underline{\beta}$: it can be easily proved that $\text{var}(\underline{\varepsilon} \mid \underline{x}) = \underline{x}'\underline{\beta}(1 - \underline{x}'\underline{\beta})$
- The errors are not normally distributed: this can be easily shown considering tat for a given $\underline{x}^*$, as $y_{\underline{x}^*}$ can either be 0 or 1, the error can either be $0 - E(y \mid \underline{x}^*)$ or $1 - E(y \mid \underline{x}^*)$.
- The LPM generates nonsensical predictions: the linear function is not limited between 0 and 1, and probabilities greater than 1 or negative can be predicted.
- The functional form of the LPM is not realistic when used for fitting a binary dependent variable. According to Long (1997) this is the major problem with LPM, which implies constant marginal effects[37]. In reality, as the value of a factor increases, it is evident that its effect on the dependent variable decreases.

Although some authors as Fomby et al. (1984) have proposed some modifications to improve LPM, most of the literature recommends to adopt a different function between the binary dependent variable and the factors. The usual choice is the cumulate of a probability distribution. In this respect, many distributions can be employed (Maddala, 1983); still the most commonly employed are the normal and the logistic distribution

[36] It is beyond the scopes of this work to provide a complete discussion on LPM. For further details it is recommended to consult Aldrich and Nelson (1984).
[37] The concept of marginal effect will be discussed below.

that generate the probit and the logit model respectively. Therefore, the probit model can be written as:

$$\Pr(y=1 \mid \underline{x}) = \int_{-\infty}^{\underline{x}'\underline{\beta}} \phi(s)ds = \Phi(\underline{x}'\underline{\beta})$$

Where $\phi(s)$ denotes the normal distribution. The logit model uses the logistic distribution instead of the Gaussian distribution:

$$\Pr(y=1 \mid \underline{x}) = \frac{e^{\underline{x}'\underline{\beta}}}{1+e^{\underline{x}'\underline{\beta}}} = \Lambda(\underline{x}'\underline{\beta})$$

The broad diffusion of the logit model is justified by the computational advantages of using the logistic distribution, especially in post-estimation analysis. The logistic distribution differs slightly from the Gaussian, and is close to a t-distribution with 7 degrees of freedom (Greene, 2003). In practice, whether to use one model or the other, it is not clear, and a definitive criterion based on theoretical ground is not available.

G.3.1 THE LATENT VARIABLE APPROACH

The adoption of the logit and probit models can be derived through the introduction of a latent, not observed, variable y^*, which leads to exactly the same results. Following the latent variable approach the observed binary variable y takes the value 0 when y^* is smaller than a certain threshold τ [38]; conversely it is 1 when y^* is greater than τ. Thus, the latent variable can be considered as the propensity of taking or not a certain decision. It could be, for instance, the propensity of buying a luxury car, which depends on a variety of factors. The relationship between that links the factors of the model and the latent variable is linear:

[38] In general it is assumed $\tau = 0$

$$y_i^* = \underline{x}_i' \underline{\beta} + \varepsilon_i$$

The distribution we assume[39] for ε_i (logistic or Gaussian) determines the model adopted (logit or probit, respectively).

G.3.2 MAXIMUM LIKELIHOOD ESTIMATION

Contrarily to the case of the linear regression model, maximizing the likelihood function of a logit (or probit) model requires approximate numerical methods. For a binary choice model, as the dependent variable can assume only the values 0 or 1, the joint probability function can be written as:

$$\Pr(y_1, \ldots y_n \mid \underline{\theta}) = \prod_{y_i=0} (1 - F(\underline{x}_i' \underline{\theta})) \prod_{y_i=1} F(\underline{x}_i' \underline{\theta})$$

which can be conveniently rewritten as:

$$\Pr(y_1, \ldots y_n \mid \underline{\theta}) = \prod_{i=1}^{n} [F(\underline{x}_i' \underline{\theta})]^{y_i} [1 - F(\underline{x}_i' \underline{\theta})]^{1-y_i}$$

where, as usual, F is the cumulative probability function of the logistic distribution in the case of a logit model and of a normal distribution if a probit model is adopted. In particular, in the case of a logit model we have:

$$F(\underline{x}_i' \underline{\theta}) = \Pr(\underline{y} = 1 \mid \underline{x}) = \frac{e^{\underline{x}'\underline{\theta}}}{1 + e^{\underline{x}'\underline{\theta}}} = \Lambda(\underline{x}' \underline{\theta})$$

[39] The latent variable is a conceptual model, and it is not observed. Therefore its variance cannot be estimated, but only assumed.

and

$$\lambda(\underline{x}'\underline{\theta}) = \frac{d\Lambda(\underline{x}'\underline{\theta})}{d(\underline{x}'\underline{\theta})} = \frac{e^{\underline{x}'\underline{\theta}}}{\left(1+e^{\underline{x}'\underline{\theta}}\right)^2} = \Lambda(\underline{x}'\underline{\theta})[1-\Lambda(\underline{x}'\underline{\theta})].$$

After taking the log of the joint probability function, the likelihood equation, after some manipulation, is:

$$\frac{\partial \ln L}{\partial \underline{\theta}} = \sum_{i=1}^{n}\left[\frac{y_i \cdot \lambda_i}{\Lambda_i} + (1-y_i)\frac{-\lambda_i}{1-\Lambda_i}\right]\cdot \underline{x}_i = \underline{0}$$

which can be written as:

$$\frac{\partial \ln L}{\partial \underline{\theta}} = \sum_{i=1}^{n}(y_i - \Lambda_i)\underline{x} = \underline{0}$$

This system of equations is non linear and, as it has been said, it is necessary to employ approximate numerical methods to estimate the vector of the parameters $\underline{\theta}$. We also observe that the constraint imposed by the MLE applied to the logit model requires the average predicted probabilities to be equal to the average observed probabilities of the sample.

A number of different numerical methods can be used to estimate the parameters. Each method has different properties regarding stability, convergence and regularity requirements. It is beyond our scopes to investigate this topic. The reader can consult specialized texts and publication, as Judge et al. (1985), Cramer (1986) or Dhrymes (1984).

G.3.3 COMPARISON BETWEEN LOGIT AND PROBIT MODELS

It is natural to ask which model should be used for fitting a binary dependent variable; however, when the sample size is not extremely large, no definitive proof on whether to use a probit or a logit model under given conditions can be found in the literature. The choice is often taken considering the different computational effort required by the use of the logit model. This model is also preferred when the practitioner is interested in using the odd ratios method[40] for interpretation.

Although the statistical significance of a parameter estimated through a logit or a probit distribution does not vary significantly, the value of the parameter itself is different, as the structural form of the two density functions differs the way it has been described above. For example, let us distinguish between the probit and the logit models applied to the same sample of data. The latent variable is described as

$$y_L^* = \underline{x}\underline{\beta}_L + \varepsilon_L$$

if the logit model is used, and as

$$y_P^* = \underline{x}\underline{\beta}_P + \varepsilon_P$$

if the probit model is used. What is the relationship between $\underline{\beta}_L$ and $\underline{\beta}_P$? The solution to this question is not univocal, and several criteria can be used, leading to slightly different findings. Long (1997) proposes to start considering the assumed variance of the error of the latent variable, which is $\text{var}(\varepsilon_P \mid \underline{x}) = 1$ in the probit model and $\text{var}(\varepsilon_L \mid \underline{x}) = \pi^2/3$ in the logit model. This way it an be easily showed that

[40] This method will be presented below

$$\underline{\beta}_L \approx \sqrt{\mathrm{var}(\varepsilon_L \mid \underline{x})}\underline{\beta}_P \quad \text{or} \quad \underline{\beta}_L \approx \delta \cdot \underline{\beta}_P$$

where $\delta \approx 1.81$. Another criterion is proposed by Amemiya (1981), setting δ in order to minimize the distance between the logistic and the normal distribution. This way $\delta \approx 1.6$.

G.3.4 INTERPRETATION

This section aims at giving the necessary tools for interpreting the results of a binary response model, being it a probit or a logit model. The linear probability model, due to its limitations and shortcomings, is not considered here. First, the effects that the parameters have on the probability curve are examined: we will see how the intercept and the slope for both logit and probit models shape the predicted. Then we will focus on the most common methods of interpretation, starting from the odds ratio method that can be used with the logit model only, to the use of the predicted probabilities, the marginal effect and the discrete change in the outcome. The main properties and shortcomings of each method are briefly presented too.

G.3.4.1 EFFECT OF THE PARAMETERS ON THE CURVE

We start examining the effect of the parameters on the curve that describes the probability of a certain event to occur, or $\Pr(y = 1 \mid \underline{x})$. Here, only the case of a single parameter and the intercept is considered; the extension to the multiparameter case is then straightforward. In general, we find that

$$\Pr(y = 1 \mid x) = F(\alpha + \beta \cdot x)$$

where α is the intercept. The effect of the intercept on the curve is a shift to the left as its value gets larger, and to the right as it gets smaller, as shows. In particular, for $\alpha = 0$ the curve passes through the point$(0;0.5)$: when the value of the parameter x equals 0,

the probabilities of y to be 0 equals the probability to be 1: the events have the same probability to occur.

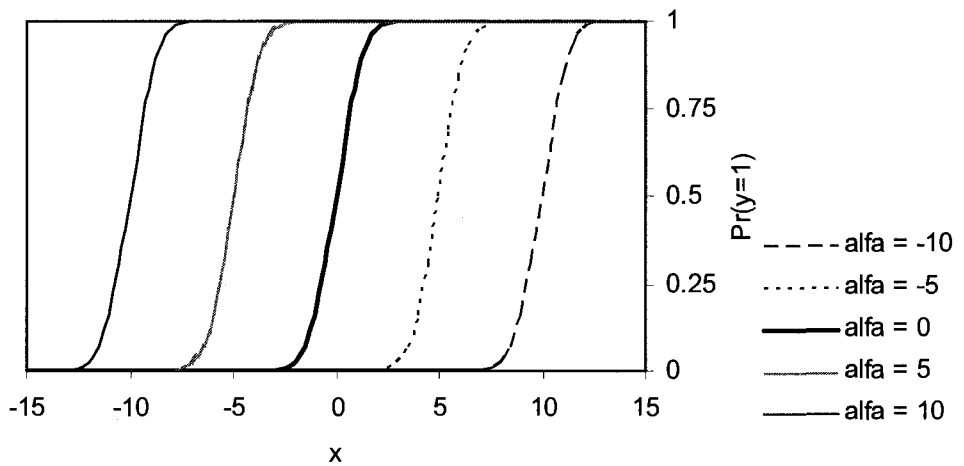

Figure G.2 : Effect of the intercept on the BRM curve.

Figure G.3 shows the effect of the parameter β on the same curve, when $\alpha = 0$: an increase in β makes the curve steeper around the point $(0;0.5)$, and the tails flat. Conversely, a decrease in β stretches the curve:

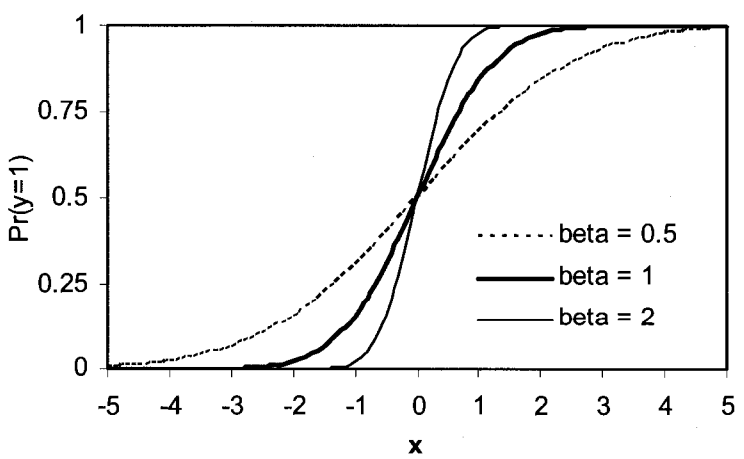

Figure G.3 : Effect of the slope on the BRM curve

G.3.4.2 THE ODDS RATIOS METHOD

This method consists in a set of considerations on the coefficients of the logit model that can be interpreted as a change in the logit for a unit change in the independent variable. To show this fact, we first define the quantity:

$$\Phi(\underline{x}) = \frac{\Pr(y=1 \mid \underline{x})}{1-\Pr(y=1 \mid \underline{x})}$$

If we introduce the specific functional form of the logistic distribution and take the log of $\Phi(\underline{x})$ we obtain, after some manipulation:

$$\ln \Phi(\underline{x}) = \underline{x}'\underline{\theta}$$

Therefore, the partial derivative of this quantity calculated with respect to the independent variable x_i is simply the parameter θ_i:

$$\frac{\partial \ln \Phi(\underline{x})}{\partial x_i} = \theta_i$$

which does not depend on the values of the other independent variables at which it is computed. The meaning of a change in the logit remains to be explained: considering the expression of $\Phi(\underline{x}) = \exp(\underline{x}'\underline{\theta})$ we can calculate this quantity for a change of δ in the independent variable x_i, keeping all the others at a certain value. We thus obtain:

$$\Phi(\underline{x}'; x_i + \delta) = \exp(\theta_0)\exp(x_1\theta_1)\cdots\exp(x_i\theta_i)\exp(\delta \cdot \theta_i)\cdots\exp(x_n\theta_n)$$

It is therefore immediate to obtain the quantity:

$$\frac{\Phi(\underline{x}; x_k + \delta)}{\Phi(\underline{x}; x_k)} = \exp(\delta \cdot \theta_k)$$

which is called the odds ratio and can be interpreted as the change in the odds for a change of δ in the independent variable. In the case of $\delta = 1$ the odds ration take the simpler expression $\exp(\theta_i)$, which is called factor change. It is worth noting that a negative effect of a variable on the outcome results in a positive odds ratio which is smaller than 1, while a positive effect is represented by an odds ratio which is greater than 1. Often the percentage change in the odds is used for interpretation. This quantity is defined as:

$$\frac{\Phi(\underline{x}; x_i + \delta) - \Phi(\underline{x}; x_i)}{\Phi(\underline{x}; x_i)} \times 100 = [\exp(\delta \cdot \theta_i) - 1] \times 100$$

G.3.4.3 PREDICTED PROBABILITIES

Interpretation using the predicted probabilities is usually the first, most simple and direct method to analyze the effect of a factor on the dependent variable. Nevertheless, when more than two factors affect the outcome, this method is no longer easy to employ, as the response cannot be visualized. The predicted probability of a given event, i.e. $y_i = 1$ is defined, for the logit model, as:

$$\Pr(y_i = 1 \mid \underline{x}_i) = \Lambda(\underline{x}_i' \underline{\hat{\theta}})$$

This way, for a given $\underline{\bar{x}}_i$ it is possible to compute the predicted probability according to the model, be it a probit or a logit. This kind of interpretation is useful when the practitioner is interested in setting the independent variables to fix the outcome at a certain desired value. In symbols, the problem is finding the value of $\underline{\bar{x}}$ for a given proportion of

the sample. In this work, we will not use this method, as we are interested in understanding the effect of the independent variables on the outcome, and not in setting the parameters to get a specific proportion of the outcome. Rather, it can be interesting to analyze the effect of each variable on the predicted probabilities. This way the practitioner can gat an understanding of the extent to which a change in a certain variable affects the predicted probability. One way to do this is to set all the independent variables but one, say x_i, at their means, and to compute the predicted probability when x_i is at his minimum and at his maximum. Thus, the predicted change in the probability for a change in x_i can be obtained as:

$$\Pr(y=1 \mid \bar{x}, \max x_i) - \Pr(y=1 \mid \bar{x}, \min x_i)$$

This way, one can have an immediate evidence of which are the most significant variables.

G.3.4.4 MARGINAL EFFECT

Interpretation using marginal effect is quite usual among econometricians, but one must be careful in employing it, as it has a number of shortcomings of which the practitioner must be aware. The marginal effect of a variable on the outcome is defined as the partial derivative of $\Pr(y=1 \mid \underline{x})$ with respect to a certain independent variable x_i, and in the case of the logit model can be expressed as:

$$\frac{\partial \Pr(y=1 \mid \underline{x})}{\partial x_i} = \frac{\partial \Lambda(\underline{x}'\underline{\theta})}{\partial x_i} = \lambda(\underline{x}'\underline{\theta}) \cdot \theta_i = \Pr(y=1 \mid \underline{x})[1 - \Pr(y=1 \mid \underline{x})]\theta_k$$

This means that the partial change of the probability for a given parameter depends on the values at which the other parameters are set. In addition, one could be interested in computing the relative magnitudes of the marginal effect of two independent variables in

the outcome. In this case, the part $\Pr(y=1\,|\,\underline{x})[1-\Pr(y=1\,|\,\underline{x})]$ simplifies to the expression:

$$\frac{\dfrac{\partial\Lambda(\underline{x}'\underline{\theta})}{\partial x_i}}{\dfrac{\partial\Lambda(\underline{x}'\underline{\theta})}{\partial x_j}}=\frac{\theta_i}{\theta_j}$$

This quantity does no longer depend on the specific value of $\underline{x}$ at which it is evaluated; it is therefore a useful tool to assess the relative effect of two variables on the binary outcome.

When using the simple marginal change, a commonly adopted choice is to set the values of the parameters at their means, and to compute the marginal change as a function of x_i. It is also possible to set the variables at the average of all observations. However, the question of at which value the other parameters should be set becomes a major issue when one or more independent variables are binary. In this case, it is inappropriate to set them at their means, and a binary variable can only assume the values 0 or 1. In this case, Long (1997) recommends using the discrete change method, presented below.

G.3.4.5 METHOD OF THE DISCRETE CHANGE

This method represents the discrete version of the marginal effect, being the partial change in the probability calculated with respect to the variable x_i defined as:

$$\frac{\Delta\Pr(y=1\,|\,\underline{x})}{\Delta x_i}=\Pr(y=1\,|\,\underline{x};x_i+\delta)-\Pr(y=1\,|\,\underline{x};x_i)$$

Some authors like Kaufman (1996) suggest computing the centered discrete change, to assess the magnitude of the probability change for a centered variation of δ around a certain value:

$$\frac{\Delta \Pr(y=1 \mid \underline{x})}{\Delta x_i} = \Pr\left(y=1 \mid \underline{x}; x_i + \frac{1}{2}\delta\right) - \Pr\left(y=1 \mid \underline{x}; x_i - \frac{1}{2}\delta\right)$$

Clearly, the discrete change becomes the marginal change as Δx_i tends toward 0. The discrete change is therefore interpreted as the change in the probability for an increment of δ in a certain independent variable. While it is not worth discussing about the geometrical properties of a discrete increment compared to the marginal effect, it is interesting to note how, the more the probability curve is linear, the more the two measures will be close to each other. In the presence of one or more binary independent variables, the advantage of using the discrete change is easy to assess, as the probabilities can be computed when the binary independent variable is at its minimum and when it is at its maximum; the difference between these two quantities is the discrete change:

$$\frac{\Delta \Pr(y=1 \mid \underline{x})}{\Delta x_i} = \Pr(y=1 \mid \underline{x}; x_i = 1) - \Pr(y=1 \mid \underline{x}; x_i = 0)$$

Some authors suggest comparing the discrete change in the probability with respect to the estimated standard deviation of the specific factor that is considered. The conceptual meaning of this procedure reminds the one employed in the ANOVA, where an F statistic is computed to evaluate the effect of one or more factors in the outcome (Montgomery, 1997). This way the discrete change is:

$$\frac{\Delta \Pr(y=1 \mid \underline{x})}{\Delta x_i} = \Pr\left(y=1 \mid \underline{x}; x_i + \frac{1}{2}s_i\right) - \Pr\left(y=1 \mid \underline{x}; x_i - \frac{1}{2}s_i\right)$$

G.3.5 HYPOTHESIS TESTING

To check whether a parameter is statistically significant or not, it is necessary to test its value with the null hypothesis that it is equal to 0. For single parameter testing, a simple z-test can be used, as MLE tends to be normally distributed as the sample grows in size. However, many softwares employ a t-test or a quasi-t-test (Cramer, 1986; Godfrey, 1988). The hypotheses to be tested are:

$$\begin{cases} H_0 : \theta_i = \theta_i^* \\ H_1 : \theta_i \neq \theta_i^* \end{cases}$$

where usually $\theta_i^* = 0$. If a z-test is employed, the statistic is:

$$z = \frac{\hat{\theta}_i - \theta_i^*}{\hat{\sigma}_{\hat{\theta}_i}}$$

where $\hat{\sigma}_{\hat{\theta}_i}$ is the estimated variance of the parameter θ_i. When the practitioner is interested in testing if several coefficients are simultaneously equal to 0 (or more complex hypotheses), then the z-test can no longer be employed, and other techniques are necessary. The most common methods to test complex hypotheses are the Wald test, the Likelihood Ratio (LR) test and the method of the Lagrange Multipliers (LM). In the following, we discuss the main issues regarding these three methods. A complete discussion on these tests is provided by many authors. Godfrey (1988) in particular, presents a detailed discussion on this topic. What the three procedures have in common is that they test the constrained model (where the constrain is introduced by the test itself) with respect to the non-constrained estimated model. The way the comparison is made differentiates one method from the others. A linear constraint for a multiparameter model can be written under the form of a vector. For example, in a model with four parameters, one may be interested in testing the following

$$\begin{cases} \theta_1 = 0 \\ \theta_3 = \theta_4 \end{cases}$$

In symbols, the null hypothesis is $H_0 : \underline{\underline{C}}\underline{\theta} = \underline{r}$ or:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

In this case, the constraint is linear, but the Wald test can be also used to test non-linear restrictions. The null hypothesis will thus be $H_0 : c(\underline{\theta}) = \underline{r}$. The Wald statistics is based on the full rank quadratic form

$$q = (\underline{x} - \underline{\mu})\Sigma^{-1}(\underline{x} - \underline{\mu})$$

which is distributed as a chi-square with degrees of freedom equal to the number of observations. The Wald statistic for a non-linear constraint is

$$W = [c(\underline{\theta}) - \underline{r}]\left(\underset{Asy.}{Var}\, c(\underline{\theta}) - \underline{r}\right)^{-1}[c(\underline{\theta}) - \underline{r}]$$

which is asymptotically distributed as a chi-square with degrees of freedom equal to the number of constraints. The meaning of the Wald test can be examined considering the terms constituting the statistics. Here the distance between the restricted and the unrestricted models (the external terms in the statistics) are weighted by the curvature of the likelihood function. This means that if the curvature of the likelihood function changes

rapidly, a small distance between the restricted and the unrestricted model will be enough to reject the null hypothesis, and vice versa. To see the equivalence of the z-test and the Wald test when the null hypothesis is $H_0 : \theta_i = \theta_i^*$, let us consider the Wald statistics for this simple situation:

$$W = \frac{\left(\hat{\theta}_i - \theta_i^*\right)^2}{\hat{\sigma}^2_{\hat{\theta}_i}}$$

that is the square of the z-statistic, and is distributed as a chi-square with 1 degree of freedom, which is the square of the normal distribution.

A second way to test complex hypothesis is the LR test. Let the vector of the estimated parameters of the model without restriction be denoted as $\underline{\hat{\theta}}_m$, and the parameters of the restricted model as $\underline{\hat{\theta}}_r$. As the likelihood function L_m of the unrestricted model reaches its maximum for $\hat{\theta} = \underline{\hat{\theta}}_m$, the function

$$\lambda = \frac{L_r}{L_m}$$

must be between 0 and 1, where L_r is the likelihood function of the restricted model. Intuitively, the closer λ is to zero, the more likely is the null hypothesis to be rejected, and vice versa. It can be shown that for a large sample, the asymptotic distribution of $-2 \cdot \ln \lambda$ is a chi-square with degrees of freedom equal to the number of restrictions imposed:

$$-2 \cdot \ln \lambda \sim \chi^2_{num.rest}$$

The main shortcoming of this method is that it requires the estimation of both the restricted and the unrestricted model, which may make the computation too onerous. By contrast, the Wald test requires the estimate of one model only, but it could encounter numerical instability if the software employed rounds the results. Moreover, the LR test cannot be used to test simple hypothesis on one parameter, as Greene (2003) shows.

The last method for complex hypotheses testing we consider is the LM procedure. It is based on the consideration that if the imposed restrictions are true, then the difference between the maximized values of the restricted and unrestricted likelihood functions will be very small. Let us indicate with $\underline{\lambda}$ a vector of Lagrange Multipliers, representing the effect of the restrictions on the roots of the likelihood function. The log-likelihood function is thus:

$$\ln L_r = \ln L_m(\underline{\theta}) + \underline{\lambda}'(c(\underline{\theta}) - \underline{r})$$

where the last term is likely to be very small if the restrictions are true. The roots $\underline{\lambda}$ and $\underline{\theta}$ of this system must be computed and, as anticipated above, the Lagrange multipliers can be directly tested as $H_0 : \underline{\lambda} = \underline{0}$. Under this null hypothesis, it can be shown that for large samples the statistic

$$LM = \left(\frac{\partial \ln L(\underline{\hat{\theta}}_r)}{\partial \underline{\hat{\theta}}_r} \right)' \left[I(\underline{\hat{\theta}}_r) \right]^{-1} \left(\frac{\partial \ln L(\underline{\hat{\theta}}_r)}{\partial \underline{\hat{\theta}}_r} \right)$$

is distributed as a chi-square with degrees of freedom equal to the number of restrictions imposed. $I(\underline{\hat{\theta}}_r)$ is the information matrix defined above. The LM test requires the estimate of just the restricted model.

Of the three procedures examined, the most commonly used is the Wald test, as it requires the estimation of one model, it takes into account the effect of the estimated variance of the parameters and it can be used also for simple hypothesis testing, as in this case it is equivalent to the usual z-test. Although some authors (Hauck and Donner, 1977) show that in some situations the Wald test has an aberrant behavior, it is not clear, from a conceptual perspective, which test is superior (Rothenberg, 1984). It is worth noting, however, that the calculations required by the Wald test need the coefficients to be expressed at full precision, as the effect of rounding could lead to numerical instability and to misleading results.

G.3.6 TEST OF FIT

After the parameters of the model have been estimated and tested, it is useful to assess the overall model fit, defined as the likelihood of the model to represent the real data. A large number of qualitative and quantitative methods can be employed; in the following, we will present some of them. If interested, the reader should consult specialized manuals. In this respect Fox (1991), Weisberg (1980), Pregibon (1981) and Long (1997) provide useful introduction and overview to these methods, also for limited dependent variables.

This section is not intended to give the theoretical basis of the residual analysis, and some knowledge of these principles in the case of the linear regression model is necessary to understand the following. We will only present the peculiar aspects, or particularly relevant, when a binary response model is used: we will not introduce, for example, the residual plot or the methods to individuate the outliers.

G.3.6.1 THE PEARSON RESIDUALS

In the case of a binary response model, the conditional mean of the probability is:

$$p_i = E(y_i = 1 \mid \underline{x}_i)$$

and, as the dependent variable can only assume the values 0 and 1, its variance is:

$$Var(y_i \mid \underline{x}_i) = p_i(1 - p_i)$$

It is thus straightforward to define the Pearson residual as:

$$r_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}}$$

Intuitively, the larger the residuals, the less likely the model to fit the observed data. For a quantitative measure of the difference between prediction and observation using the Pearson residual, a statistic can de defined as:

$$X^2 = \sum_{i=1}^{N} r_i$$

Conceptually, it is preferred to employ the standardized Pearson residual r_i^S (Hosmer and Lemeshow, 1989; Long, 1997), defined as:

$$r_i^S = r_i / \sqrt{1 - h_{ii}}$$

where h_{ii} is used to estimate the variance of the Pearson residual, and is:

$$h_{ii} = \hat{p}_i(1 - \hat{p}_i) \cdot \underline{x}_i \cdot Var(\underline{\hat{\theta}}) \cdot \underline{x}_i'$$

The use of this measure alone to assess the fit of a certain model could be misleading (McCullagh and Nelder, 1989). Actually, it is not even clear how X^2 is distributed:

some authors say that it is distributed as a chi-square, some others as a Gaussian (McCullagh, 1986).

G.3.6.2 PSEUDO-R^2 MEASURES

Scalar measures of fit are a tool to investigate the likelihood of the model to accurately describe the data generating mechanism. Nevertheless, the practitioner should not be blind in judging the quality of a model by considering the values of those measures. In fact, "there is no convincing evidence that selecting a model that maximizes the value of a given measure results in a model that is optimal in any sense" (Long and Freese, 2006). In the following, a number of scalar measures of fit for binary dependent variable models are presented. Each one of these coefficients represents a variation of the coefficient of correlation R^2.

The well known statistic R^2 in the case of the linear regression model provides information on the proportion of the variation in y that can be explained by the model's regressors. It is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$

Where in general $\sum_{i=1}^{N} (y_i - \hat{y}_i)^2$ is referred as RSS, or sum of the squared residuals and $\sum_{i=1}^{N} (y_i - \bar{y})^2$ as TSS, or total sum of squares. The adjusted R^2 takes into account the number of the independent variables K:

$$R^2_{adj} = \left(R^2 - \frac{K}{N-1} \right) \left(\frac{N-1}{N-K-1} \right)$$

R^2 can be derived in a variety of ways (Goldberger, 1991; Pindyck and Rubinfeld, 1998): using the estimated variance, the likelihood function or the F-test for example. Each one can be modified to be employed for a binary response model. The simplest extension of R^2 for a BRM is obtained by simply substituting the observed $\hat{y}_i$ with the estimated conditional probability p_i. This procedure was first introduced by Efron (1978) and the so-called Efron statistic is thus:

$$R^2_{Efron} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{p}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}.$$

Considering R^2 as obtained from the ratio of the estimated variances, McKelvey and Zavoina (1975) proposed an extension to the models for count variables, successively adapted by Laitila (1993) to the case of a binary response model:

$$R^2_{MZ} = \frac{Var(\hat{y}^*)}{Var(\hat{y}^*) + Var(\varepsilon)}$$

Obviously, this expression poses an important issue. It has been said that the latent variable is not directly observed, and its variance is assumed, as it cannot be estimated. None the less, the authors propose another way to compute the variance of the latent variable:

$$Var(\hat{y}^*) = \underline{\hat{\theta}}' Var(\underline{x}) \underline{\hat{\theta}}$$

If we consider the derivation of R^2 as the ratio between the likelihood function of the model with just the intercept and the likelihood function of the full model, we obtain, for a BRM, known as the Cox-Snell R^2:

$$R^2_{ML} = 1 - \left[\frac{L(M_{\text{int}})}{L(M_{full})} \right]^{2/N}$$

that can also be expressed as (Maddala, 1983):

$$R^2_{ML} = 1 - \exp\left(-G^2/N\right)$$

where G^2 is the likelihood ratio:

$$G^2 = -2 \cdot \ln\left[\frac{L(M_{\text{int}})}{L(M_{full})} \right]$$

Cragg and Uhler (1970) build on Maddala (1983) and show that R^2_{ML} has a specific maximum, whose expression is:

$$\max\left(R^2_{ML}\right) = 1 - L(M_{\text{int}})^{2/N}$$

Therefore, they propose to norm R^2_{ML} with respect to its maximum, obtaining:

$$R^2_{CU} = \frac{R^2_{ML}}{1 - L(M_{\text{int}})^{2/N}}$$

The last scalar measure of fit we consider is maybe the most popular among econometricians and has first been introduced by McFadden (1973). This measure too is based on the ratio between the likelihood function of the model with just the intercept and the full model, but the definition is different than R^2_{ML}:

$$R^2_{MF} = 1 - \frac{\ln L(M_{full})}{\ln L(M_{\text{int}})}$$

Unfortunately, as R^2, also R^2_{MF} increases as the number of parameters grows; it is therefore necessary to define the R^2_{MF} adjusted, to overcome this shortcoming. Thus, the measure of fit becomes:

$$R^2_{MF,adj} = 1 - \frac{\ln L(M_{full}) - K}{\ln L(M_{\text{int}})}$$

A number of other scalar measures of fit exist[41] that we will not consider here. It is worth reminding that a scalar measure in itself is not enough to state with certitude that a model is better than another is; however, it can give a general idea of the behavior of the model. A useful framework to follow is to compute different R^2's to compare the models using several approaches. This will provide the practitioner with a deeper understanding of the overall behavior of the models and will give him a better tool to choose. Unfortunately, when sampling weights are employed, the majority of these coefficients cannot be computed. In this work, we only use the McKelvey and Zavoina measure of fit.

G.4 MODELS FOR COUNTS

A count variable is a non-negative integer that measures the number of times an event occurs over a given period of time. Processes involving this type of variables require specific models to be employed. Cameron and Trivedi's (1998) monograph provide a complete and detailed discussion on this type of models, and can be assumed as a reference for the practitioner interested in analyzing count data.

[41] For a review, see Veall and Zimmermann (1996).

An early example of the application of models for counts is provided by Bortkiewicz (1898), who investigated the factors affecting the number of soldiers in the Prussian army dead annually by being kicked by mules. Since their introduction, models for counts have been developed in order to become flexible enough to take into account specific data structures, as zero-inflation, truncation, overdispersion etc... The main reason for not using a classical linear regression for count is that often in reality the hypothesis of identically and independently distributed errors is too strong, and the variance of the conditional mean is not constant over the range of variation of the independent variables.

This section is structured in a different way than the preceding, as many of the concepts introduced before can be easily applied to the case of the models for count variables. Each subsection deals with the estimation, interpretation and test of fit of a specific model, in order to highlight its peculiar characteristics.

In the following, two models for counts are presented and examined: the Poisson Regression Model (PRM) and the Negative Binomial Regression (NBR) with its main variations. All these models, as in the case of BRM, are usually estimated by maximizing a log-Likelihood function (in particular, a pseudo-MLE is employed to estimate the covariance matrix); therefore, what has been presented in the preceding section will remain valid, and will not be repeated here. First, we will consider the basic count regression, represented by the Poisson Regression. Although its use is not widespread, it is useful to examine it in order to introduce the main issues a count regression implies. We will then focus on the NBR and on the interpretation of the parameters. Finally, testing hypothesis and test of fit are presented.

G.4.1 POISSON REGRESSION

The assumption of the Poisson Regression is that the conditional probability density of the outcome is independently Poisson distributed:

$$f(y_i \mid \underline{x}_i) = \frac{e^{-\mu_i} \cdot \mu_i^{y_i}}{y_i!}$$

Clearly, this condition is not enough to fit the model, as there is no relationship between μ_i and the vector $\underline{x}$. One could suggest many different relationships, but the most commonly used is an exponential:

$$\mu_i = \exp(\underline{x}_i' \underline{\theta})$$

where $\underline{\theta}$ is the vector of the parameters to be estimated. This expression can also be written in the equivalent form of a log-linear:

$$\ln(\mu_i) = \underline{x}' \underline{\theta}$$

As the outcome is Poisson distributed, the conditional mean and the conditional variance have the same value, and are equal to:

$$E[y_i \mid \underline{x}_i] = V[y_i \mid \underline{x}_i] = \exp(\underline{x}_i' \underline{\theta}).$$

Clearly, the model is intrinsically heteroscedastic, as the variance grows exponentially with the independent variables.

As anticipated, the Poisson Regression Model is usually estimated maximizing the log-likelihood function:

$$\ln L(\underline{\theta} \mid \underline{y}, \underline{x}) = \sum_{i=1}^{N} \ln[\Pr(y_i \mid \mu_i)] = \sum_{i=1}^{N} [y_i \underline{x}_i' \underline{\theta} - \exp(\underline{x}_i' \underline{\theta}) - \ln(y_i!)]$$

As in the case of the logit and probit models, an analytical solution to maximize this function is not available, and approximate numerical methods must be used.

G.4.1.1 INTERPRETATION

Once the parameters are estimated, their interpretation could not be straightforward, as the relationship between the conditional mean of the observations and the values of the independent variables is not linear. Figure G.4 shows the effect of the mean on the conditional probability distribution:

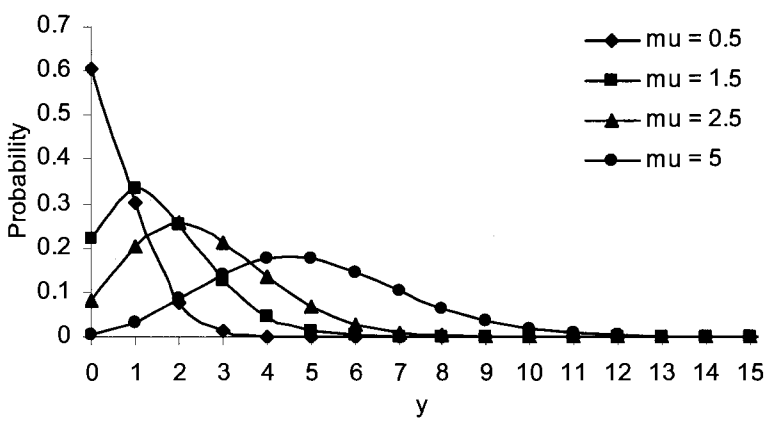

Figure G.4 : Effect of μ on the conditional probability distribution

As Long (1997) notes, different approaches are possible for interpreting a Poisson model, depending on whether the practitioner is interested in computing the probability of a count for a given value of the independent variable (or variables), or in assessing the expected value of the count for a change in the regressor (or regressors).

Let us first consider the meaning of the parameters $\underline{\theta}$ influencing the conditional mean. As it has been said, econometricians usually use an exponential conditional mean:

$$E[y \mid \underline{x}] = \exp(\underline{x}'\underline{\theta})$$

The analogous of the marginal effect examined for the logit and probit models has in this case a different meaning, as the derivative with respect to x_i is:

$$\frac{\partial E(y \mid \underline{x})}{\partial x_i} = \theta_i \cdot E(y \mid \underline{x}) = \theta_i \cdot \exp(\underline{x}'\underline{\theta})$$

The effect of a unit change in the independent variable depends on the values at which the other variables are set. This poses the usual question on which value the other variables should be set at to compute the marginal effect. The most common solution to this issue consists on setting all the variables at their means. Cameron and Trivedi (1998) propose to average the partial changes with respect of a given regressor over all observations, obtaining the following:

$$Pc = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial E[y_i \mid \underline{x}_i]}{\partial x_{ij}} = \frac{1}{N}\sum_{i=1}^{N}\theta_j \cdot \exp(\underline{x}_j'\underline{\theta})$$

The partial change can be also used to assess the relative magnitude of two parameters, independently from the values of the other variables at which it is computed:

$$\frac{\partial E[y \mid \underline{x}]/\partial x_i}{\partial E[y \mid \underline{x}]/\partial x_j} = \frac{\theta_i \cdot \exp(\underline{x}'\underline{\theta})}{\theta_j \cdot \exp(\underline{x}'\underline{\theta})} = \frac{\theta_i}{\theta_j}$$

If one of the regressors is a limited variable (binary or integer), this measures do not make sense because the derivative cannot be defined, and the discrete change is preferred:

$$\frac{\Delta E(y \mid \underline{x})}{\Delta x_i} = E\left(y \mid \underline{x}, x_i = \bar{x} + \frac{\delta}{2}\right) - E\left(y \mid \underline{x}, x_i = \bar{x} - \frac{\delta}{2}\right)$$

As usual, δ can be chosen depending on the needs and the focus of the practitioner. Some common choices are the unitary change, the estimated variance change or maximum-minimum change (which, in the case of a binary variable, is a change from 0 to 1). Clearly, also this measure depends on the level of all the variables in the model. A concise expression for the discrete change can be obtained by taking the ratio of the two expected instead of subtracting them:

$$\frac{E[y \mid \underline{x}, x_{bin} = 1]}{E[y \mid \underline{x}, x_{bin} = 0]} = \frac{\exp(\underline{x}'\underline{\theta} + \theta_{bin})}{\exp(\underline{x}'\underline{\theta})} = \exp(\theta_{bin})$$

where θ_{bin} is the parameter of the binary variable x_{bin}. This measure can be interpreted as follows: the effect of the change in a binary variable from value 0 to 1 is to multiply the conditional mean by a factor equal to $\exp(\theta_{bin})$.

Interpretation using the odds ratios is possible, and the same procedure can be followed as in the case of the logit model. The conditional mean can be written in a product form:

$$E[y \mid \underline{x}, x_i] = \exp(\theta_0) \cdot \exp(x_1\theta_1) \cdots \exp(x_i\theta_i) \cdots \exp(x_n\theta_n)$$

and, for a change of δ in the regressor x_i we obtain:

$$E[y \mid \underline{x}, x_i + \delta] = \exp(\theta_0) \cdot \exp(x_1\theta_1) \cdots \exp(x_i\theta_i)\exp(\delta \cdot \theta_i) \cdots \exp(x_n\theta_n)$$

The odds ratio is therefore:

$$\frac{E[y \mid \underline{x}, x_i + \delta]}{E[y \mid \underline{x}, x_i]} = \exp(\delta \cdot \theta_i)$$

Thus, all the other variables held constant, a change of δ in x_i increases y by a factor equal to $\exp(\delta \cdot \theta_i)$. The percentage change in the odds is:

$$\frac{E(y \mid \underline{x}; x_i + \delta) - E(y \mid \underline{x}; x_i)}{E(y \mid \underline{x}; x_i)} \times 100 = [\exp(\delta \cdot \theta_i) - 1] \times 100 .$$

Finally, we examine how the estimated parameters can be used to obtain the predicted probabilities. Let us suppose we are interested in computing the probability that the count y is equal to a certain integer c for a given $\underline{\bar{x}}$. It is first necessary to calculate the estimated conditional mean as

$$\hat{\mu} = \exp(\underline{\bar{x}}' \underline{\hat{\theta}})$$

and then put it into

$$\Pr(y = c \mid \underline{\bar{x}}) = \frac{\exp(-\hat{\mu}) \cdot \hat{\mu}^c}{c!}$$

In the case of a count dependent variable analysis, it may be useful to compute the *mean predicted probability*, which can be evaluated for each count c as:

$$\Pr(y = c) = \frac{1}{N} \sum_{i=1}^{N} \frac{\exp(-\hat{\mu}_i) \cdot \hat{\mu}_i^c}{c!}$$

A useful example of these last measures is provided by Long (1997) through an analysis of the factors affecting the number of doctoral publications.

G.4.1.2 TESTING HYPOTHESES AND TEST OF FIT

The application of the theory for simple and complex hypothesis testing considered for the case of binary response models is straightforward to models for counts. The single parameter can easily be tested through a z-test if the sample is large enough or with a t-test or pseudo-t-test when the sample is small. As well, complex hypothesis testing requires the use of the procedures examined above: Wald test, LR test or the method of the Lagrange Multipliers. The specific characteristics, advantages and shortcomings of each test remain clearly the same. The only difference is in the expression of the log-likelihood function. Among practitioners, Wald test is the most commonly used, due to the limited computational effort required. For a brief review on these methods, refer to the preceding section or, for further details, consult specialized manuals listed in bibliography.

In the following, model diagnostic is considered: residual and pseudo-R^2 measures are presented with applications to models for counts.

Overall model evaluation procedures are, as usual, divided in two broad categories: residual plot and analysis and scalar measures of fit. Conceptually, the same techniques examined for Logit and Probit models apply to the case of models for counts. However, the practitioner should be more careful when dealing with counts: as Cameron and Trivedi (1998) show, even if a model fits data in a very satisfying way, it will be rejected by any test at a conventional significance level, as the sample grows. The main concern with the analysis of residual in the case of models for counts is the definition of residual, which is not unique. Let us first consider that the asymptotic conditional distribution of the observation is Poisson, with mean and variance μ_i. This means that the residuals are not normally distributed, are not homoscedastic and are not symmetric. The definition of

the Pearson residuals provide a correction for heteroscedasticity, weighting the raw residuals $r_i = (y_i - \hat{\mu}_i)$ by the estimate of the variance of y_i:

$$p_i = \frac{(y_i - \hat{\mu}_i)}{\sqrt{\hat{s}_i}}$$

As usual, the Pearson statistic P is defined as:

$$P = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{s}_i}$$

which is the sum of the squares of the Pearson residuals. When applied to the Poisson regression, this measure deserves a particular attention, as it can be compared to the number of the total degrees of freedom $(n-k)$: considering that the expected value of the square of the Pearson residual is unitary, the expected value of the Pearson statistic is equal to the number of observations n. Now, as the Poisson distribution is equidispersed (the variance equals the mean), we can obtain a test for overdispersion by substituting $\hat{s}_i$ with $\hat{\mu}_i$ in the expression of the Pearson statistic, obtaining

$$P_P = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}.$$

This way a value of $P_P \geq (n-k)$ indicates that the data are overdispersed (the conditional variance is larger than the conditional mean). Conversely, if $P_P \leq (n-k)$ then data are underdispersed (the conditional variance is smaller than the conditional mean). The main shortcoming of using this procedure as a measure of overdispersion is that is requires that the conditional mean $\hat{\mu}_i$ is correctly estimated. If this is not the case, a

value of $P_P \neq (n-k)$ does not necessarily mean that the model is misspecified. Overdispersion is very frequent in real data: to consider this, different modifications to the Poisson regression are proposed, as we will see in the next subsection.

Another residual definition can be employed under certain hypotheses (Cameron and Trivedi, 1998). Let us take the log of the density of the observations y taken at $\mu = \hat{\mu}$ and denote it as $l(\hat{\mu})$; let us take the log density of the observations y evaluated at $\mu = y$ and denote it as $l(y)$. This way the deviance residual can be defined as:

$$r_{dev} = sign(y_i - \hat{\mu}_i) \cdot \sqrt{2 \cdot [l(y_i) - l(\hat{\mu}_i)]}$$

This measure is useful when employing the Generalized Method of Moments (GLM) for model estimation. As in this work GLM's are not employed, we do not provide further details. If interested, the reader can consult specialized texts listed in bibliography.
The last residual measure we present here is the *Anscombe residual*, which implies a transformation in the observations, to make their distribution as close as possible to the normal distribution. McCullagh and Nedler (1989) develop this method and provide the following definition:

$$r_a = 1.5 \cdot \frac{y_i^{2/3} - \mu_i^{2/3}}{\mu_i^{1/6}}$$

Other more generalized definitions of residuals can be found in the literature. Cameron and Trivedi (1998) provide a review of this methods with applications to models for counts. For more details, specialized texts and articles can be consulted, as Cox and Snell (1968), (Gourieoux et al., 1987a), (Gourieoux et al., 1987b) and (Pierce and Schafer, 1986).

Along with the analysis of the residuals, scalar measures of fit are commonly employed. As models for count variables are not linear, the definition of R^2 must be modified and adapted to the characteristics of the fitted model. As we saw for binary response models, different pseudo-R^2 can be defined. The first derivation we provide builds on the decomposition of the Total Sum of Squares TSS in the Explained Sum of Squares ESS and the Residual Sum of Squares RSS. This way

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

In particular, the ratio in the second form can be interpreted as the ratio between the deviation from the mean that is due to the residuals and the total deviance. The term "deviance" is commonly used in the GLM approach, and represents the generalization of the sum of squares. Here we do not define it rigorously; if interested, the reader can consult Cameron and Windmeijer (1996). It can be shown that in the case of a Poisson model, the deviance can be expressed as (Bishop et al., 1975):

$$D_{Poi} = \sum_{i=1}^{n} \left[y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i) \right]$$

This way the pseudo-R^2 based on the deviance is expressed as:

$$R^2_{D,Poi} = \frac{\sum_{i=1}^{n} y_i \cdot \ln\left(\frac{\hat{\mu}_i}{\bar{y}}\right) - (y_i - \hat{\mu}_i)}{\sum_{i=1}^{n} y_i \cdot \ln\left(\frac{\hat{\mu}_i}{\bar{y}}\right)}$$

It is to be noted that if $y = 0$, than it is assumed that $y \cdot \ln(y) = 0$.

Another common scalar measure of fit considers the difference in the log-likelihood functions of the model estimated with just the intercept and the log-likelihood of the full model. When this measure, as it has been shown for the binary response models, is calculated with respect to the maximum value that can be achieved by the likelihood function we obtain, in the case of the Poisson model (Merkle and Zimmermann, 1992):

$$R^2_{ML,Adj} = \frac{\ln L_{full} - \ln L_{\text{int}}}{\ln L_{MAX} - \ln L_{\text{int}}}$$

However, although in the case of the logit model it is easy to calculate the maximum value achievable by the log-likelihood function, in the case of the Poisson model this is not straightforward. Actually, for some generalizations of the Poisson model, this value is not defined, and the log-likelihood function can grow to infinity; in this case, $R^2_{ML,Adj}$ is always equal to 0.

As it has been said in the preceding subsection, it would be necessary to analyze both the residuals and the scalar measures of fit to assess the goodness of fit in order to compare and choose between two or more models. However, even after a complete analysis based on quantitative measures, the practitioner should not be blind in his choice, and should also examine the qualitative aspects of the model (the outliers, for example) before deciding to keep or to discard it.

G.4.2 NEGATIVE BINOMIAL REGRESSION

The main concern with the Poisson regression is that it does not take into account overdispersion in the data, which is frequent indeed in real data. As it has been observed, overdispersion can be the cause of a biased estimation of the Pearson statistic, and produce misleading results and wrong interpretation. It can be shown that overdispersion causes a downward bias on the standard deviation estimates (Cameron and Trivedi, 1986) and the effect of the parameters are overestimated. This limitation of the Poisson

regression model can be overcome by modifying the conditional probability density adding a parameter that takes overdispersion (or underdispersion) into account. In other words, the negative binomial model does not assume equality between the conditional mean and the conditional variance. The most common implementation employs a quadratic relationship between the conditional mean and the conditional variance:

$$V[y \mid \underline{x}] = \mu + \alpha \cdot \mu^2$$

where $\alpha > 0$. Clearly, for $\alpha = 0$ we obtain the Poisson regression model. The conditional probability density function is:

$$\Pr(y \mid \underline{x}, \alpha, \underline{\theta}) = \frac{\Gamma(y + \alpha^{-1})}{\Gamma(y+1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu} \right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1} + \mu} \right)^{y}$$

where $\Gamma(\arg)$ denotes the gamma distribution. Once the density function is defined, the likelihood function can be easily computed

$$L(\underline{\theta} \mid \underline{y}, \underline{x}) = \prod_{i=1}^{n} \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i} \right)^{y_i}$$

where an exponential conditional mean is employed. Cameron and Trivedi (1998) show that, after some manipulation, the log-likelihood function can be written as:

$$\ln L(\underline{\theta} \mid \underline{y}, \underline{x}) = \sum_{i=1}^{n} \left\{ \left(\sum_{j=0}^{y_i - 1} \ln(j + \alpha^{-1}) \right) - \ln(y_i!) - (y_i + \alpha^{-1}) \ln(1 + \alpha \cdot \exp(\underline{x}' \underline{\theta})) \right.$$
$$\left. + y_i \cdot \ln \alpha + y_i \underline{x}' \underline{\theta} \right\}$$

As usual, approximate numerical methods must be used to estimate the parameters. Different variations of the negative binomial model exist, depending on the assumed relationship between the conditional variance and the conditional mean. For example, we could set the more general form

$$V[y \mid \underline{x}] = \mu + \alpha \cdot \mu^m$$

In this case, two extra parameters are introduced in the model. When $m = 2$ we obtain the usual form of negative binomial regression model introduced above. When $\alpha = 0$ we obtain instead the Poisson regression model. A common choice is to set $m = 1$ obtaining what Cameron and Trivedi (1998) call the NB1 model.

As the negative binomial regression model is used in order to relax the equidispersion hypothesis in the conditional probability density, it is important to test for overdispersion. A simple and intuitive way to do this is through the comparison between the value that maximizes the log-likelihood function of the negative binomial model and the value that maximizes the log-likelihood function of the Poisson regression model (which is the negative binomial when α is set to 0). This way we obtain the statistic:

$$G^2 = 2 \cdot (\ln L_{NBM} - \ln L_{NBR})$$

that must be compared with a critical value of X^2_{2p} where p is the confidence level[42]. Alternatively, the null hypothesis $H_0 : \alpha = 0$ can be directly tested through a one-tailed z-test .

This test can be avoided when overdispersion is obvious, as in the case of an evident excess of 0's.

[42] As $\alpha > 0$ the test is one-tailed.

G.4.2.1 INTERPRETATION

The methods examined for the Poisson regression model remain the same in the case of the NBR, and will not be explained here. The only difference is, of course, in using predicted probabilities, which for the negative binomial model are defined as:

$$\Pr(y \mid \underline{x}) = \frac{\Gamma\left(y + \hat{\alpha}^{-1}\right)}{\Gamma(y+1)\Gamma\left(\hat{\alpha}^{-1}\right)}\left(\frac{\hat{\alpha}^{-1}}{\hat{\alpha}^{-1} + \hat{\mu}}\right)^{\alpha^{-1}}\left(\frac{\hat{\mu}}{\hat{\alpha}^{-1} + \hat{\mu}}\right)^{y}$$

It is interesting to qualitatively explain the effect of the parameter α on the conditional probability distribution for a given value of μ. This effect is qualitatively the same as a change in the parameter μ has on the conditional distribution in the case of the Poisson regression model. This way, for a given value of the independent variable, the parameter α stretches or compresses the probability function. In particular, as the value of the parameter α grows, the conditional distribution is compressed towards the vertical axes: the probability of the count to be 0 grows.

The methods for testing simple hypotheses, complex hypotheses and the overall behavior of the model are the same as in the case of Poisson model. The only interesting difference is in the definition of the pseudo-R^2 coefficient calculated considering the generalized sum of squares, or deviances. For the negative binomial model, if $\alpha = 2$ we obtain the following:

$$D_{NBR} = \sum_{i=1}^{n}\left\{y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - \left(y_i + \alpha^{-1}\right)\ln\left(\frac{y_i + \alpha^{-1}}{\hat{\mu}_i + \alpha^{-1}}\right)\right\}$$

This way, the so-called deviance pseudo-R^2 for the NBR model is

$$R^2_{D,NBR} = 1 - \frac{\sum_{i=1}^{n}\left\{ y_i \ln\left(\frac{\hat{\mu}_i}{\bar{y}}\right) - (y_i + \hat{\alpha})\ln\left(\frac{y_i + \hat{\alpha}}{\hat{\mu}_i + \hat{\alpha}}\right)\right\}}{\sum_{i=1}^{n}\left\{ y_i \ln\left(\frac{\hat{y}_i}{\bar{y}}\right) - (y_i + \hat{\alpha})\ln\left(\frac{y_i + \hat{\alpha}}{\bar{y} + \hat{\alpha}}\right)\right\}}$$

One could be tempted to compare the deviance pseudo-R^2 of the Poisson regression with the deviance pseudo-R^2 of the negative binomial regression. This procedure is wrong: the two measures have different denominators and are indeed differently defined for each model.